

PhD Dissertation

Nonparametric inference with directional and linear data

Eduardo García Portugués

Departamento de Estatística e Investigación Operativa

Universidade de Santiago de Compostela

September 2014

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Nonparametric inference with directional and linear data

Eduardo García Portugués

Supervised by Wenceslao González Manteiga and Rosa M. Crujeiras

This work has been supported by FPU grant AP2010–0957 from *Ministerio de Educación*, graduate grant for research stays from *Fundación Barrié*, Project MTM2008–03010 from *Ministerio de Ciencia e Innovación*, Project 10MDS207015PR from *Dirección Xeral de I+D da Xunta de Galicia* and IAP network StUDyS from Belgian Science Policy. The author gratefully acknowledges the computational resources used at the SVG cluster of the CESGA Supercomputing Center that allowed to run most of the simulations.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Informe de los directores

Don Wenceslao González Manteiga, Catedrático del Departamento de Estatística e Investigación Operativa de la Universidade de Santiago de Compostela y Doña Rosa M. Crujeiras Casais, Profesora Contratada Doctora en el mismo departamento, como directores de la tesis titulada **"Nonparametric inference with directional and linear data"**

Por la presente, DECLARAN

Que la tesis doctoral presentada por Don Eduardo García Portugués es idónea para ser presentada, de acuerdo con el artículo 41 del Reglamento de Estudios de Doctorado, por la modalidad de compendio de ARTÍCULOS, en los que el doctorando tuvo participación en el peso de la investigación y su contribución fue decisiva para llevar a cabo este trabajo.

Y que está en conocimiento de los coautores, tanto doctores como no doctores, participantes en los artículos, que ninguno de los trabajos reunidos en esta tesis serán presentados por ninguno de ellos en otra tesis de doctorado, lo que firmamos bajo nuestra responsabilidad.

En Santiago de Compostela, a 3 de septiembre de 2014.

Los directores:

Prof. Dr. Wenceslao González Manteiga

Prof. Dr. Rosa M. Crujeiras Casais

El doctorando:

Eduardo García Portugués

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Agradecimientos

Quiero comenzar estos agradecimientos expresando mi profunda gratitud por el apoyo recibido por parte de mis directores, Wenceslao González Manteiga y Rosa M. Crujeiras, durante todo el proceso de realización de esta tesis doctoral. A *Wences* le agradezco el haber despertado mi interés por la estadística en mis últimos años de licenciatura, la confianza que depositó en mí desde mis primeros pasos como investigador y la continua motivación que me ha transmitido. Estoy muy agradecido a Rosa por su apoyo diario, tanto profesional como personal, a lo largo de estos años. Valoro especialmente sus innumerables y exhaustivas revisiones críticas, que han servido para perfeccionar notablemente el trabajo desarrollado en esta tesis. La verdad es que me siento enormemente afortunado por haberla tenido como codirectora. I would like also to thank Ingrid Van Keilegom for her hospitality and her guidance in research during my stay at Louvain-la-Neuve.

Gracias a mi madre y a mi padre, a vosotros os lo debo todo en esta vida. Gracias también a mi familia por haberme apoyado en esta y en otras etapas y, en especial, a mi abuela Isabel Del Río Galza y a mi abuelo Rafael Portugués Carrera. Que esta tesis sirva como mi pequeño homenaje a los grandes esfuerzos que ambos realizaron por el bienestar de la familia.

Finalmente, agradezco a mis compañeros del Departamento de Estadística e Investigación Operativa y de la Facultad de Matemáticas por los buenos momentos pasados durante estos años.

Santiago de Compostela, 3 de septiembre de 2014.
Eduardo García Portugués

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Contents

Informe de los directores **iii**

Agradecimientos **v**

Contents **x**

List of Figures **xvi**

List of Tables **xviii**

1 Introduction **1**
1.1 What is directional data? . . . 1
1.2 Contributions of the thesis . . . 5
1.3 Real datasets . . . 8
1.3.1 Wind direction . . . 9
1.3.2 Hipparcos dataset . . . 10
1.3.3 Portuguese wildfires . . . 11
1.3.4 Protein angles . . . 12
1.3.5 Text mining . . . 13
1.4 Manuscript distribution . . . 15
References . . . 16

2 Density estimation by circular-linear copulas **21**
2.1 Introduction . . . 21
2.2 Background . . . 24
2.2.1 Some circular and circular-linear distributions . . . 24
2.2.2 Some notes on copulas . . . 25
2.3 Estimation algorithm . . . 27
2.3.1 Some simulation results . . . 30
2.4 Application to wind direction and SO_2 concentration . . . 35
2.4.1 Exploring SO_2 and wind direction in 2004 and 2011 . . . 36
2.5 Final comments . . . 36
References . . . 38

3 Kernel density estimation for directional-linear data **41**
3.1 Introduction . . . 42
3.2 Background on linear and directional kernel density estimation . . . 43
3.2.1 Kernel density estimation for directional data . . . 43

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3.2.2 Kernel density estimation for directional-linear data . . . 47
3.3 Main results . . . 47
3.4 Error measurement and optimal bandwidth . . . 49
3.4.1 MISE for directional and directional-linear kernel density estimators . . . 49
3.4.2 Some exact MISE calculations for mixture distributions . . . 51
3.5 Conclusions . . . 55
3.A Some technical lemmas . . . 56
3.B Proofs of the main results . . . 57
3.C Proofs of the technical lemmas . . . 70
References . . . 76

4 Bandwidth selectors for kernel density estimation with directional data 79
4.1 Introduction . . . 80
4.2 Kernel density estimation with directional data . . . 81
4.2.1 Available bandwidth selectors . . . 84
4.3 A new rule of thumb selector . . . 85
4.4 Selectors based on mixtures . . . 86
4.4.1 Mixtures fitting and selection of the number of components . . . 89
4.5 Comparative study . . . 90
4.5.1 Directional models . . . 90
4.5.2 Circular case . . . 92
4.5.3 Spherical case . . . 94
4.5.4 The effect of dimension . . . 96
4.6 Data application . . . 97
4.6.1 Wind direction . . . 97
4.6.2 Position of stars . . . 97
4.7 Conclusions . . . 98
4.A Proofs . . . 99
4.B Models for the simulation study . . . 99
4.C Extended tables for the simulation study . . . 100
References . . . 104

5 A nonparametric test for directional-linear independence 107
5.1 Introduction . . . 108
5.2 Background to kernel density estimation . . . 110
5.2.1 Linear kernel density estimation . . . 110
5.2.2 Directional kernel density estimation . . . 110
5.2.3 Directional-linear kernel density estimation. . . . 111
5.3 A test for directional-linear independence . . . 112
5.3.1 The test statistic . . . 112
5.3.2 Calibration of the test . . . 113
5.3.3 Simulation study . . . 115
5.4 Real data analysis . . . 120
5.4.1 Data description . . . 120
5.4.2 Results . . . 121
5.5 Discussion . . . 122
5.A Proof of Lemma 5.1 . . . 124

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

References . . . 125

6 Central limit theorems for directional and linear random variables with applications **129**
6.1 Introduction . . . 130
6.2 Background . . . 131
6.3 Central limit theorem for the integrated squared error . . . 133
6.3.1 Main result . . . 133
6.3.2 Extensions of Theorem 6.1 . . . 134
6.4 Testing independence with directional random variables . . . 135
6.5 Goodness-of-fit test with directional random variables . . . 136
6.5.1 Testing a simple null hypothesis . . . 136
6.5.2 Composite null hypothesis . . . 137
6.5.3 Calibration in practise . . . 138
6.5.4 Extensions to directional-directional models . . . 138
6.6 Simulation study . . . 139
6.7 Real data application . . . 142
6.A Sketches of the main proofs . . . 143
6.A.1 CLT for the integrated squared error . . . 143
6.A.2 Testing independence with directional data . . . 145
6.A.3 Goodness-of-fit test for models with directional data . . . 146
References . . . 147

7 Testing parametric models in linear-directional regression **151**
7.1 Introduction . . . 152
7.2 Nonparametric linear-directional regression . . . 153
7.2.1 The projected local estimator . . . 154
7.2.2 Properties . . . 155
7.3 Goodness-of-fit test for linear-directional regression . . . 156
7.3.1 Bootstrap calibration . . . 158
7.4 Simulation study . . . 160
7.5 Application to text mining . . . 162
7.A Main results . . . 165
7.A.1 Projected local estimator properties . . . 165
7.A.2 Asymptotic results for the goodness-of-fit test . . . 168
References . . . 172

8 Future research **175**
8.1 Bandwidth selection in nonparametric linear-directional regression . . . 175
8.2 Kernel density estimation under rotational symmetry . . . 180
8.3 `R` package `DirStats` . . . 185
8.4 A goodness-of-fit test for the Johnson and Wehrly copula structure . . . 187
References . . . 189

A Supplement to Chapter 6 **193**
A.1 Technical lemmas . . . 193
A.1.1 CLT for the ISE . . . 193
A.1.2 Testing independence with directional data . . . 210

A.1.3 Goodness-of-fit test for models with directional data . . . 221
A.1.4 General purpose lemmas . . . 223
A.2 Further results for the independence test . . . 227
A.2.1 Closed expressions . . . 227
A.2.2 Extension to the directional-directional case . . . 228
A.2.3 Some numerical experiments . . . 229
A.3 Extended simulation study . . . 229
A.3.1 Parametric models . . . 229
A.3.2 Estimation . . . 231
A.3.3 Simulation . . . 232
A.3.4 Alternative models . . . 235
A.3.5 Bandwidth choice . . . 235
A.3.6 Further results . . . 235
A.4 Extended data application . . . 239
References . . . 240

B Supplement to Chapter 7 **243**
B.1 Particular cases of the projected local estimator . . . 243
B.1.1 Local constant . . . 243
B.1.2 Local linear with $q = 1$ . . . 244
B.1.3 Local linear with $q = 2$ . . . 245
B.2 Technical lemmas . . . 246
B.2.1 Projected local estimator properties . . . 246
B.2.2 Asymptotic results for the goodness-of-fit test . . . 252
B.2.3 General purpose lemmas . . . 262
B.3 Further simulation results . . . 263

Resumen en castellano **271**
Bibliografía del resumen . . . 282

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

List of Figures

1.1 Polar coordinates in Ω_1 and spherical coordinates in Ω_2 (just an octant of the sphere is represented for a better display). 2

1.2 Von Mises densities on Ω_1 and Ω_2. From right to left, densities corresponding to $q = 1$ with κ_1 and κ_2 and to $q = 2$ with κ_1 and κ_2, respectively. The parameters are $\boldsymbol{\mu} = (\mathbf{0}_q, 1)$, $\kappa_1 = 2$ and $\kappa_2 = 5$, with $\mathbf{0}_q$ denoting a vector of q zeros. Samples of size $n = 250$ are drawn. 4

1.3 From right to left, densities corresponding to: MS $\left(\frac{3\pi}{2}, 5, 0, \frac{1}{2}, -\frac{3}{4}, \frac{3}{2}\right)$; density (1.1) with a vM $\left(\frac{5\pi}{4}, \frac{3}{2}\right)$ link and marginals $\mathcal{N}(0, 1)$ and circular uniform; density (1.1) with a vM$(\pi, 7)$ link and marginals JP $\left(0, \frac{1}{2}, 1\right)$ and uniform circular; WNT $\left(0, \frac{\pi}{6}, \frac{3}{2}, \frac{1}{4}, 0\right)$. Samples of size $n = 300$ are drawn. 5

1.4 Diagram with the contributions of the thesis (marked in bold) and a short state of the art with the main references in the scope of the thesis. Future research lines described in Chapter 8 are included in *italics*. 6

1.5 Rose diagrams for wind direction in station B2 for January 2004 (left), January 2011 (center) and in station A mourela in June 2012 (right). The first two ones show the average SO_2 concentration in $\mu g/m^3$ associated with winds coming from each partition of the windrose. 9

1.6 Number of observed stars per square degree, in galactic coordinates (cell size $2° \times 2°$), extracted from Figure 3.2.1 of Perryman (1997). The higher concentrations of stars are located around the equator (galactic plane) and two spots that represent the Orion's arm (left) and the Gould's Belt (right). 11

1.7 Number of hectares burnt in 1985–2005 in each watershed delineated by Barros et al. (2012) (left plot) and main orientation of a wildfire obtained from the PC1 of the perimeter (right plot; extracted from Barros et al. (2012)). 12

1.8 Representation of the dihedral angles in the primary protein structure (left, adapted from Richardson (2011)) and Ramachadran plot of the dihedral angles of the alanine-alanine-alanine segments contained in the `ProteinsAAA` object of Fernández-Durán and Gregorio-Domínguez (2013). 13

1.9 View of the *Slashdot* website on July 27, 2014. 14

2.1 Locations of monitoring station (circle) and power plant in Galicia (NW-Spain). Location of station B2: 7° 44' 10" W, 43° 32' 05" N. Power plant location: 7° 51' 45" W, 43° 26' 26" N. 22

2.2 Rose diagrams for wind direction in station B2 for 2004 (left plot) and 2011 (right plot), with average SO_2 concentration. 23

2.3 Copula density surfaces. Left plot: J&W copula with von Mises joining density with parameters $\mu = \pi$ and $\kappa = 2$. Middle plot: QS-copula with $\alpha = (2\pi)^{-1}$. Right plot: reflected Frank copula with $\alpha = 10$. 27

2.4 Illustration for data reflection for copula density estimation. The central square in each plot corresponds with the original data ranks. Left plot: mirror reflection from Gijbels and Mielniczuk (1990). Right plot: circular-mirror reflection for estimator (2.11). . . . 29

2.5 Density surfaces for the simulation study. Top left: Example 2.1 with $\mu = \pi$ and $\kappa = 2$. Top right: Example 2.2 with $\mu' = \pi$, $\kappa' = 5$, $\mu_2 = \pi/2$ and $\kappa_2 = 2$. Bottom left: Example 2.3 with $\alpha = (2\pi)^{-1}$, $\mu_3 = \pi/2$ and $\kappa_3 = 0.5$. Bottom right: Example 2.4 with $\alpha = 10$, $\mu_4 = \pi/2$ and $\kappa_4 = 0.5$. 31

2.6 Boxplots of the ISE $\times$ 100 for Example 2.1 (top left), Example 2.2 (top right), Example 2.3 (bottom left) and Example 2.4 (bottom right) for $n = 500$ and different estimation procedures. 34

2.7 Circular-linear density estimator for wind direction and SO_2 concentrations in monitoring station B2. Left column: 2004. Right column: 2011. 37

3.1 Left: contour plot of a von Mises density $\text{vM}(\boldsymbol{\mu}, \kappa)$, with $\boldsymbol{\mu} = (0, 0, 1)$ and $\kappa = 3$. Right: contour plot of the mixture of von Mises densities (3.14). 46

3.2 From left to right: circular-linear mixture (3.15) and corresponding circular and linear marginal densities, respectively. Random samples of size $n = 200$ are drawn. 52

3.3 From left to right: exact MISE and AMISE for the linear mixture $\frac{2}{5}\mathcal{N}\left(0, \frac{1}{4}\right) + \frac{2}{5}\mathcal{N}(1, 1) + \frac{1}{5}\mathcal{N}(2, 1)$ and the circular and spherical mixtures (3.14), for a range of bandwidths between 0 and 1. The black curves are for the MISE, whereas the red ones are for the AMISE. Solid curves correspond to $n = 100$ and dotted to $n = 1000$. Vertical lines represent the bandwidth values minimizing each curve. 53

3.4 Upper plot, from left to right: exact MISE versus AMISE for the circular-linear mixture (3.15) for $n = 100$ and $n = 1000$. Lower plot, from left to right: spherical-linear mixture (3.15) for $n = 100$ and $n = 1000$. The solid curves are for the MISE, where the dashed ones are for the AMISE. The pairs of bandwidths that minimizes each surface error are denoted by $(h, g)_{\text{MISE}}$ and by $(h, g)_{\text{AMISE}}$. 54

4.1 The effect of the "extra term" in h_{ROT}. Left plot: logarithm of the curves of MISE(h_{TAY}), MISE(h_{ROT}) and MISE(h_{MISE}) for sample size $n = 250$. The curves are computed by 1000 Monte Carlo samples and h_{MISE} is obtained exactly. The abscissae axis represents the variation of the parameter $\theta \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$, which indexes the reference density $\frac{1}{2}\text{vM}\left((0, 1), 2\right) + \frac{1}{2}\text{vM}\left((\cos(\theta), \sin(\theta)), 2\right)$. Right plot: logarithm of h_{TAY}, h_{ROT}, h_{MISE} and their corresponding MISE for different values of κ, with $n = 250$. 86

4.2 Simulation scenarios for the circular case. From left to right and up to down, models M1 to M20. For each model, a sample of size 250 is drawn. 93

4.3 Simulation scenarios for the spherical case. From left to right and up to down, models M1 to M20. For each model, a sample of size 250 is drawn. 95

4.4 Left: density of the wind direction in the meteorological station of A Mourela. Right: density of the stars collected in the Hipparcos catalogue, represented in galactic coordinates and Aitoff projection. 97

5.1 Descriptive maps of wildfires in Portugal with the 102 watersheds delineated by Barros et al. (2012). The left map shows the number of hectares burnt from fire perimeters associated with each watershed. Each fire perimeter is associated with the watershed that contains its centroid. The center map represents the mean slope of the fires of each watershed, where the slope is measured in degrees (0° stands for plain slope and 90° for a vertical one). Finally, the right map shows watersheds where fires display preferential alignment according to Barros et al. (2012). 109
5.2 Random samples of $n = 500$ points for the simulation models in the circular-linear case, with $\delta = 0.50$ (situation with dependence). From left to right and up to down, M1 to M6. M1, M2 and M3 present a deviation from the independence in terms of the conditional expectation; M4 and M5 account for a deviation in terms of the conditional variance and M6 includes deviations both in conditional expectation and variance. 116
5.3 p-values from the independence test for the first principal component PC1 of the fire perimeter and the burnt area (on a log scale), by watersheds. From left to right, the first and second maps represent the circular-linear p-values (PC1 in $\mathbb{R}^2$) and their corrected versions using the FDR, respectively. The third and fourth maps represent the spherical-linear situation (PC1 in $\mathbb{R}^3$), with uncorrected and corrected p-values by FDR, respectively. 121
5.4 Left: density contour plot for fires in watershed number 31, the watershed in the second plot on the left of Figure 5.3 with p-value = 0.000. The number of fires in the watershed is $n = 1543$. The contour plot shows that the size of the area burnt is related with the orientation of the fires in the watershed. Right: scatter plot of the fires slope and the burnt area for the whole dataset, with a nonparametric kernel regression curve showing the negative correlation between fire slope and size. 122

6.1 Density models for the simulation study in the circular-linear case. From left to right and up to down, models CL1 to CL12. 140
6.2 Density models for the simulation study in the circular-circular case. From left to right and up to down, models CC1 to CC12. 140
6.3 Empirical size and power of the circular-linear (left, model CL1) and circular-circular (right, model CC8) goodness-of-fit tests for a 10×10 logarithmic spaced grid. Lower surface represents the empirical rejection rate under $H_{0.00}$ and upper surface under $H_{0.15}$. Green colour indicates that the percentage of rejections is in the 95% confidence interval of $\alpha = 0.05$, blue that is smaller and orange that is larger. Black points represent the empirical size and power obtained with the median of the LCV bandwidths. 141
6.4 Left: parametric fit (model from Mardia and Sutton (1978)) to the circular mean orientation and mean log-burnt area of the fires in each of the 102 watersheds of Portugal. Right: parametric fit (model from Fernández-Durán (2007)) for the dihedral angles of the alanine-alanine-alanine segments. 142

7.1 QQ-plot comparing the quantiles of the asymptotic distribution given by Theorem 7.3 with the sample quantiles for $\left\{nh^{\frac{1}{2}}\left(T_n^j - \frac{\sqrt{\pi}}{4}nh\right)\right\}_{j=1}^{500}$ with $n = 10^2$ (left) and $n = 5 \times 10^5$ (right). 159
7.2 Parametric regression models for scenarios S1 to S4, for circular and spherical cases. . . 161
7.3 Empirical sizes (first row) and powers (second row) for significance level $\alpha = 0.05$ for the different scenarios, with $p = 0$ (solid line) and $p = 1$ (dashed line). From left to right, columns represent dimensions $q = 1, 2, 3$ with sample size $n = 100$. 162
7.4 Significance trace of the local constant goodness-of-fit test for the constrained linear model. 163

8.1 Directional regression models proposed for the simulation study. From left to right and up to down, models used in scenarios S1 to S12, with the first three rows fo the circular versions and the last three for the spherical ones. 179
8.2 Directional densities to be considered in the simulation study. From left to right and up to down, the first two columns represent the directional densities D1 to D6 in the circular case. The corresponsing spherical versions are in the third and fourth columns. For each density, a sample of size $n = 250$ is drawn. 180
8.3 From right to left: vM$(\boldsymbol{\theta}, \kappa)$ density, rotational kernel density estimators $\hat{f}_{h_{\mathrm{LCV}},\boldsymbol{\theta}}$ and $\hat{f}_{h_{\mathrm{LCV}},\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}}$ and usual kernel density estimator $\hat{f}_{h_{\mathrm{LCV}}}$. Sample size is $n = 100$, $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$ and $\kappa = 2$. 184
8.4 From right to left: contourplots for copula C_G, copula C_{G_n} and empirical copula C_n, for a sample size of $n = 50$ with $q = -1$ and $g(\theta) = (2\pi\mathcal{I}_0(\kappa))^{-1} e^{\kappa\cos(\theta-\mu)}$ a circular von Mises density with $\mu = \frac{\pi}{2}$ and $\kappa = 2$. 189

A.1 Comparison of the asymptotic and empirical distributions of $(nh_n^q g_n)^{\frac{1}{2}}(T_n - A_n)$ for sample sizes $n = 5^j \times 10^k$, $j = 0, 1$, $k = 2, 3, 4, 5$. Black curves represent a kernel estimation from 1000 simulations, green curves represent a normal fit to the unknown density and red curves represent the theoretical asymptotic distribution. 230
A.2 Empirical size and power of the goodness-of-fit tests for a 10×10 grid of bandwidths. First two rows, from left to right and up to down: models CL1, CL5, CL7, CL8, CL9 and CL11. Last two rows: CC1, CC5, CC7, CC8, CC9 and CC11. Lower surface represents the empirical rejection rate under $H_{0.00}$ and upper surface under $H_{0.15}$. Green colour represent that the empirical rejection is in the 95% confidence interval of $\alpha = 0.05$, blue that is lower and orange that is larger. Black points represent the sized and powers obtained with the median of the LCV bandwidths (for model CC1 under H_0 is outside the grid). 236
A.3 Upper row, from left to right: parametric fit (model from Mardia and Sutton (1978)) to the circular mean orientation and mean log-burnt area of the fires in each of the 102 watersheds of Portugal; parametric fit (model from Fernández-Durán (2007)) for the dihedral angles of the alanine-alanine-alanine segments. Lower row: p-values of the goodness-of-fit tests for a 10×10 grid, with the LCV bandwidth for the data. 239

B.1 From left to right: directional densities for scenarios S1 to S4 for circular and spherical cases. 264
B.2 From left to right: deviations Δ_1 and Δ_2 and conditional standard deviation function σ_2 for circular and spherical cases. 264
B.3 Densities of the response Y under the null (solid line) and under the alternative (dashed line) for scenarios S1 to S4 (columns, from left to right) and dimensions $q = 1, 2, 3$ (rows, from top to bottom). 265
B.4 Empirical sizes for $\alpha = 0.01$ (first row), $\alpha = 0.05$ (second row) and $\alpha = 0.10$ (third row) for the different scenarios, with $p = 0$ (solid line) and $p = 1$ (dashed line). From left to right, columns represent dimensions $q = 1, 2, 3$ with sample size $n = 100$. 266
B.5 Empirical sizes for $\alpha = 0.01$ (first row), $\alpha = 0.05$ (second row) and $\alpha = 0.10$ (third row) for the different scenarios, with $p = 0$ (solid line) and $p = 1$ (dashed line). From left to right, columns represent dimensions $q = 1, 2, 3$ with sample size $n = 250$. 267
B.6 Empirical sizes for $\alpha = 0.01$ (first row), $\alpha = 0.05$ (second row) and $\alpha = 0.10$ (third row) for the different scenarios, with $p = 0$ (solid line) and $p = 1$ (dashed line). From left to right, columns represent dimensions $q = 1, 2, 3$ with sample size $n = 500$. 268

B.7 Empirical powers for the different scenarios, with $p = 0$ (solid line) and $p = 1$ (dashed line). From top to bottom, rows represent sample sizes $n = 100, 250, 500$ and from left to right, columns represent dimensions $q = 1, 2, 3$. 269

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

List of Tables

2.1 MISE × 100 for estimating the circular-linear density in Examples 2.1 and 2.2. Relative efficiencies for JWSP, JWNP, CSP and CNP are taken with respect to JWP. 33

2.2 MISE × 100 for estimating the circular-linear density in Examples 2.3 and 2.4. 33

4.1 Comparative study for the circular case, with sample size $n = 500$. Columns of the selector • represent the MISE(•) × 100, with bold type for the minimum of the errors. The standard deviation of the ISE × 100 is given between parentheses. 94

4.2 Ranking for the selectors for the circular and spherical cases, for sample sizes $n = 100, 250, 500, 1000$. The higher the score in the ranking, the better the performance of the selector. Bold type indicates the best selector. 94

4.3 Comparative study for the spherical case, with sample size $n = 500$. Columns of the selector • represent the MISE(•) × 100, with bold type for the minimum of the errors. The standard deviation of the ISE × 100 is given between parentheses. 96

4.4 Ranking for the selectors for dimensions $q = 3, 4, 5$ and sample size $n = 1000$. The larger the score in the ranking, the better the performance of the selector. Bold type indicates the best selector. 96

4.5 Directional densities considered in the simulation study. 100

4.6 Comparative study for the circular case, with up to down blocks corresponding to sample sizes 100, 250 and 1000, respectively. Columns of the selector • represent the MISE(•) × 100, with bold type for the minimum of the errors. The standard deviation of the ISE×100 is given between parentheses. 101

4.7 Comparative study for the spherical case, with up to down blocks corresponding to sample sizes 100, 250 and 1000, respectively. Columns of the selector • represent the MISE(•) × 100, with bold type for the minimum of the errors. The standard deviation of the ISE×100 is given between parentheses. 102

4.8 Comparative study for higher dimensions with sample size $n = 1000$: up to down blocks correspond to dimensions $q = 3, 4, 5$. Columns of the selector • represent the MISE(•) × 100, with bold type for the minimum of the errors. The standard deviation of the ISE×100 is given between parentheses. 103

5.1 Proportion of rejections for the R_n^2, U_n, λ_{4n}^*, T_n^{LCV} and T_n^{BLCV} tests of independence for sample sizes $n = 50, 100, 200$ for a nominal significance level of 5%. For the six different models the values of the deviation from independence parameter are $\delta = 0$ (independence), 0.25 and 0.50. Each proportion was calculated using $B = 1000$ permutations for each of $M = 1000$ random samples of size n simulated from the specified model. 118

5.2 Proportion of rejections for the R_n^2, U_n, λ_{4n}^*, T_n^{LCV} and T_n^{BLCV} tests of independence for sample sizes $n = 500, 1000$ for a nominal significance level of 5%. For the six different models the values of the deviation from independence parameter are $\delta = 0$ (independence), 0.25 and 0.50. Each proportion was calculated using $B = 1000$ permutations for each of $M = 1000$ random samples of size n simulated from the specified model. 119
5.3 Computing times (in seconds) for T_n^{LCV} and T_n^{BLCV} as a function of sample size and dimension q, with $q = 1$ for the circular-linear case and $q = 2$ for the spherical-linear case. The tests were run with $B = 1000$ permutations and the times were measured in a 3.5 GHz core. 119

6.1 Empirical size and power of the circular-linear and circular-circular goodness-of-fit tests for models CL1–CL12 and CC1–CC12 (respectively) with significance level $\alpha = 0.05$ and different sample sizes and deviations. 141

7.1 Specification of simulation scenarios for model (7.3). 161
7.2 Fitted constrained linear model on the Slashdot dataset, with $R^2 = 0.25$. The significances of each coefficient are lower than 0.002. 164

8.1 Simulation scenarios for comparing the bandwidth selectors. f_{SN} is the Skew Normal distribution of Azzalini (1985), while the rest of notations can be seen in Subsection 4.5.1. 181

A.1 Notation for the densities described in Tables A.2 and A.3. 231
A.2 Circular-linear models. 233
A.3 Circular-circular models. 234
A.4 Empirical size and power of the circular-linear goodness-of-fit test for models CL1–CL12 with different sample sizes, deviations and significance levels. 237
A.5 Empirical size and power of the circular-circular goodness-of-fit test for models CC1–CC12 with different sample sizes, deviations and significance levels. 238

Chapter 1

Introduction

An introduction to basic concepts and statistical tools for analysing directional data is provided in this chapter. Section 1.1 presents a short introduction to the analysis of directional data and to the most well known directional distributions. The contributions of this thesis are summarized in Section 1.2, with a brief state of the art on nonparametric inference with directional and linear data. Section 1.3 describes the real datasets used along the manuscript. Finally, Section 1.4 presents the thesis distribution and organization, with short abstracts describing the contents of each chapter.

Contents

1.1 What is directional data? 1
1.2 Contributions of the thesis 5
1.3 Real datasets 8
1.3.1 Wind direction 9
1.3.2 Hipparcos dataset 10
1.3.3 Portuguese wildfires 11
1.3.4 Protein angles 12
1.3.5 Text mining 13
1.4 Manuscript distribution 15
References 16

1.1 What is directional data?

The term *directional data* was coined in the first book of Mardia (1972) to refer to data whose support is a circumference, a sphere or, generally, an hypersphere of arbitrary dimension. This kind of data appears naturally in several applied disciplines: proteomics (angles in the structure of proteins; see for example Hamelryck et al. (2012)); environmental sciences (wind direction (Johnson and Wehrly, 1978), direction of waves (Jona-Lasinio et al., 2012)); biology (animal orientation, see Batschelet (1981) for several examples); cyclic phenomena (arrival times at a care unit (Fisher, 1993, page 239), seasonality in freezing and thawing (Oliveira et al., 2013)); astronomy (position of stars, see Sections 1.2.8 and 1.5.3 of Perryman (1997)); image analysis (Dryden, 2005) or even in text mining (analysis of word frequency in texts, see for example Banerjee et al. (2005)). The collection of statistical techniques intended to analyse directional

data was named as *directional statistics* by the homonym book of Mardia and Jupp (2000), the revised reprint of Mardia (1972).

Directional observations are represented as points in the Euclidean hypersphere of dimension q, denoted by $\Omega_q = \{\mathbf{x} \in \mathbb{R}^{q+1} : ||\mathbf{x}|| = 1\}$ (also referred to as $\mathbb{S}^q$), where the simplest cases correspond to the circumference ($q = 1$) and the sphere ($q = 2$). Inference with directional data is indeed *constrained inference*, since all the methods used for statistical analysis should take into account the special nature of Ω_q, something that is not required with the usual *linear* (*i.e.* Euclidean) data. A pedagogical example that illustrates this problem is the definition of an appropriate directional mean for the simplest situation: for two observations $\mathbf{X}_1$ and $\mathbf{X}_2$ in the circumference Ω_1. A first attempt could be to consider the Euclidean mean $\bar{\mathbf{X}} = \frac{\mathbf{X}_1+\mathbf{X}_2}{2}$, but then $\bar{\mathbf{X}}$ is not guaranteed to belong to Ω_1. Another possibility is to consider polar coordinates (see Figure 1.1), compute the usual mean of the corresponding angles $\theta_1, \theta_2 \in [0, 2\pi)$ and then set the directional mean as the point $(\cos\bar{\theta}, \sin\bar{\theta})$, where $\bar{\theta} = \frac{\theta_1+\theta_2}{2}$. The problem with this approach is that if, for example, $\theta_1 = \frac{\pi}{4}$ and $\theta_2 = \frac{7\pi}{4}$, then $\bar{\theta} = \pi$, which produces an output in the opposite direction of the obvious mean, corresponding this one to $\bar{\theta} = 0$. A reasonable definition for a directional mean is obtained by $\frac{\bar{\mathbf{X}}}{||\bar{\mathbf{X}}||}$ (if $\bar{\mathbf{X}} \neq 0$, not defined otherwise), see Mardia and Jupp (2000) for further details.

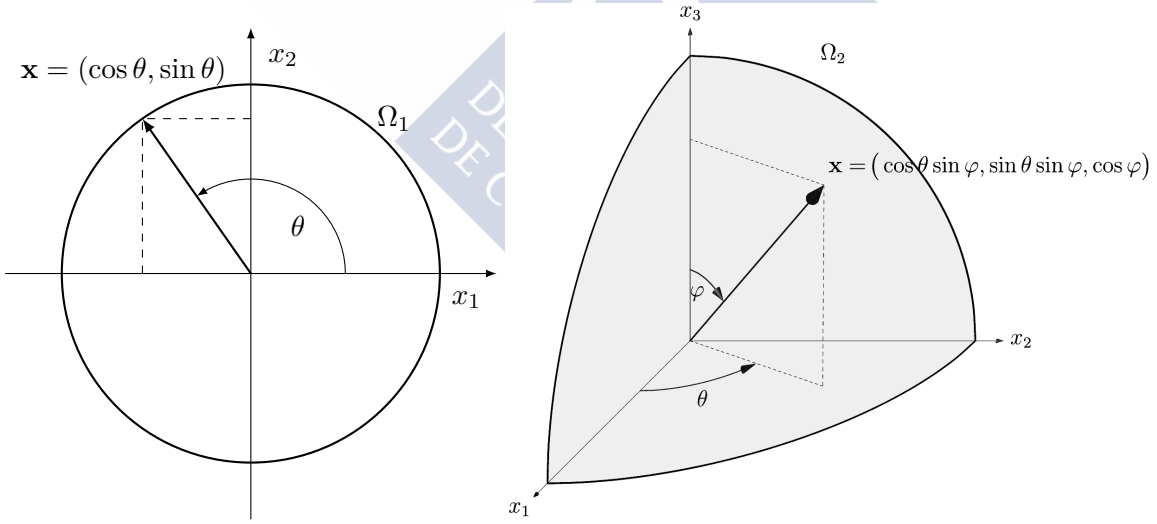


Figure 1.1: Polar coordinates in Ω_1 and spherical coordinates in Ω_2 (just an octant of the sphere is represented for a better display).

Two main approaches have been followed in the statistical literature for the analysis of directional data, differing in the kind of representation. The first one is based on polar and spherical coordinates (see Figure 1.1) to develop methods which are specifically designed to treat circular and spherical data, respectively, the most common types of directional data in practise. This is the approach followed by the books of Fisher (1993), Jammalamadaka and SenGupta (2001) and Pewsey et al. (2013) for circular data and of Fisher et al. (1993) for spherical. Unfortunately, the extensions of these methods to an arbitrary dimension q are not straightforward due to the nature of the spherical coordinates for higher dimensions. The second approach relies only in the Cartesian coordinates of the points in Ω_q, without assuming any particular dimension and

thus ensuring more generality. This is the approach followed in the thesis, except in Chapter 2, where the first one is employed.

Perhaps the most popular directional distribution is the *von Mises-Fisher* density (see Watson (1983) and Mardia and Jupp (2000)), or shortly the von Mises. The von Mises density, denoted by $\text{vM}(\boldsymbol{\mu}, \kappa)$ (or by $\text{vM}(\mu, \kappa)$, if $q = 1$ and $\boldsymbol{\mu} = (\cos\mu, \sin\mu)$), is given by

$$f_{\text{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = C_q(\kappa) \exp\left\{\kappa \mathbf{x}^T \boldsymbol{\mu}\right\}, \quad C_q(\kappa) = \frac{\kappa^{\frac{q-1}{2}}}{(2\pi)^{\frac{q+1}{2}} \mathcal{I}_{\frac{q-1}{2}}(\kappa)},$$

being $\boldsymbol{\mu} \in \Omega_q$ the directional mean, $\kappa \geq 0$ the concentration parameter around the mean ($\kappa = 0$ gives the uniform density on Ω_q) and $\mathcal{I}_\nu$ the modified Bessel function of the first kind and order ν, which can be written as (see equation 10.32.2 of Olver et al. (2010))

$$\mathcal{I}_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\pi^{1/2}\Gamma\left(\nu + \frac{1}{2}\right)} \int_{-1}^{1} (1 - t^2)^{\nu - \frac{1}{2}} e^{zt}\, dt.$$

This distribution is considered as the Gaussian analogue for directional data for two main reasons. First, it presents the same Maximum Likelihood Estimator (MLE) characterization property that the Gaussian distribution has in the Euclidean case: it is the only directional distribution whose MLE of the location parameter is the directional sample mean (see Bingham and Mardia (1975) for a proof). Second, the von Mises density can be obtained from a random vector normally distributed conditioned to have unit norm. That is, for a normal random vector

$$\mathbf{X} \sim \mathcal{N}_{q+1}\left(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_{q+1}\right), \text{ with } \boldsymbol{\mu} \in \mathbb{R}^{q+1} \backslash \{\mathbf{0}\} \text{ and } \sigma^2 > 0,$$

it happens that

$$\mathbf{Y} = (\mathbf{X} \,|\, ||\mathbf{X}|| = 1) \sim \text{vM}\left(\frac{\boldsymbol{\mu}}{||\boldsymbol{\mu}||}, \frac{||\boldsymbol{\mu}||}{\sigma^2}\right).$$

This result shows that the inverse of the concentration parameter κ of a von Mises can be identified with the variance of a multivariate normal with covariance matrix proportional to the identity. See Gatto (2011) for a proof of this result in a more general situation.

Another remarkable directional distribution is the one given by Jones and Pewsey (2005), which is denoted by $\text{JP}(\boldsymbol{\mu}, \kappa, \psi)$. Originally motivated for the circular case, its density can be also defined in Ω_q for an arbitrary dimension q:

$$f_{\text{JP}}(\mathbf{x}; \boldsymbol{\mu}, \kappa, \psi) = \frac{|\sinh(\kappa\psi)|^{\frac{q-1}{2}} \left(\cosh(\kappa\psi) + \sinh(\kappa\psi)\mathbf{x}^T\boldsymbol{\mu}\right)^{\frac{1}{\psi}}}{2^{\frac{q-1}{2}}\Gamma\left(\frac{q+1}{2}\right) P^{-\frac{q-1}{2}}_{\frac{1}{\psi}+\frac{q-1}{2}}(\cosh(\kappa\psi))},$$

where $\boldsymbol{\mu} \in \Omega_q$ is the location parameter, $\kappa \geq 0$ is the concentration around $\boldsymbol{\mu}$, $\psi \in \mathbb{R}$ is a shape parameter that controls a kind of negative kurtosis with respect to a $\text{vM}(\boldsymbol{\mu}, \kappa)$ ($\psi < 0$ stands for more peaked densities while $\psi > 0$ produces flatter ones) and P^μ_ν is the Legendre function of the first kind, order μ and degree ν (see equation 14.12.4 of Olver et al. (2010)). This parametric family has the interesting property of containing as particular cases the $\text{vM}(\boldsymbol{\mu}, \kappa)$ (corresponding to $\psi \to 0$) and, with $q = 1$ and taking polar coordinates, the Cardioid ($\psi = 1$), the Wrapped

Cauchy ($\psi = -1$) and the Cartwright's power-of-cosine ($\psi < 0$, $\kappa \to \infty$). See Section 2 of Jammalamadaka and SenGupta (2001) for more information on these distributions.

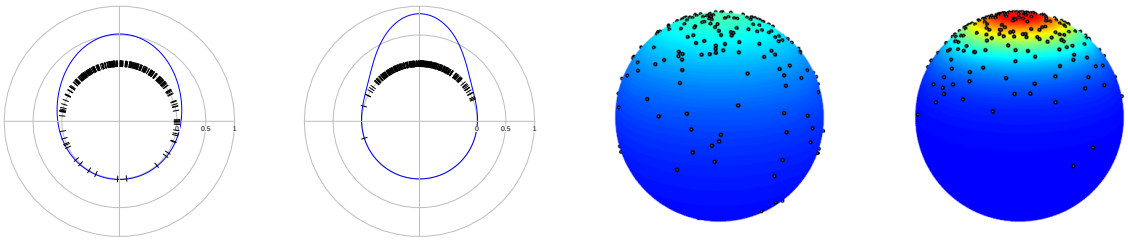


Figure 1.2: Von Mises densities on Ω_1 and Ω_2. From right to left, densities corresponding to $q = 1$ with κ_1 and κ_2 and to $q = 2$ with κ_1 and κ_2, respectively. The parameters are $\boldsymbol{\mu} = (\mathbf{0}_q, 1)$, $\kappa_1 = 2$ and $\kappa_2 = 5$, with $\mathbf{0}_q$ denoting a vector of q zeros. Samples of size $n = 250$ are drawn.

In many situations, directional random variables appear together with a linear or another directional variable, being the circular-linear (support on the cylinder $\Omega_1 \times \mathbb{R}$) and circular-circular (support on the torus $\Omega_1 \times \Omega_1$) data the most common situations. See Figure 1.3 for some examples of densities in these supports. For both cases, a semiparametric model for the joint density was proposed in Johnson and Wehrly (1978) and Wehrly and Johnson (1979):

$$f_{\Theta,X}(\theta, x) = 2\pi g\left(2\pi(F_\Theta(\theta) \pm F_X(x))\right) \times f_\Theta(\theta) f_X(x), \tag{1.1}$$

being Θ a circular variable with density f_Θ and distribution function F_Θ and X either a linear or a circular variable with associated f_X and F_X. g is a circular density that acts as the link function between the two marginal densities, given by f_Θ and f_X, either considering a positive (negative dependency) or a negative sign (positive dependency) in $\pm$. g can be interpreted in terms of copulas (see Nelsen (2006) for an introduction to copulas), since the copula density of (Θ, X) is given by $c_{\Theta,X}(u, v) = 2\pi g\left(2\pi(u \pm v)\right)$. Different parametric models can be obtained from this semiparametric structure, for example the bivariate von Mises model of Shieh and Johnson (2005).

However, not all parametric densities in this context satisfy (1.1). For example, the circular-linear density of Mardia and Sutton (1978) (denoted by $\text{MS}(\mu, \kappa, m, \rho_1, \rho_2, \sigma)$) or the circular-circular Wrapped Normal Torus given in Example 7.3 of Johnson and Wehrly (1977) (denoted by $\text{WNT}(m_1, m_2, \sigma_1, \sigma_2, \rho)$) are parametric densities that do not verify (1.1). The expressions of these densities are, respectively:

$$f_{\text{MS}}(\theta, x; \mu, \kappa, m, \sigma, \rho_1, \rho_2) = f_{\text{vM}}(\theta; \mu, \kappa) \times f_{\mathcal{N}}(z; m(\theta; \mu, \kappa, m, \sigma, \rho_1, \rho_2), \sigma(1 - \rho_1 - \rho_2)),$$

$$f_{\text{WNT}}(\theta, \psi; m_1, m_2, \sigma_1, \sigma_2, \rho) = \sum_{p_1=-\infty}^{\infty} \sum_{p_2=-\infty}^{\infty} f_{\mathcal{N}}(\theta + 2\pi p_1, \psi + 2\pi p_2; m_1, m_2, \sigma_1, \sigma_2, \rho),$$

where $m(\theta; \mu, \kappa, m, \sigma, \rho_1, \rho_2) = m + \sigma\kappa^{\frac{1}{2}}\{\rho_1(\cos(\theta) - \cos(\mu)) + \rho_2(\sin(\theta) - \sin(\mu))\}$, $f_{\mathcal{N}}(\cdot; m, \sigma)$ stands for the density of a $\mathcal{N}(m, \sigma^2)$ and $f_{\mathcal{N}}(\cdot, \cdot; m_1, m_2, \sigma_1, \sigma_2, \rho)$ represents the density of a bivariate normal with mean vector $(m_1, m_2)^T$, marginal variances σ_1^2 and σ_2^2 and correlation coefficient ρ. Among others, these two distributions will be employed along the different chapters. For example, the last two ones appear in Chapter 6 and the relation (1.1) is crucial for Chapter 2.

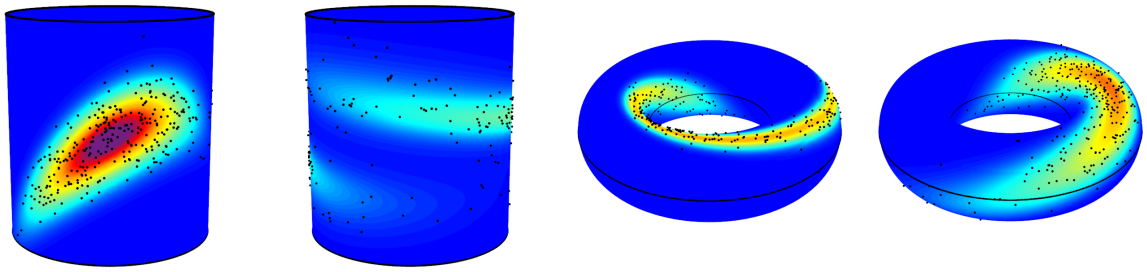


Figure 1.3: From right to left, densities corresponding to: MS $\left(\frac{3\pi}{2}, 5, 0, \frac{1}{2}, -\frac{3}{4}, \frac{3}{2}\right)$; density (1.1) with a vM $\left(\frac{5\pi}{4}, \frac{3}{2}\right)$ link and marginals $\mathcal{N}(0,1)$ and circular uniform; density (1.1) with a vM$(\pi, 7)$ link and marginals JP $\left(0, \frac{1}{2}, 1\right)$ and uniform circular; WNT $\left(0, \frac{\pi}{6}, \frac{3}{2}, \frac{1}{4}, 0\right)$. Samples of size $n = 300$ are drawn.

Finally, it should be noted that directional data also arise related to or as particular cases of more general spaces, as it happens for example in *statistical shape analysis* (see Dryden and Mardia (1998) and Kendall et al. (1999) for a review on the topic) or when considering *statistics on Riemannian manifolds* (see Bhattacharya and Bhattacharya (2012) and references therein).

1.2 Contributions of the thesis

Parametric methods have played a predominant role in the development of statistical inference for the analysis of directional data (see Mardia (1972) and Watson (1983)). Later publications, such as Fisher (1993), Fisher et al. (1993), Mardia and Jupp (2000), Jammalamadaka and SenGupta (2001) and Pewsey et al. (2013) also devoted much of their attention to the use of parametric techniques. These methods rely on the assumption that a certain parametric hypothesis in the stochastic generating process holds. For example, inference on the unknown density of a directional random variable is usually done by assuming a certain density model, up to the determination of some unknown parameters which are estimated from the data. Whereas this procedure leads to optimal results (in terms of efficiency) if the parametric assumption holds, the estimation can be totally misleading if the assumption fails.

On the other hand, nonparametric methods do not rely on strong parametric assumptions on the stochastic generating process, except for some mild smoothness conditions. The main advantage is that nonparametric methods always provide reasonable solutions for inference in general, no matter if a parametric assumption holds or not. Obviously, a nonparametric method is not optimal compared with a parametric competitor designed *ad hoc* for a parametric scenario, but still very useful. For example, the comparison of a parametric and a nonparametric fit leads to a so called *goodness-of-fit test*, which formally checks if the parametric hypothesis is plausible given the sample information.

The aim of this thesis is to provide new methodological tools for nonparametric inference with directional and linear data. Specifically, nonparametric methods are obtained for both estimation and testing, for the density and the regression curves, in situations where directional random variables are present, that is, directional, directional-linear and directional-directional random variables. In what follows, short states of the art on these topics are given jointly with the contributions of the thesis, referring to the papers providing Chapters 2–7, the main core of the manuscript. See also Figure 1.4 for a diagram with the main references and contributions.

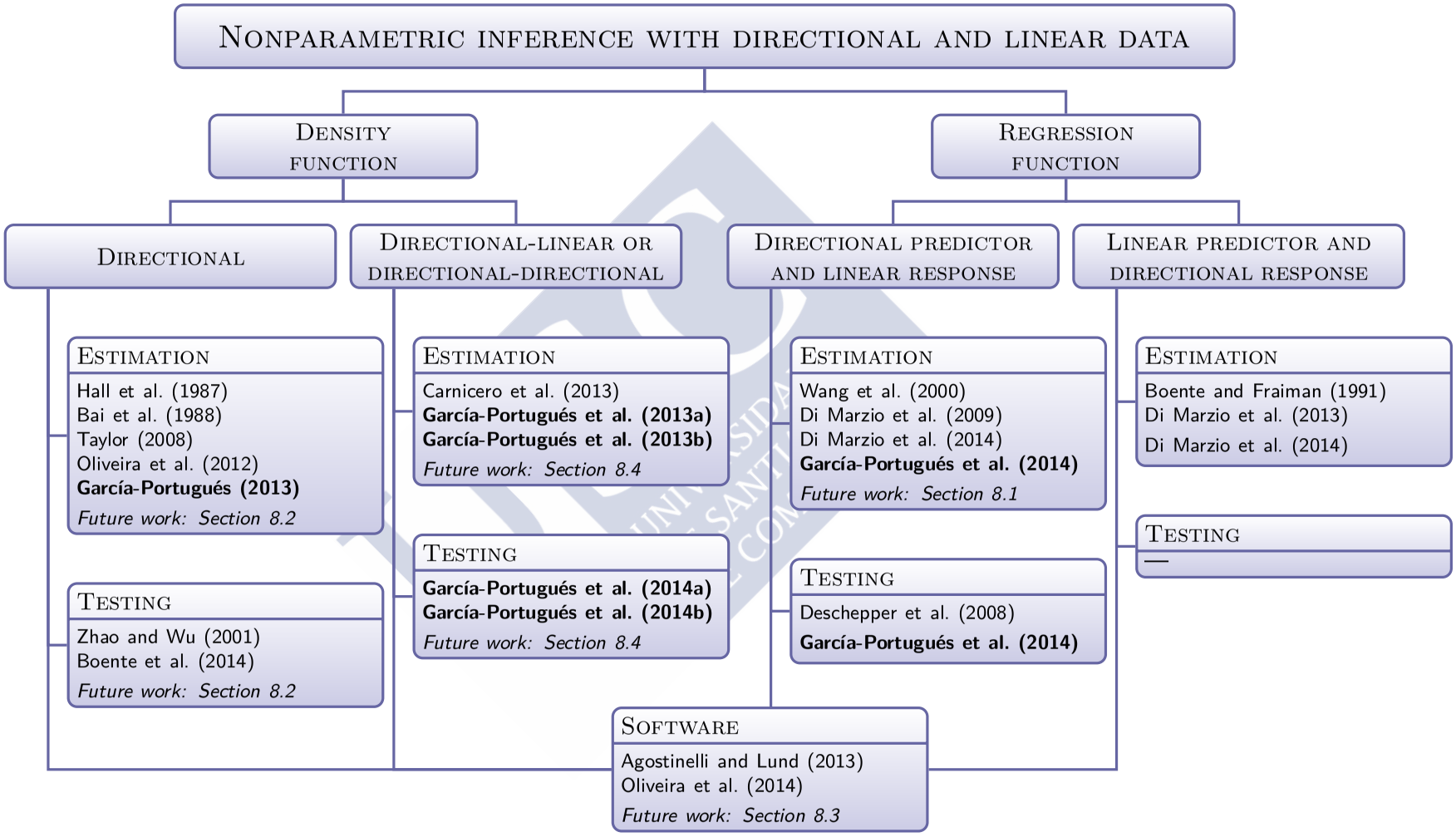


Figure 1.4: Diagram with the contributions of the thesis (marked in bold) and a short state of the art with the main references in the scope of the thesis. Future research lines described in Chapter 8 are included in italics.

A. Density function. Consider $\mathbf{X}$ and $\mathbf{Y}$ two directional variables and Z a linear one. Let $f_{\mathbf{X}}$, $f_{\mathbf{X},Z}$ and $f_{\mathbf{X},\mathbf{Y}}$ be their directional, directional-linear and directional-directional density functions, respectively.

A.1. **Estimation**. The objective is the estimation of $f_{\mathbf{X}}$, $f_{\mathbf{X},Z}$ and $f_{\mathbf{X},\mathbf{Y}}$ by kernel smoothing methods. The copula density is also estimated for circular-linear and circular-circular cases.

- **State of the art**. Kernel density estimation for $f_{\mathbf{X}}$ was firstly considered by Hall et al. (1987) and Bai et al. (1988), who established its basic asymptotic properties, and was later studied by Klemelä (2000). Taylor (2008) and Oliveira et al. (2012) proposed bandwidth selection rules for the circular case, the latter based in the asymptotic error expression stated in Di Marzio et al. (2009). For the estimation of the densities $f_{\mathbf{X},Z}$ and $f_{\mathbf{X},\mathbf{Y}}$ with circular variables, Fernández-Durán (2007) introduced a parametric method based on the copula structure given by Johnson and Wehrly (1978). A fully nonparametric estimation of the copula density employing Bernstein polynomials was given in Carnicero et al. (2013).
- **Contributions**. A new nonparametric method for estimating circular-linear and circular-circular densities from the estimation of the copula structure of Johnson and Wehrly (1978) is presented in García-Portugués et al. (2013a). The method considers, among other approaches, a modification of the kernel density estimator of Gijbels and Mielniczuk (1990). Since this procedure is hardly extensible to higher dimensions, in García-Portugués et al. (2013b) a new kernel density estimator for $f_{\mathbf{X},Z}$ is given, which avoids the estimation via copulas. Exact error expressions are derived for the density estimators of $f_{\mathbf{X},Z}$ and $f_{\mathbf{X},\mathbf{Y}}$. These expressions are used in García-Portugués (2013) to set up new bandwidth selection rules for the kernel density estimator of $f_{\mathbf{X}}$.

A.2. **Testing**. The two goals are: *i*) test if $\mathbf{X}$ and Z are independent, *i.e.*, test $H_0 : f_{\mathbf{X},Z}(\cdot,\cdot) = f_{\mathbf{X}}(\cdot)f_Z(\cdot)$ holds; *ii*) test if $f_{\mathbf{X},Z}$ has a particular parametric form, *i.e.*, if $H_0 : f_{\mathbf{X},Z} \in \{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ holds. Similarly with $f_{\mathbf{X},\mathbf{Y}}$ instead of $f_{\mathbf{X},Z}$.

- **State of the art**. Up to the author's knowledge, the only goodness-of-fit test for parametric directional densities was proposed by Boente et al. (2014). It is based on the Central Limit Theorem (CLT) of Zhao and Wu (2001) for the Integrated Squared Error (ISE) of the kernel density estimator for $f_{\mathbf{X}}$. These papers can be regarded as directional analogues of Fan (1994) (or Bickel and Rosenblatt (1973)) and Hall (1984), respectively. There exist also correlation based tests for detecting circular-linear association, such as the ones given by Mardia (1976), Johnson and Wehrly (1977) and Fisher and Lee (1981).
- **Contributions**. A test for assessing the independence between a directional and a linear random variable (also adaptable to the directional-directional case) is given in García-Portugués et al. (2014a). The test statistic can be seen as an analogue of Rosenblatt and Wahlen (1992) (or Rosenblatt (1975)), since it considers the squared distance between the estimator of $f_{\mathbf{X},Z}$ and the product of the estimators of $f_{\mathbf{X}}$ and f_Z. The CLT for the ISE of the directional-linear and directional-directional estimator is obtained in García-Portugués et al. (2014b). This serves as a keystone to derive the asymptotic distribution of the independence test and

also goodness-of-fit tests for directional-linear and directional-directional parametric densities. The consistency of a bootstrap resampling method for calibration is also proved.

B. **Regression function**. Let $\mathbf{X}$ be a directional variable and Y and ε two linear variables. It is assumed that a regression model for Y over $\mathbf{X}$ holds, that is $Y = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$, with $m(\cdot) = \mathbb{E}[Y|\mathbf{X} = \cdot]$ and $\sigma^2(\cdot) = \mathbb{V}\mathrm{ar}[Y|\mathbf{X} = \cdot]$.

B.1. **Estimation**. The objective is the estimation of m by kernel smoothing methods using a local linear estimator.

- **State of the art**. An adaptation of the Nadaraya-Watson estimator for the regression function m was given by Wang et al. (2000) who derived its law of the iterated logarithm. In Di Marzio et al. (2009) a local polynomial estimator for m, when the predictor is circular, is presented. This approach was lately considered in Di Marzio et al. (2013) for regression with circular response. Di Marzio et al. (2014) proposed a different local linear estimator with either directional predictor or response based on Taylor expansions constructed with the tangent normal decomposition. An earlier definition of a local estimator with directional response and linear predictor was given in Boente and Fraiman (1991).
- **Contributions**. A projected local linear estimator for the regression function m is considered in García-Portugués et al. (2014). The estimator is motivated by a modified Taylor expansion designed to avoid the overparametrization that appears when considering the classical local linear estimator. Asymptotic bias, variance and normality of the estimator are provided, as well as the equivalent kernel formulation. Particular cases of the estimator include the one from Wang et al. (2000) and the local linear estimator with circular predictor of Di Marzio et al. (2009).

B.2. **Testing**. The goal is to test if m belongs to a class of parametric regression functions, *i.e.*, if the hypothesis $H_0 : m \in \{m_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ holds.

- **State of the art**. Deschepper et al. (2008) provided a test for the significance of a linear response on a circular predictor, which is, up to the author's knowledge, the only nonparametric test in the regression setting with directional variables. A resampling mechanism for the calibration of the test statistic was also proposed in the mentioned reference.
- **Contributions**. A goodness-of-fit test for parametric regression models with directional predictor and linear response is presented in García-Portugués et al. (2014). The projected local linear estimator is used to construct a test statistic that measures the squared distance between this nonparametric estimator and a smoothed version of the parametric one (similar to the one of Härdle and Mammen (1993)), using either local constant or linear fits. The asymptotic distribution of the test statistic and its power against local alternatives are addressed, together with a consistent resampling procedure.

1.3 Real datasets

Along the thesis, different data examples have been considered to motivate and illustrate the new methodologies. In this section an exhaustive description of the real datasets, most of them

original (except for the protein angles and part of the Portuguese wildfires) is provided. For a collection of classical datasets on directional data see Fisher (1993), Fisher et al. (1993) and Mardia and Jupp (2000).

1.3.1 Wind direction

The coal power plant of As Pontes (7° 51' 45" W, 43° 26' 26" N), located in the northwest of Spain, has one of the highest electricity power generation capacity among the power plants in the country. The power plant is able to generate up to 2200 megawatts, but unfortunately at the expense of a considerable emission of pollutants. Due to the high concentration of pollutants and the serious consequences in the environment that acid rain causes (produced by high concentrations of sulphur dioxide), a series of precautionary measures designed to reduce the emissions of the power plant were applied since 2005. The measurement of different pollutants' concentration, including sulphur dioxide (SO_2), is controlled by means of a network of monitoring stations located around the plant, that also measure meteorological variables of interest. Among these variables, the direction in which the wind blows is recorded, since it plays a predominant role on the dissemination of particles in the atmosphere.

The data application in Chapter 2 is focused on the relation between SO_2 concentration and wind direction in a monitoring station located at the northeast of the power plant (B2 station, 7° 44' 10" W, 43° 32' 05" N). The aim is to check if wind blowing from the power plant carries higher concentrations of SO_2 and the effectiveness of the implemented precautionary measures. For this purpose, two datasets were obtained for the months of January 2004 and 2011 from minutely recordings at station B2. After that, the following steps were performed: 1) not available observations were omitted; 2) data were hourly averaged in order to mitigate serial dependence; 3) a slight perturbation was applied to avoid serial repeated data arising from limitations of the measuring devices; 4) the SO_2 sample was transformed using a Box-Cox transformation to mitigate its skewness. This procedure results in a pair of datasets with 736 and 743 observations for 2004 and 2011, respectively.

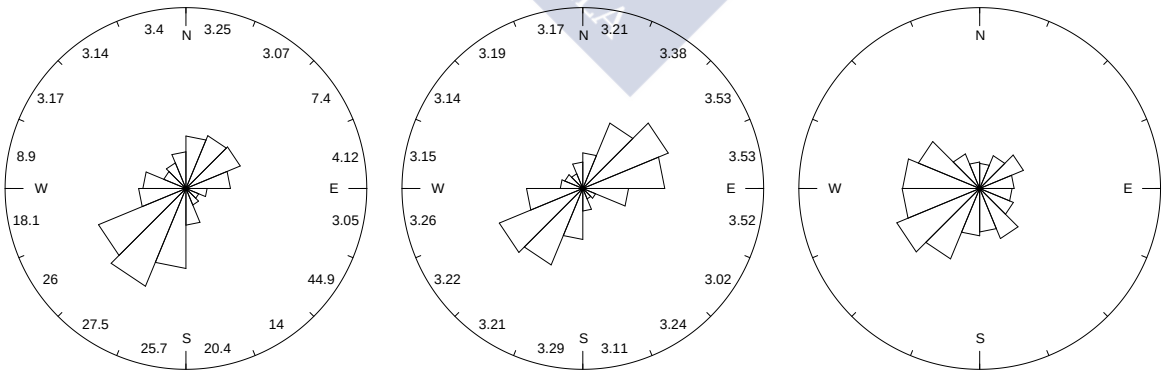


Figure 1.5: Rose diagrams for wind direction in station B2 for January 2004 (left), January 2011 (center) and in station A mourela in June 2012 (right). The first two ones show the average SO_2 concentration in $\mu g/m^3$ associated with winds coming from each partition of the windrose.

On the other hand, the data application in Chapter 4 includes the study of the wind direction in a station closer to the power plant (A Mourela station, 7° 51' 21.91" W, 43° 25' 52.35" N), in order to determine the directions where the pollutants are more likely to be spread. The

data acquisition was performed as described previously but recording only the wind direction, omitting steps 3) and 4) and considering measurements from June 2012.

Figure 1.5 shows three rose diagrams (circular histograms) summarizing the information of the datasets. The first two ones are from the station B2 before and after the precautionary measures. While in 2004 there is a high SO_2 concentration associated to the south and southwest winds (coming from the power plant), this is not the case for 2011, where average SO_2 concentrations are roughly constant.

1.3.2 Hipparcos dataset

The Hipparcos space astrometry mission was a carried out by the European Space Agency between 1989–1993 in order to pinpoint the position of more than one hundred thousand stars. The massive enumeration of stars obtained from the mission were collected in the Hipparcos catalogue (Perryman, 1997) and made openly accessible. A decade later, a new revision of the raw data was carried out by van Leeuwen (2007). During this period, the advances on the measuring techniques allowed to determine with higher precision the exact position of the satellite during the mission. As a consequence, this revised version of the dataset presents a significant improvement in the overall reliability of the astrometric catalogue. This is the dataset that has been considered in the application in Chapter 4 and can be downloaded from the VizieR catalogue service (Ochsenbein et al., 2000). The number of stars in the dataset is $n = 117955$.

Since stars are objects that are continuously moving, the measurements in the Hipparcos catalogue were done with respect to their positions in a common reference date. This concept is known as *epoch* (Ep) in astronomy and it was fixed to the median year with respect to the duration of the mission, Ep 1991.5. The position of stars is referred to the position that occupy in the celestial sphere, *i.e.*, the location in the earth surface that arises as the intersection with the imaginary line that joins the centre of the idealized earth (perfectly spherical) with the star. This is parametrized by a couple of angles (λ, β), $\lambda \in [-\pi, \pi)$, $\beta \in [0, \pi)$, so that

$$\begin{cases} x_1 = \cos\beta\cos\lambda, \\ x_2 = \cos\beta\sin\lambda, \\ x_3 = \sin\beta. \end{cases}$$

The centre of the celestial sphere is placed at the centre of the Milky Way and its equator corresponds to the *galactic plane*, this is, the rotation plane of the galaxy. Point representation in this coordinate system is known as *galactic coordinates* and is very popular in astronomy due to its easy interpretation.

The usual way of representing spherical surfaces in the Hipparcos catalogue (and in astronomy in general) is the Aitoff projection. This transformation projects the sphere surface inside an ellipse with major semi-axis twice the minor semi-axis, whose longitude is R. The point (x, y) inside the ellipse is given by

$$\begin{cases} x = -2R\cos\beta\sin(\lambda/2)\,/\sqrt{1+\cos\beta\cos(\lambda/2)}, \\ y = R\sin\beta/\sqrt{1+\cos\beta\cos(\lambda/2)}. \end{cases}$$

This projection does not conserve distances but it does preserve the area (the proportions between areas of regions in the sphere and areas of the projected regions remain constant).

Figure 1.6 shows the Aitoff projection of the histogram of the stars counts as presented in the Hipparcos catalogue. A smoothed version of this figure can be obtained by replacing the histogram with a kernel density estimator for spherical data that takes into account the continuous nature of the observations. This is done in Chapter 4 employing a suitable bandwidth selector in the kernel density estimator.

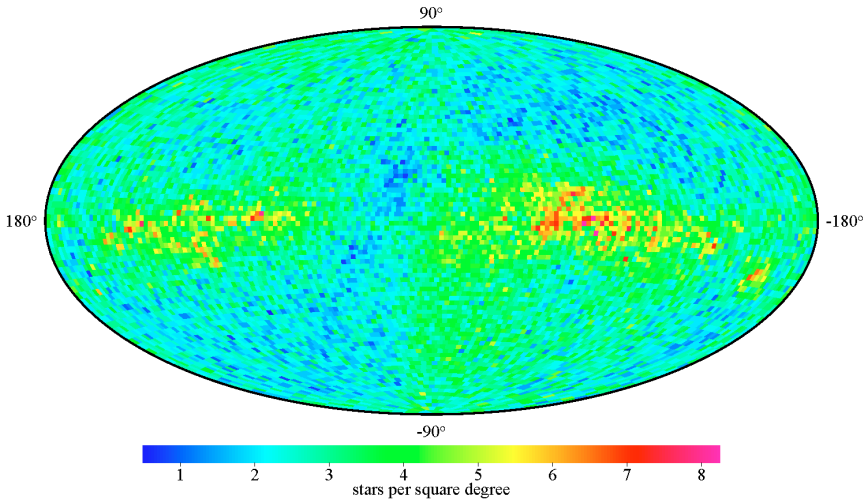


Figure 1.6: Number of observed stars per square degree, in galactic coordinates (cell size $2° \times 2°$), extracted from Figure 3.2.1 of Perryman (1997). The higher concentrations of stars are located around the equator (galactic plane) and two spots that represent the Orion's arm (left) and the Gould's Belt (right).

1.3.3 Portuguese wildfires

Chapters 5 and 6 analyse directional data arising from the main orientations of wildfires occurred in Portugal from 1985 to 2005. This massive data collection contains the $n = 26870$ fire perimeters (see right plot of Figure 1.7) together with their (log) burnt areas and was acquired from the imagery of the Landsat satellites. Imagery covering the mainland of Portugal was obtained before and after the fire season, providing a snapshot of the fires that occurred during the season. Annual fire perimeters were derived through a semi-automatic procedure that starts with supervised image classification and is followed by manual editing (Barros et al., 2012). The Minimum Mapping Unit (MMU) of the satellite is 5 hectares, which, although is not able to capture the smallest fires, allows to map up to the 90% of total burnt area.

Watersheds play an important role on the study of wildfires and their orientation. In Barros et al. (2012) the authors delimited 102 watersheds (see left plot of Figure 1.7) in which the wildfires are grouped, studying which of them showed a preferential alignment with the fires orientation. The orientation of the different object perimeters (either watersheds or wildfires) is determined by the first principal component (PC1) obtained from the points that constitute the object's boundary, either in bidimensional space defined by each vertex's latitude and longitude coordinates, or in tridimensional space, taking also into account the altitude. Then, the PC1 corresponds to an axis that passes through the object mass centre and that maximizes the variance of the projected vertices, represented in $\mathbb{R}^2$ or in $\mathbb{R}^3$ (see right plot of Figure 1.7).

In the two-dimensional case, the orientation is an *axial* object (the orientation N/S is also S/N). These orientations can be encoded by an angular variable $\theta \in [0, \pi)$ with period π, so 2θ is a usual circular variable. With this codification, the angles 0, $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$ represent the E/W, NE/SW, N/S and NW/SE orientations, respectively. In the three-dimensional space, the orientation is coded by a pair of angles (θ, φ) using spherical coordinates (see Figure 1.1), where $\theta \in [0, \pi)$ plays the same role as the previous case and $\varphi \in [0, \frac{\pi}{2}]$ measures the inclination ($\Phi = \frac{\pi}{2}$ for flat slope and $\varphi = 0$ for vertical; only positive angles are considered since negative ones lead to the same inclination). Therefore, points with spherical coordinates $(2\theta, \varphi)$, which lie on the upper semisphere, can be considered as a realization of a spherical variable.

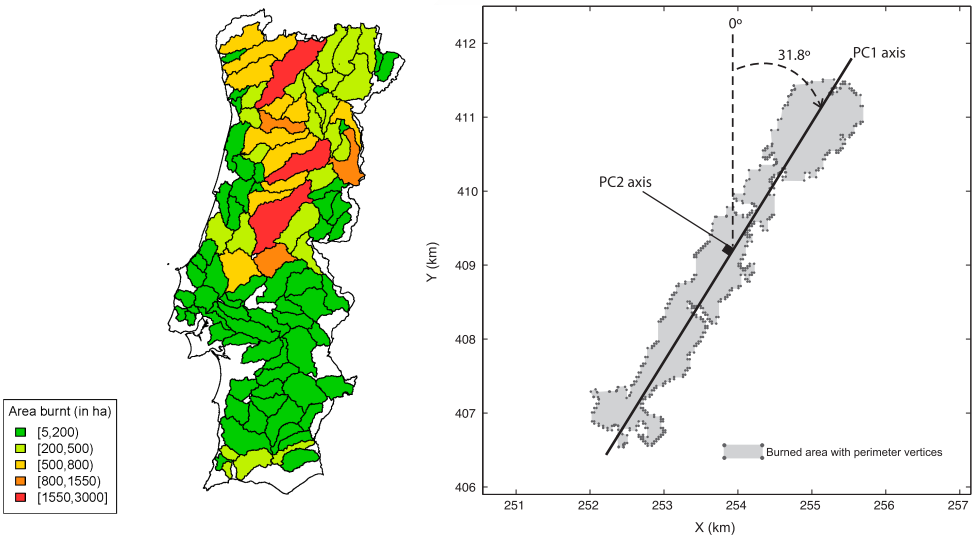


Figure 1.7: Number of hectares burnt in 1985–2005 in each watershed delineated by Barros et al. (2012) (left plot) and main orientation of a wildfire obtained from the PC1 of the perimeter (right plot; extracted from Barros et al. (2012)).

In Chapter 5 the independence between circular or spherical orientations of the different objects and burnt areas is tested with the whole dataset. Chapter 6 analyses a reduced dataset obtained by considering the average wildfire circular orientation in each of the watersheds and the mean of the burnt area, yielding a dataset of size $n = 102$. A goodness-of-fit test is applied for a parametric circular-linear model to assess its suitability for explaining the dataset.

1.3.4 Protein angles

A scientific field where directional statistics is called to play an important role is proteomics. Biomolecular structures like proteins are often expressed in terms of the dihedral angles that describe the rotations of the backbone around the bonds between atoms N-C_α (angle ϕ) and C_α-C (angle Ψ). The scatterplot of these pairs of angles in a protein, known as the *Ramachadran plot*, provides an easy way to view the allowed torsion combinations of the backbone. The distribution of the dihedral angles and its modelling is a key step in the study of the so-called *protein folding problem*, one of the main open problems in biology nowadays. See Hamelryck et al. (2012) and references therein for deeper insights on proteomics and directional methods used in the field.

The dataset analysed in Chapter 6 contains pairs of dihedral angles in the primary structure of 1932 proteins. This dataset was initially studied in Fernández-Durán (2007) by the use of the copula structure of Wehrly and Johnson (1979) and parametric models for the marginal densities and link function. The application in Chapter 6 concerns a goodness-of-fit test for such parametric models. The dataset is formed by 233 pairs of angles from segments of the type alanine-alanine-alanine in alanine amino acids, extracted from a representative sample of 1932 proteins retrieved from the July 2003 list of recommended proteins from the Protein Data Bank (Berman et al., 2000). The dataset is available in the `ProteinsAAA` object of the `R` package `CircNNTSR` (Fernández-Durán and Gregorio-Domínguez, 2013).

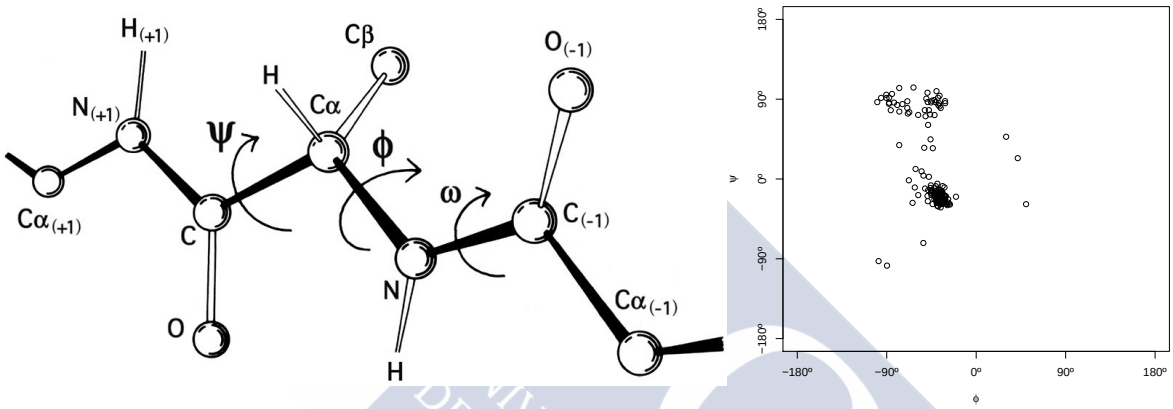


Figure 1.8: Representation of the dihedral angles in the primary protein structure (left, adapted from Richardson (2011)) and Ramachadran plot of the dihedral angles of the alanine-alanine-alanine segments contained in the `ProteinsAAA` object of Fernández-Durán and Gregorio-Domínguez (2013).

1.3.5 Text mining

Directional statistics has also interesting applications in high dimensional settings. A good example is text mining, where documents are usually represented as normalized vectors in Ω_{D-1} being D the number of words in a dictionary. This concept is the celebrated *vector space model*: a collection of documents (known as *corpus*) $\mathbf{d}_1, \ldots, \mathbf{d}_n$ is codified by the set of vectors $\{(d_{i1}, \ldots, d_{iD})\}_{i=1}^{n}$ (the *document-term matrix*) with respect to a dictionary $\{w_1, \ldots, w_D\}$, such that d_{ij} represent the frequency of the dictionary's j-th word in the document d_i. In this scenario a normalization is required to avoid unbalances between large and small documents. Consider for example the case where a document is produced by copying and pasting N times another document: both vectors will have the same direction but the length of the former will be N times larger, thus make them appear to be rather different elements in $\mathbb{R}^D$, although they share the same information. Taking the Euclidean norm then $\mathbf{d}_i / ||\mathbf{d}_i|| \in \Omega_{D-1}$ and hence the corpus can be regarded as a sample of directional data.

The data application in Chapter 7 concerns a corpus extracted from the news aggregator *Slashdot* (`wwww.slashdot.org`): a goodness-of-fit test is applied for a linear model to assess its suitability for explaining the popularity of the news from their content. The website publishes daily news/stories that are submitted and evaluated by users. The site is mainly focused on technology and science, but it also covers news related with politics or digital rights. Its main structure

can be seen in Figure 1.9, which essentially can be regarded as a series of news entries. Each of them includes a title, a short abstract of the news and information regarding the submission (user, date, keywords,...). Under the abstract a link shows the number of comments in the discussion section attached to the entry. Obviously, the degree of participation in the discussion section varies notably according to the news topic. For example, news related with politics and with proprietary software tend to be more controversial and generate more discussion, whereas news with high specific scientific content often have fewer comments.

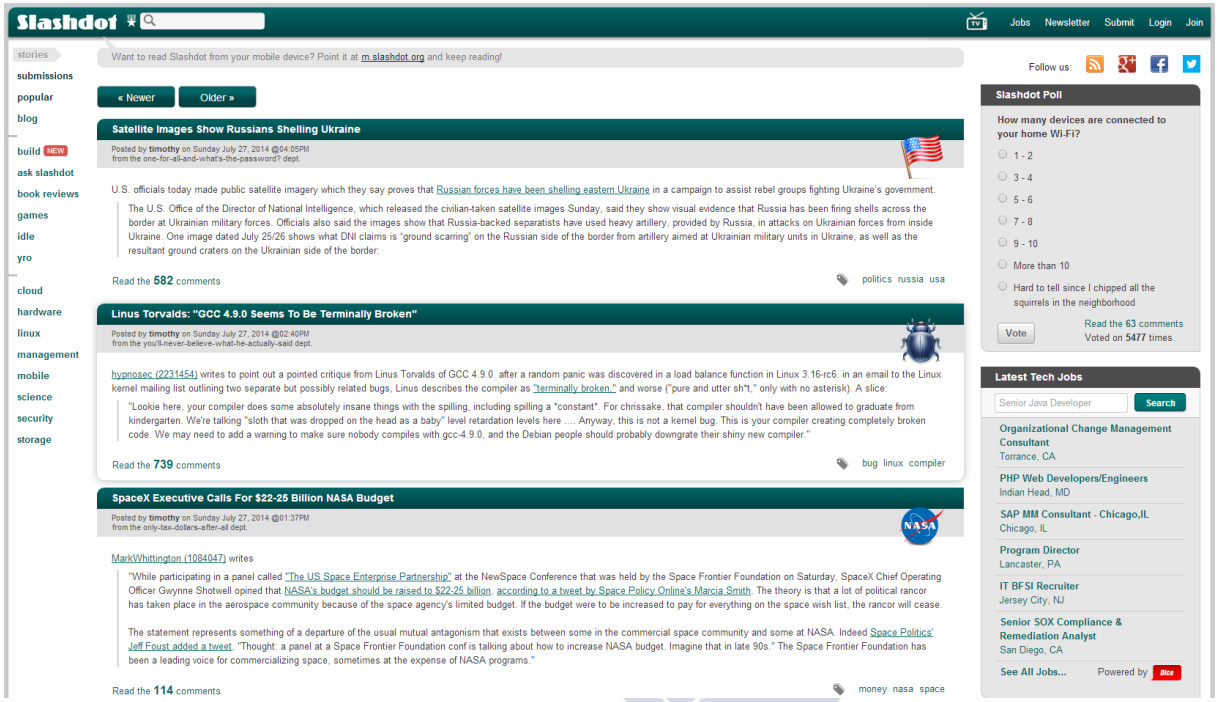


Figure 1.9: View of the *Slashdot* website on July 27, 2014.

The dataset was acquired as follows. First, titles, summaries and number of comments in all the news appeared in 2013 were downloaded by parsing automatically the news archive, obtaining a collection of $n = 8121$ documents. After that, the next steps were performed using the text mining `R` library `tm` (Meyer et al., 2008) and self-programmed code: 1) merge titles and summaries in the same document, omitting user submission details; 2) deletion of `HTML` codes; 3) conversion to lowercase; 4) deletion of stop words (retrieved from the stop words lists given in `tm` and `MySQL`), punctuation, white spaces and numbers; 5) stemming of words to reduce them to their root form; 6) pruning to remove words too rare or too frequent (more than the 50% of the processed words only appeared in a single document). The last step was performed by considering only the words that appear within 58 and 1096 documents, which were $D = 1508$. These quantities correspond to quantiles 95% and 99.95% of the *document frequency* (*i.e.*, the number of documents containing a particular word) empirical distribution. Finally, the corpus was stored as a normalized document-term matrix using the dictionary formed by the D words.

1.4 Manuscript distribution

The main core of the thesis are Chapters 2 to 7. Each of them contains an original article about a specific topic related with the main guiding theme: nonparametric inference with directional and linear data. Therefore, each chapter is presented as a self-contained paper with proper abstract, sections, appendices and references, just in the same way as it was published, accepted for publication or submitted. The reference of the paper that gives rise to the chapter is included in the front page. At the time of presenting this manuscript, the papers from Chapters 2–5 have been published, the paper from Chapter 6 has been accepted and the one from Chapter 7 has been submitted. Short summaries of all the chapters and their associated papers are given below.

Chapter 1: Introduction. An introduction to the field of directional statistics is presented in this first chapter. The introduction describes the state-of-the art and main references in the topics where the thesis presents new contributions. For a better understanding, an explanatory diagram of the contributions of this thesis is given. The chapter also describes the data sets used along the manuscript and the structure of the thesis.

Chapter 2: Density estimation by circular-linear copulas (García-Portugués et al., 2013a). This chapter presents different approaches to the estimation of a circular-linear or circular-circular density by the use of copulas and the dependence structure of Johnson and Wehrly (1978). A nonparametric procedure is used to analyse the relationship between the wind direction and the SO_2 concentration in a monitoring station near As Pontes coal power plant, in order to study the effectiveness of precautionary measures for pollutants reduction.

Chapter 3: Kernel density estimation for directional-linear data (García-Portugués et al., 2013b). A natural alternative to the nonparametric method given in the previous chapter is a kernel density estimator applied on the data directly, *i.e.* without requiring copula modelling. The kernel density estimator for directional-linear data is presented in this chapter, providing results for bias, variance and asymptotic normality. Exact error expressions are obtained for the directional-linear kernel density estimator but also for the directional one, setting the basis for Chapter 4.

Chapter 4: Bandwidth selectors for kernel density estimation with directional data (García-Portugués, 2013). Based on the asymptotic and exact error expressions given in Chapter 3, three new bandwidth selectors are proposed. The first selector is a natural analogue of the circular selector given in Taylor (2008) while the other two arise from combining mixtures of von Mises distributions with the asymptotic or exact error criteria. The performance of the proposed selectors are compared in an extensive simulation study and the best selector is illustrated with the datasets from wind directions and from the Hipparcos satellite.

Chapter 5: A nonparametric test for directional-linear independence (García-Portugués et al., 2014a). Using the estimator given in Chapter 3, a test based on the squared distance between the joint kernel estimator and the product of marginal directional and linear kernel density estimators is considered. A closed expression is given for the statistic and a resampling method based on permutations is employed to calibrate the test statistic. The test performance is analysed in a simulation study under a variety of situations and is applied to account for the influence of the orientation of the wildfires on their size in the Portuguese wildfires dataset.

Chapter 6: Central limit theorems for directional and linear random variables with applications (García-Portugués et al., 2014b). This chapter is devoted to derive a CLT for the ISE of the kernel density estimator in Chapter 3. The result is used to establish the convergence in distribution of the independence test given in Chapter 5 and to analyse a new goodness-of-fit test for parametric families of directional-linear densities. A consistent bootstrap strategy is given for the goodness-of-fit test and illustrated in an extensive simulation study. The goodness-of-fit test is applied to the protein angles and Portuguese wildfires datasets.

Chapter 7: Testing parametric models in linear-directional regression (García-Portugués et al., 2014). A new local linear estimator is proposed for the estimation of the regression function with directional predictor and linear response, establishing its different properties. Based on this estimator, a goodness-of-fit test is constructed to check the null hypothesis that the unknown regression function belongs to a certain parametric family. The asymptotic distribution of the test statistic is obtained jointly with the power under local alternatives and a consistent bootstrap algorithm. The test is illustrated in a simulation study and is applied to the Slashdot dataset.

Chapter 8: Future research. Different ideas for future projects are outlined in this chapter: new bandwidth selectors in nonparametric linear-directional regression, a kernel density estimator for directional data under rotational symmetry, an `R` package implementing the methods described in this thesis and a goodness-of-fit test for the Johnson and Wehrly (1978) copula structure.

Appendix A: Supplement to Chapter 6. This supplement contains the proofs of the technical lemmas used in Chapter 6, exhaustive details of the simulation study, further results for the independence test and an extended data application.

Appendix B: Supplement to Chapter 7. Particular cases of the projected local linear estimator, proofs of the technical lemmas and further results for the simulation study for Chapter 7 are collected in this appendix.

Resumen en castellano. This last part presents a short summary of the thesis in Spanish.

References

Agostinelli, C. and Lund, U. (2013). *R package* `circular`*: circular statistics (version 0.4-7).*

Bai, Z. D., Rao, C. R., and Zhao, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.*, 27(1):24–39.

Banerjee, A., Dhillon, I. S., Ghosh, J., and Sra, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions. *J. Mach. Learn. Res.*, 6:1345–1382.

Barros, A. M. G., Pereira, J. M. C., and Lund, U. J. (2012). Identifying geographical patterns of wildfire orientation: a watershed-based analysis. *Forest. Ecol. Manag.*, 264:98–107.

Batschelet, E. (1981). *Circular statistics in biology.* Academic Press, Inc., London-New York.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Res.*, 28(1):235–242.

Bhattacharya, A. and Bhattacharya, R. (2012). *Nonparametric inference on manifolds*, volume 2 of *Institute of Mathematical Statistics Monographs*. Cambridge University Press, Cambridge.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.*, 1(6):1071–1095.

Bingham, M. S. and Mardia, K. V. (1975). Maximum likelihood characterization of the von Mises distribution. In Patil, G. P., Kotz, S., and Ord, J. K., editors, *A modern course on statistical distributions in scientific work*, volume 17 of *NATO Advanced Study Institutes Series. Series C, Mathematical and Physical Sciences.*, pages 387–398. D. Reidel, Dordrecht.

Boente, G. and Fraiman, R. (1991). Nonparametric regression for directional data. *Trabajos de Matemática*, 176.

Boente, G., Rodríguez, D., and González-Manteiga, W. (2014). Goodness-of-fit test for directional data. *Scand. J. Stat.*, 41(1):259–275.

Carnicero, J. A., Ausín, M. C., and Wiper, M. P. (2013). Non-parametric copulas for circular-linear and circular-circular data: an application to wind directions. *Stoch. Environ. Res. Risk Assess.*, 27(8):1991–2002.

Deschepper, E., Thas, O., and Ottoy, J. P. (2008). Tests and diagnostic plots for detecting lack-of-fit for circular-linear regression models. *Biometrics*, 64(3):912–920.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.*, 79(19):2066–2075.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2013). Non-parametric regression for circular responses. *Scand. J. Stat.*, 40(2):238–255.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2014). Nonparametric regression for spherical data. *J. Amer. Statist. Assoc.*, 109(506):748–763.

Dryden, I. L. (2005). Statistical analysis on high-dimensional spheres and shape spaces. *Ann. Statist.*, 33(4):1643–1665.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical shape analysis*. Wiley Series in Probability and Statistics. Probability and Statistics. John Wiley & Sons, Chichester.

Fan, Y. (1994). Testing the goodness of fit of a parametric density function by kernel method. *Economet. Theor.*, 10(2):316–356.

Fernández-Durán, J. J. (2007). Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics*, 63(2):579–585.

Fernández-Durán, J. J. and Gregorio-Domínguez, M. M. (2013). *CircNNTSR: an R package for the statistical analysis of circular data using NonNegative Trigonometric Sums (NNTS) models.* R package version 2.1.

Fisher, N. I. (1993). *Statistical analysis of circular data*. Cambridge University Press, Cambridge.

Fisher, N. I. and Lee, A. J. (1981). Nonparametric measures of angular-linear association. *Biometrika*, 68(3):629–636.

Fisher, N. I., Lewis, T., and Embleton, B. J. J. (1993). *Statistical analysis of spherical data*. Cambridge University Press, Cambridge.

García-Portugués, E. (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron. J. Stat.*, 7:1655–1685.

García-Portugués, E., Barros, A. M. G., Crujeiras, R. M., González-Manteiga, W., and Pereira, J. (2014a). A test for directional-linear independence, with applications to wildfire orientation and size. *Stoch. Environ. Res. Risk Assess.*, 28(5):1261–1275.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013a). Exploring wind direction and SO_2 concentration by circular-linear density estimation. *Stoch. Environ. Res. Risk Assess.*, 27(5):1055–1067.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013b). Kernel density estimation for directional-linear data. *J. Multivariate Anal.*, 121:152–175.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2014b). Central limit theorems for directional and linear data with applications. *Statist. Sinica*, to appear.

García-Portugués, E., Van Keilegom, I., Crujeiras, R. M., and González-Manteiga, W. (2014). Testing parametric models in linear-directional regression. *arXiv:1409.0506*.

Gatto, R. (2011). The generalized von Mises-Fisher distribution. In Wells, M. T. and SenGupta, A., editors, *Advances in directional and linear statistics*, pages 51–68. Physica, Heidelberg.

Gijbels, I. and Mielniczuk, J. (1990). Estimating the density of a copula function. *Comm. Statist. Theory Methods*, 19(2):445–464.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.*, 14(1):1–16.

Hall, P., Watson, G. S., and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762.

Hamelryck, T., Mardia, K., and Ferkinghoff-Borg, J., editors (2012). *Bayesian methods in structural bioinformatics*. Statistics for Biology and Health. Springer, Berlin.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, 21(4):1926–1947.

Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in circular statistics*, volume 5 of *Series on Multivariate Analysis*. World Scientific Publishing, River Edge.

Johnson, R. A. and Wehrly, T. (1977). Measures and models for angular correlation and angular-linear correlation. *J. Roy. Statist. Soc. Ser. B*, 39(2):222–229.

Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *J. Amer. Statist. Assoc.*, 73(363):602–606.

Jona-Lasinio, G., Gelfand, A., and Jona-Lasinio, M. (2012). Spatial analysis of wave direction data using wrapped gaussian processes. *Ann. Appl. Stat.*, 6(4):1478–1498.

Jones, M. C. and Pewsey, A. (2005). A family of symmetric distributions on the circle. *J. Amer. Statist. Assoc.*, 100(472):1422–1428.

Kendall, D. G., Barden, D., Carne, T. K., and Le, H. (1999). *Shape and shape theory*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester.

Klemelä, J. (2000). Estimation of densities and derivatives of densities with directional data. *J. Multivariate Anal.*, 73(1):18–40.

Mardia, K. V. (1972). *Statistics of directional data*, volume 13 of *Probability and Mathematical Statistics*. Academic Press, London.

Mardia, K. V. (1976). Linear-circular correlation coefficients and rhythmometry. *Biometrika*, 63(2):403–405.

Mardia, K. V. and Jupp, P. E. (2000). *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, second edition.

Mardia, K. V. and Sutton, T. W. (1978). A model for cylindrical variables with applications. *J. Roy. Statist. Soc. Ser. B*, 40(2):229–233.

Meyer, D., Hornik, K., and Feinerer, I. (2008). Text mining infrastructure in R. *J. Stat. Softw.*, 25(5):1–54.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition.

Ochsenbein, F., Bauer, P., and Marcout, J. (2000). The VizieR database of astronomical catalogues. *Astron. Astrophys. Supp. Ser.*, 143(23):23–32.

Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2012). A plug-in rule for bandwidth selection in circular density estimation. *Comput. Statist. Data Anal.*, 56(12):3898–3908.

Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2013). Nonparametric circular methods for exploring environmental data. *Environ. Ecol. Stat.*, 20(1):1–17.

Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2014). NPCirc: nonparametric circular methods. *J. Stat. Softw.*, to appear.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F., and Clark, C. W., editors (2010). *NIST handbook of mathematical functions*. Cambridge University Press, Cambridge.

Perryman, M. A. C. (1997). *The Hipparcos and Tycho catalogues*, volume 1200 of *ESA SP*. ESA Publication Division, Noordwijk.

Pewsey, A., Neuhäuser, M., and Ruxton, G. D. (2013). *Circular statistics in R*. Oxford University Press, Oxford.

Richardson, J. S. (2011). Protein backbone PhiPsiOmega drawing. Retrieved August 16, 2014 from `http://commons.wikimedia.org/wiki/File:Protein_backbone_PhiPsiOmega_drawing.jpg`.

Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.*, 3(1):1–14.

Rosenblatt, M. and Wahlen, B. E. (1992). A nonparametric measure of independence under a hypothesis of independent components. *Statist. Probab. Lett.*, 15(3):245–252.

Shieh, G. S. and Johnson, R. A. (2005). Inferences based on a bivariate distribution with von Mises marginals. *Ann. Inst. Statist. Math.*, 57(4):789–802.

Taylor, C. C. (2008). Automatic bandwidth selection for circular density estimation. *Comput. Statist. Data Anal.*, 52(7):3493–3500.

van Leeuwen, F. (2007). *Hipparcos, the new reduction of the raw data*, volume 350 of *Astrophysics and Space Science Library*. Springer, Dordrecht.

Wang, X., Zhao, L., and Wu, Y. (2000). Distribution free laws of the iterated logarithm for kernel estimator of regression function based on directional data. *Chinese Ann. Math. Ser. B*, 21(4):489–498.

Watson, G. S. (1983). *Statistics on spheres*, volume 6 of *University of Arkansas Lecture Notes in the Mathematical Sciences*. John Wiley & Sons, New York.

Wehrly, T. E. and Johnson, R. A. (1979). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, 67(1):255–256.

Zhao, L. and Wu, C. (2001). Central limit theorem for integrated square error of kernel estimators of spherical density. *Sci. China Ser. A*, 44(4):474–483.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Chapter 2

Density estimation by circular-linear copulas

Abstract

The study of environmental problems usually requires the description of variables with different nature and the assessment of relations between them. In this work, an algorithm for flexible estimation of the joint density for a circular-linear variable is proposed. The method is applied for exploring the relation between wind direction and SO_2 concentration in a monitoring station close to a power plant located in Galicia (NW-Spain), in order to compare the effectiveness of precautionary measures for pollutants reduction in two different years.

Reference

García-Portugués, E., Crujeiras, R. M. and González-Manteiga, W. (2013). Exploring wind direction and SO_2 concentration by circular-linear density estimation. *Stoch. Environ. Res. Risk Assess.*, 27:1055–1067.

Contents

2.1 Introduction **21**
2.2 Background **24**
2.2.1 Some circular and circular-linear distributions 24
2.2.2 Some notes on copulas 25
2.3 Estimation algorithm **27**
2.3.1 Some simulation results 30
2.4 Application to wind direction and SO_2 concentration **35**
2.4.1 Exploring SO_2 and wind direction in 2004 and 2011 36
2.5 Final comments **36**
References **38**

2.1 Introduction

Air pollution studies require the investigation of relationships between emission sources and pollutants concentration in nearby sites. In addition, the effectiveness of environmental policies

with the aim of pollution reduction should be checked, at least in a descriptive way, in order to assess whether the implemented precautionary measurements had worked out.

Different statistical methods have been considered for the study of the relation between wind direction and pollutants concentration, both for exploratory and for inferential analysis, taking into account that wind direction is a circular variable requiring a proper statistical treatment. Wind potential assessment using descriptive methods and spectral analysis has been carried out by Shih (2008). In addition, wind direction has been proved to play a significant role in detection of emission sources (see Chen et al. (2012)) and air quality studies (see Bayraktar et al. (2010)), although the wind direction is not treated as a circular variable, but discretized as a factor. Jammalamadaka and Lund (2006) considered regression models for the pollutants concentration (linear response) over the wind direction (circular explanatory variable), constructing the regression function in terms of the sine and cosine components of the circular variable. Recently, Deschepper et al. (2008) introduced a graphical diagnostic tool, jointly with a test, for fit assessment in parametric circular-linear regression, illustrating the technique in an air quality environmental study.

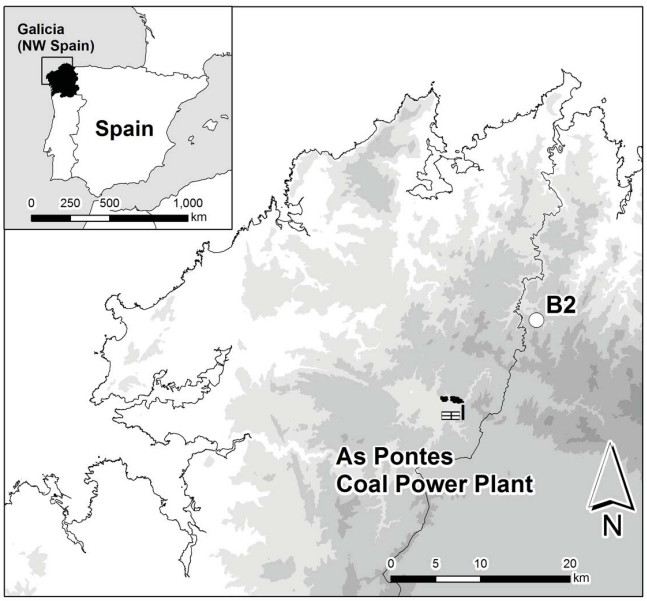


Figure 2.1: Locations of monitoring station (circle) and power plant in Galicia (NW-Spain). Location of station B2: 7° 44' 10" W, 43° 32' 05" N. Power plant location: 7° 51' 45" W, 43° 26' 26" N.

From a more technical perspective, Johnson and Wehrly (1978) and Wehrly and Johnson (1979) presented a method for obtaining joint circular-linear and circular-circular densities with specified marginals, respectively. Fernández-Durán (2004) introduced a new family of circular distributions based on nonnegative trigonometric sums, and this idea is used in Fernández-Durán (2007) in the construction of circular-linear densities, adapting the proposal of Johnson and Wehrly (1978). The introduction of nonnegative trigonometric sums for modelling the circular distributions involved in the formulation of Johnson and Wehrly (1978) allows for more flexible models, that may present skewness or multimodality, features that cannot be reflected through

the von Mises distribution (the classical model for circular variables). However, the flexibility claimed in this proposal can be also obtained by a completely nonparametric approach.

In this work, a procedure for modelling the relation between a circular and a linear variable is proposed. The relation is specified by a circular-linear density which, represented in terms of copulas, can be estimated nonparametrically. The estimation algorithm can be also adapted to a semiparametric framework, when an underlying model for the marginal distributions can be imposed, and it also comprises the classical Johnson and Wehrly (1978) model. It also enables the construction of an estimation framework without imposing an underlying parametric model. The copula approach presents some computational advantages, in order to carry out a simulation study.

The practical aim of this work is to explore the relation between wind incidence direction (wind blowing from this direction) and SO_2 levels in a monitoring station close to a power plant located in Galicia (NW-Spain). The monitoring station and the thermal power plant locations are shown in Figure 2.1. In the power plant, energy was usually produced from the combustion of local coal, which also generates pollutants as sulphur dioxide (SO_2). In order to reduce the emission of SO_2 to the atmosphere, and to comply with European regulations, coal with less sulphur content has been used since 2005. In addition, the power plant changed the energy production system by settling a combined process of coal and gas burning in 2008. These measures were aimed to reduce the SO_2 emissions.

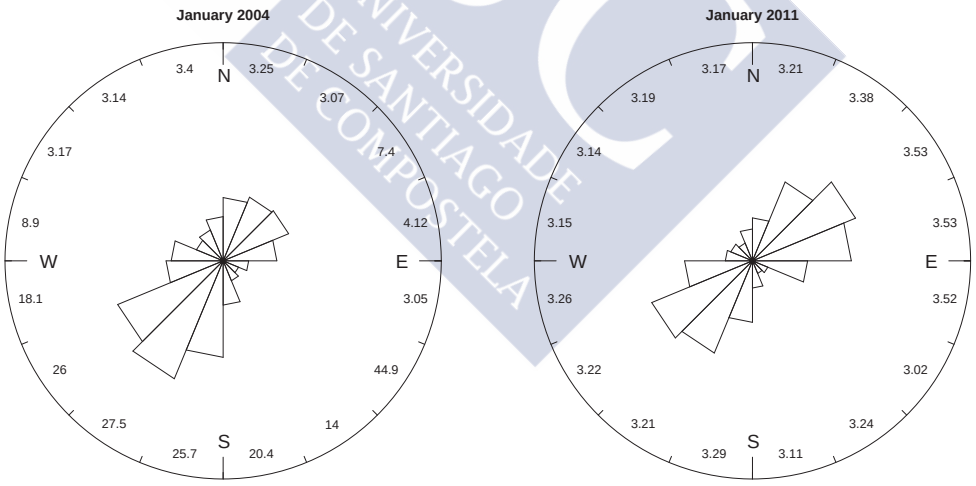


Figure 2.2: Rose diagrams for wind direction in station B2 for 2004 (left plot) and 2011 (right plot), with average SO_2 concentration.

The analysed data corresponds to SO_2 and wind incidence direction measurements taken during January 2004 and January 2011, with one minute frequency. The monitoring station B2 (see Figure 2.1) is located in a wind farm 13.4 kilometres in the NE direction with respect to the power plant. In Figure 2.2, rose diagrams for wind direction are shown, including also the average SO_2 concentration for each wind direction sector. It can be seen that higher values of SO_2 are shown in 2004. Note also that there are two dominant wind incidence directions, specifically, blowing from SW and from NE. In addition, note that the average SO_2 values for

winds blowing from SW are remarkably high for 2004, which is explained by the position of the monitoring station with respect to the power plant. However, the relation between wind direction and SO_2 levels is not clear from these representations, and the dependency (or lack of dependency) between them should be investigated.

This work is organized as follows. Section 2.2 provides some background on circular and circular-linear random variables and a brief review on copula methods. The algorithm for estimating a circular-linear density is detailed and discussed in Section 2.3. The finite sample properties of the algorithm, combining parametric and nonparametric approaches, are illustrated by a simulation study. The completely nonparametric version of this algorithm is applied for analysing the relation between wind direction and SO_2 concentration in Section 2.4. Some final comments are given in Section 2.5.

2.2 Background

The main goal of this work is to describe the relation between wind direction and SO_2 concentration in a monitoring station close to a power plant at two different time moments, before and after precautionary measurements to reduce SO_2 emissions were applied. Bearing in mind the different nature of the variables and noticing that measurements from wind direction are angles, some background on circular and circular-linear random variables is introduced. This methodology will be needed in order to describe the wind direction itself and the joint relation between the two variables. For that purpose, the Johnson and Wehrly (1978) family (J&W in what follows) of circular-linear distributions will be introduced. A general procedure, based on the copula representation of a density, allows for a more flexible estimation framework. The goal of this copula representation is twofold: firstly, the classical J&W family can be written in such a way, just involving univariate (circular and linear) densities; secondly, with the copula representation, flexible circular-linear relations beyond this specific model are also possible.

2.2.1 Some circular and circular-linear distributions

Denote by Θ a circular random variable with support in the unit circle $\mathbb{S}^1$. A circular distribution Ψ for Θ assigns a probability to each direction $(\cos(\theta), \sin(\theta))$ in the plane $\mathbb{R}^2$, characterized by the angle $\theta \in [0, 2\pi)$ (see Mardia and Jupp (2000) for a survey on statistical methods for circular data). The von Mises distribution is the analogue of the normal distribution in circular random variables. This family of distributions, usually denoted by $\mathrm{vM}(\mu, \kappa)$, is characterized by two parameters: $\mu \in [0, 2\pi)$, the circular mean and $\kappa \geq 0$, a circular concentration parameter around μ. The corresponding density function is given by

$$\varphi_{\mathrm{vM}}(\theta; \mu, \kappa) = \frac{1}{2\pi \mathcal{I}_0(\kappa)} e^{\kappa \cos(\theta - \mu)}, \quad \theta \in [0, 2\pi), \tag{2.1}$$

being $\mathcal{I}_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos \omega} d\omega$ the modified Bessel function of first kind and order zero. The von Mises cumulative distribution function, considering the zero angle as the starting point, is defined as:

$$\Psi_{\mathrm{vM}}(\theta; \mu, \kappa) = \int_0^{\theta} \varphi_{\mathrm{vM}}(\omega; \mu, \kappa)\, d\omega, \quad \theta \in [0, 2\pi).$$

The uniform circular distribution,

$$\varphi_{\mathrm{U}}(\theta) = \frac{1}{2\pi}, \quad \theta \in [0, 2\pi), \tag{2.2}$$

is obtained as a particular case of the von Mises family, for $\kappa = 0$. Circular density estimation can be performed by parametric methods, such as Maximum Likelihood Estimation (MLE), or using nonparametric techniques, based on kernel approaches, as will be seen later.

In order to explain the relation between a circular and a linear random variable (in our example, wind direction and SO_2 concentration), the construction of a joint circular-linear density will be considered. A circular-linear random variable (Θ, X) is supported on the cylinder $\mathbb{S}^1 \times \mathbb{R}$ or in a subset of it and a circular-linear density for (Θ, X), namely p, must satisfy a periodicity condition in the circular argument, that is:

$$p(\theta, x) = p(\theta + 2\pi k, x), \quad \theta \in [0, 2\pi),\, x \in \mathbb{R},\, k \in \mathbb{Z},$$

as well as the usual assumptions on taking nonnegative values and integrating one. Johnson and Wehrly (1978) proposed a method for obtaining circular-linear densities with specified marginals. Denote by φ and f the circular and linear marginal densities, respectively, and by Ψ and F their corresponding distribution functions. Let also g be another circular density. Then

$$p(\theta, x) = 2\pi g\left(2\pi\left(\Psi(\theta) - F(x)\right)\right) \times \varphi(\theta) f(x) \tag{2.3}$$

is a density for a circular-linear distribution for a random variable (Θ, X), with specified marginal densities φ and f (see Johnson and Wehrly (1978), Theorem 5). Circular-linear densities with specified marginals can be also obtained considering the sum of the marginal distributions in the argument of the joining density in (2.3). Circular-linear densities may include von Mises and Gaussian marginals, but the dependence between them will be specified by the joining density g. In fact, the independence model corresponds to taking φ_U in (2.2) as the joining density. From a data sample of (Θ, X), assuming that the joint density can be represented as in (2.3), an estimator of p could be obtained by the estimations of the marginals and the joining density. Wehrly and Johnson (1979) proved that the construction of circular-circular distributions (that is, distributions on the torus) can be done similarly to (2.3), just considering prespecified circular marginal distributions. Note that (2.3) is just a construction method and not a characterization of circular-linear densities. In addition, there are no available testing procedures for checking if a certain dataset follows such a distribution. Hence, a more general approach for circular-linear density construction would be helpful.

In the next section, some background on copulas will be introduced, allowing for a more flexible procedure for obtaining circular-linear densities with specified marginals, where the representation in (2.3) fits as a particular case. With this proposal, a fully nonparametric estimation procedure can be applied.

2.2.2 Some notes on copulas

Copula functions are multivariate distributions with uniform marginals (see Nelsen (2006) for a complete review on copulas). One of the main results in copula theory is Sklar's theorem, which, in the bivariate case, states that if F is a joint distribution function with marginals F_1 and F_2 then there exists a copula C such that:

$$F(x, y) = C(F_1(x), F_2(y)), \quad x, y \in \mathbb{R}. \tag{2.4}$$

If F_1 and F_2 are continuous distributions, then C is unique. Conversely, if C is a copula and F_1 and F_2 are distribution functions, then F defined by (2.4) is a bivariate distribution with marginals F_1 and F_2.

If the marginal random variables are absolutely continuous, Sklar's result can be interpreted in terms of the corresponding densities. Denoting by c the copula density, the bivariate density of F in (2.4) can be written as

$$f(x,y) = c(F_1(x), F_2(y)) \times f_1(x) f_2(y), \quad x, y \in \mathbb{R}.$$

As pointed out by Nazemi and Elshorbagy (2012), copula modelling provides a simple but powerful way for describing the interdependence between environmental variables. In our particular setting, the nature of the variables is different, being the variables of interest linear (SO_2 concentration) and circular (wind direction). Circular-linear copulas, that will be denoted by $C_{\Theta,X}$, can be also defined taking into account the characteristics of the circular marginal, satisfying $c_{\Theta,X}(0,v) = c_{\Theta,X}(1,v)$, $\forall v \in [0,1]$, where $c_{\Theta,X}$ is the corresponding circular-linear copula density. Hence, a circular-linear density with marginals φ and f is given by:

$$p(\theta, x) = c_{\Theta,X}(\Psi(\theta), F(x)) \times \varphi(\theta) f(x), \quad \theta \in [0, 2\pi),\ x \in \mathbb{R}. \tag{2.5}$$

Note that J&W's proposal can be seen as a particular case of (2.5). For a certain joining density g in (2.3), the corresponding copula density is given by:

$$c_{\Theta,X}\left(\Psi(\theta), F(x)\right) = 2\pi g\left(2\pi\left(\Psi(\theta) \pm F(x)\right)\right), \tag{2.6}$$

where the sign $\pm$ refers to the possibility of considering the sum or the difference of the marginal distributions in the argument of g, as it was previously mentioned.

A circular-circular density, following the result of Wehrly and Johnson (1979), can be constructed similarly just considering circular marginals for (Θ, Ω) and guaranteeing that the copula density satisfies $c_{\Theta,\Omega}(0,v) = c_{\Theta,\Omega}(1,v)$ and $c_{\Theta,\Omega}(u,0) = c_{\Theta,\Omega}(u,1)$, $\forall u, v \in [0,1]$. For simplicity, the copula density in (2.6) will be denoted as J&W copula (see Figure 2.3, left plot).

The representation of a circular-linear density in (2.5) enables the construction of new families of circular-linear distributions. From Corollary 3.2.5. in Nelsen (2006), a new family of circular-linear copulas with quadratic section (QS copula) in the linear component can be constructed. The copula densities in this family are given by

$$c^{\alpha}_{\Theta,X}(u,v) = 1 + 2\pi\alpha\cos(2\pi u)(1-2v), \quad |\alpha| \leq (2\pi)^{-1}. \tag{2.7}$$

The QS copula family is parametrized by α, which accounts for the deviation from the independence copula corresponding to $\alpha = 0$. Figure 2.3 (middle plot) shows the wavy surface corresponding to $\alpha = (2\pi)^{-1}$. The position of the three modes in the density, centred along $u = 0$, $u = 0.5$ and $u = 1$, as well as their concentration, is controlled by the value of α.

A possible way to derive new copulas is through mixtures of other copulas (see Nelsen (2006)). Thus, for any copula $\tilde{c}$, the mixture

$$c_{\Theta,X}(u,v) = \frac{1}{4}\left(\tilde{c}(u,v) + \tilde{c}(1-u,v) + \tilde{c}(u,1-v) + \tilde{c}(1-u,1-v)\right), \tag{2.8}$$

leads to a new copula satisfying $c_{\Theta,X}(0,v) = c_{\Theta,X}(1,v)$, $\forall v \in [0,1]$. This construction also satisfies $c_{\Theta,X}(u,0) = c_{\Theta,X}(u,1)$, $\forall u \in [0,1]$, and provides a circular-circular copula (which is also circular-linear). A parametrized copula density $c^{\alpha}_{\Theta,X}$ can be obtained considering, for example, the parametric Frank copula:

$$\tilde{c}_{\alpha}(u,v) = \frac{\alpha(1-e^{-\alpha})e^{-\alpha(u+v)}}{\left((1-e^{-\alpha}) - (1-e^{-\alpha u})(1-e^{-\alpha v})\right)^2}, \quad \alpha \neq 0.$$

The mixture copula (2.8) will be referred to as the *reflected* copula of $\tilde{c}$. The parameter α also measures the deviation from independence, which is a limit case as α tends to zero. The copula density surface can be seen in Figure 2.3 (right plot). In this example, the copula density surface shows five modes concentrated in the corners and the middle point of the unit square, and the peakedness of the modes increases as α grows.

These three families will be considered in the simulation study. It should be also mentioned that the copula representation poses some computational advantages in order to reproduce by simulation data samples from circular-linear distributions (see Section 2.3.1). Finally, although this work is focused on the circular-linear case, some comments will be also made about circular-circular distributions.

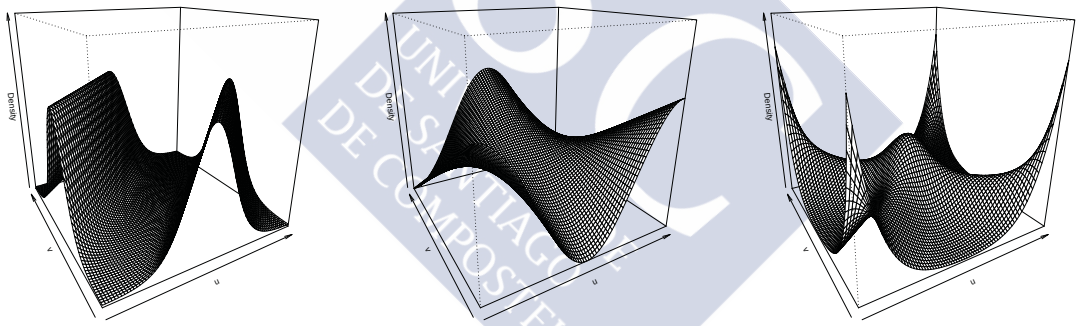


Figure 2.3: Copula density surfaces. Left plot: J&W copula with von Mises joining density with parameters $\mu = \pi$ and $\kappa = 2$. Middle plot: QS-copula with $\alpha = (2\pi)^{-1}$. Right plot: reflected Frank copula with $\alpha = 10$.

2.3 Estimation algorithm

Denote by $\{(\Theta_i, X_i)\}_{i=1}^n$ a random sample of data for the circular-linear random variables (Θ, X) and consider the copula representation for p in (2.5). In this joint circular-linear density model, three density functions must be estimated: the marginal densities φ and f (and also the corresponding distributions) and the copula density $c_{\Theta,X}$. A new natural procedure for estimating p is given in the following algorithm.

Algorithm 2.1 (Estimation algorithm)**.**

i. Obtain estimators for the marginal densities $\hat{\varphi}$, $\hat{f}$ and the corresponding marginal distributions $\hat{\Psi}$, $\hat{F}$.

ii. Compute an artificial sample $\{(\hat{\Psi}(\Theta_i), \hat{F}(X_i))\}_{i=1}^n$ *and estimate the copula density* $\hat{c}_{\Theta,X}$.

iii. Obtain the circular-linear density estimator as $\hat{p}\,(\theta, x) = \hat{c}_{\Theta,X}(\hat{\Psi}(\theta), \hat{F}(x)) \times \hat{\varphi}(\theta)\hat{f}(x)$.

The estimation of the marginal densities in step i can be done by parametric methods or by nonparametric procedures. For instance, a parametric estimator for $\hat{f}$ (respectively, for $\hat{F}$) can be obtained by MLE. In the circular case, that is, for obtaining $\hat{\varphi}$, MLE approaches are also possible (see Jammalamadaka and SenGupta (2001), Chapter 4). These estimators are consistent, although restricted to parametric families such as the von Mises distribution or mixtures of von Mises.

Nonparametric kernel density estimation for linear random variables was introduced by Parzen and Rosenblatt (see Wand and Jones (1995) for references on kernel density estimation) and the properties of this estimator have been well studied in the statistical literature. Consider $\{X_i\}_{i=1}^n$ a random sample of a linear variable X with density f. The kernel density estimator of f in a point $x \in \mathbb{R}$ is given by

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{2.9}$$

where K is a kernel function (usually a symmetric and unimodal density) and h is the bandwidth parameter. One of the crucial problems in kernel density estimation is the bandwidth choice. There exist several alternatives for obtaining a global bandwidth minimizing a certain error criterion, usually the Mean Integrated Squared Error (MISE), such as the rule-of-thumb, least-squares cross-validatory procedures (see Wand and Jones (1995)) or other plug-in rules, like the one proposed by Sheather and Jones (1991).

Hall et al. (1987) introduced a nonparametric kernel density estimator for directional data in the q-dimensional sphere $\mathbb{S}^q$. For the circular case ($q = 1$), denoting by Θ a random variable with density φ, the circular kernel density estimation from a sample $\{\Theta_i\}_{i=1}^n$ is given by

$$\hat{\varphi}_\nu(\theta) = \frac{c_0(\nu)}{n} \sum_{i=1}^{n} L\left(\nu \cos(\theta - \Theta_i)\right), \quad \theta \in [0, 2\pi), \tag{2.10}$$

where L is the circular kernel, ν is the circular bandwidth and $c_0(\nu)$ is a constant such that $\hat{\varphi}_\nu$ is a density. Some differences should be noted in contrast to the linear kernel density estimator in (2.9). First, the kernel function L must be a rapidly varying function, such as the exponential (see Hall et al. (1987)). Secondly, the behaviour of ν is opposite to h: in linear kernel density estimation, small values of the bandwidth h produce undersmoothed estimators (small values of ν oversmooth the density), whereas large values of h give oversmoothed curves (large values of ν produce undersmoothing). See Hall et al. (1987) for a detailed description of the estimator and its properties.

As in the linear case, bandwidth selection is also an issue in circular kernel density estimation. Although in the linear case it is a well-studied problem, for circular density estimation there are still some open questions. Hall et al. (1987) proposed selecting the smoothing parameter by maximum likelihood cross-validation. There are other recent proposals, such as the automatic bandwidth selection method introduced by Taylor (2008), but based on his results none of the selectors proposed seems to show a superior behaviour.

Although for the marginal distributions it may be reasonable to assume a parametric model, it is not that clear for the copula function, regarding the dependence structure between the marginals. Hence, in a general situation, the copula estimation in step *ii* should be carried out by a nonparametric procedure that will be explained below. However, for the J&W density in (2.3), the copula density $c_{\Theta,X}$ is linked with a joining circular density g in (2.6) and this circular density can be estimated in the same way as the marginal circular density. Note that, in this family, all the estimators involved in the algorithm are obtained in a strictly univariate way, which simplifies their computation.

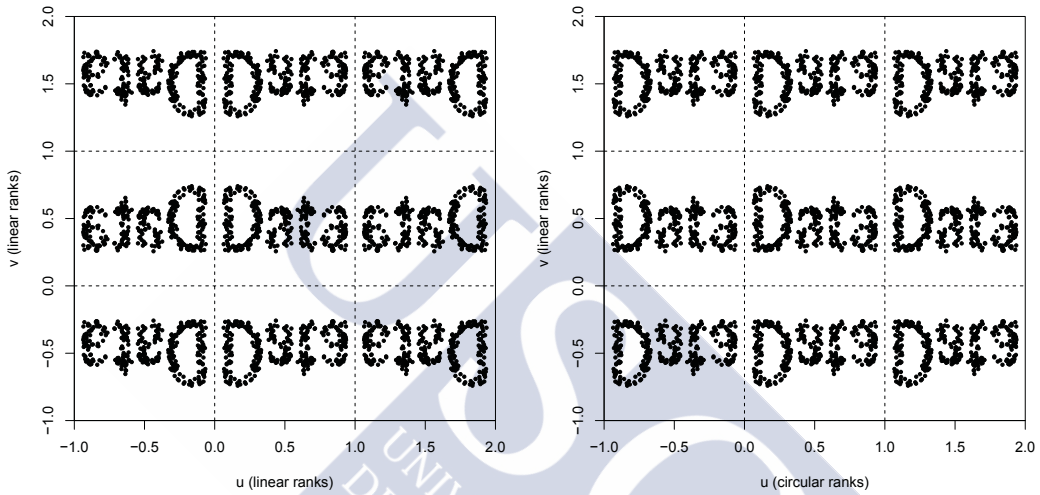


Figure 2.4: Illustration for data reflection for copula density estimation. The central square in each plot corresponds with the original data ranks. Left plot: mirror reflection from Gijbels and Mielniczuk (1990). Right plot: circular-mirror reflection for estimator (2.11).

Nonparametric copula density estimation can be also done by kernel methods, as proposed by Gijbels and Mielniczuk (1990). The proposed estimator is similar to the classical bivariate kernel density estimator, with a product kernel and with a mirror image data modification. This mirror image (see Figure 2.4, left plot) consists in reflecting the data with respect to all edges and corners of the unit square, in order to reduce the edge effect. In our particular case of circular-linear copula densities, the reflection must be done accounting for the circular nature of the first component, as shown in Figure 2.4, right plot.

The copula density kernel estimator can be defined as:

$$\hat{c}_{\Theta,X}(u,v) = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{9}\tilde{K}_B\left(u-\hat{\Psi}\left(\Theta_i^{(l)}\right), v-\hat{F}\left(X_i^{(l)}\right)\right), \tag{2.11}$$

where $\hat{\Psi}$ and $\hat{F}$ are the marginal kernel distribution functions, $\tilde{K}_B(u,v) = |B|^{-1}\tilde{K}(B^{-1/2}(u,v))$ is a bivariate rescaled kernel (see Ruppert and Wand (1994) for notation) and B is the bandwidth matrix ($|B|$ denotes the determinant). A plug-in rule for selecting the bandwidth matrix B was proposed by Duong and Hazelton (2003). The reflected data sample in each square $l = 1,\ldots,9$, namely $\{(\hat{\Psi}(\Theta_i^{(l)}), \hat{F}(X_i^{(l)}))\}_{i=1}^n$ is obtained by rows in the 3×3 plot (Figure 2.4, right plot) as follows: $(u-1,-v),(u,-v),(u+1,-v)$ (bottom row); $(u-1,v),(u,v),(u+1,v)$ (middle row); $(u-1,2-v),(u,2-v),(u+1,2-v)$ (top row).

As pointed out by Omelka et al. (2009), the estimator proposed by Gijbels and Mielniczuk (1990) in the linear case still suffers from corner bias, which could be corrected by bandwidth shrinking. This issue is not so important in the circular-linear setting, given the periodicity condition. It should be also noticed that, with this construction, the estimator obtained in step *iii*, with the proposed reflection, is guaranteed to be a density as long as the distribution and density estimators satisfy that $\hat{F}' = \hat{f}$ and $\hat{\Psi}' = \hat{\varphi}$.

The estimation algorithm can be extended for circular-circular densities, just with suitable (circular) density estimators for the marginals and a slight modification of the data reflection for the copula estimation. Specifically, reflection in 3×3 scheme as the one shown in Figure 2.4, the middle one corresponding to the original circular-circular data quantiles will be done as follows: $(u-1, v), (u, v), (u+1, v)$ (bottom row); $(u-1, v), (u, v), (u+1, v)$ (middle row); $(u-1, v), (u, v), (u+1, v)$ (top row).

2.3.1 Some simulation results

In order to check the performance of the estimation algorithm for circular-linear densities, the following scenarios are reproduced. Examples 2.1 and 2.2 were proposed by Johnson and Wehrly (1978). Example 2.3 corresponds to the QS-copula family given by (2.7) and Example 2.4 to the reflected Frank copula constructed in (2.8).

Example 2.1 (J&W copula with circular uniform and normal marginal distributions)**.** *Let φ_{U} denote the circular uniform density (2.2) and ϕ the standard normal density (Φ the standard normal distribution). Take the joining density $g = \varphi_{\mathrm{vM}}(\cdot; \mu, \kappa)$. A circular-linear density with marginals φ and ϕ is given by*

$$p_1(\theta, x) = \frac{1}{\mathcal{I}_0(\kappa)} \exp\{\kappa \cos(\theta - 2\pi\Phi(x) - \mu)\} \times \varphi_{\mathrm{U}}(\theta)\phi(x).$$

Example 2.2 (J&W copula with von Mises and normal marginal distributions)**.** *Consider $g = \varphi_{\mathrm{vM}}(\cdot; \mu', \kappa')$ and the density marginals ϕ and $\varphi_{\mathrm{vM}}(\cdot; \mu_2, \kappa_2)$ (with corresponding distributions Φ and $\Psi_{\mathrm{vM}}(\cdot; \mu_2, \kappa_2)$, respectively). A joint circular-linear density is given by*

$$p_2(\theta, x) = \frac{1}{\mathcal{I}_0(\kappa')} \exp\left\{\kappa' \cos\left(2\pi(\Psi_{\mathrm{vM}}(\theta; \mu_2, \kappa_2) - \Phi(x)) - \mu'\right)\right\} \times \varphi_{\mathrm{vM}}(\theta; \mu_2, \kappa_2)\phi(x).$$

Example 2.3 (QS-copula with von Mises and normal marginal distributions)**.** *Take $\varphi_{\mathrm{vM}}(\cdot; \mu_3, \kappa_3)$ and ϕ as marginals. The circular-linear density with copula (2.7) and $\alpha = (2\pi)^{-1}$ is given by*

$$p_3(\theta, x) = \left[1 + \frac{1}{2\pi}\cos(2\pi\Psi_{\mathrm{vM}}(\theta; \mu_3, \kappa_3))(1 - 2\Phi(x))\right] \times \varphi_{\mathrm{vM}}(\theta; \mu_3, \kappa_3)\phi(x).$$

Example 2.4 (Reflected Frank copula with von Mises and normal marginal distributions)**.** *Take $\varphi_{\mathrm{vM}}(\cdot; \mu_4, \kappa_4)$ and ϕ as marginals. The circular-linear density with reflected Frank copula ($\alpha = 10$) is obtained by the mixture construction (2.8):*

$$p_4(\theta, x) = c^{\alpha}_{\Theta, X}\left(\Psi_{\mathrm{vM}}(\theta; \mu_4, \kappa_4), \Phi(x)\right) \times \varphi_{\mathrm{vM}}(\theta; \mu_4, \kappa_4)\phi(x).$$

As commented in Section 2, the formulation of the joint circular-linear density in terms of copulas simplifies the simulation of random samples. The general idea is to split the joint distribution

P by Sklar's theorem in a copula $C_{\Theta,X}$ and marginals Ψ and F. Therefore, if we simulate a sample from (U, V) (uniform random variables with copula $C_{\Theta,X}$) and we apply the marginal quantiles transformations, then $(\Psi^{-1}(U), F^{-1}(V))$ will be a sample from the distribution P.

The simulation of (U, V) values from the copula $C_{\Theta,X}$ can be performed by the conditional method for simulating multivariate distributions (see Johnson (1987)). The conditional distribution of V given $U = u$, denoted by C_u, can be expressed as

$$C_u(v) = \frac{\partial C_{\Theta,X}(u, v)}{\partial u} = \int_0^v c_{\Theta,X}(u, t)\, dt, \tag{2.12}$$

where the first equality is an immediately property of copulas. So, for simulating random samples for the examples, or more generally, for simulating random samples of circular-linear random variables with density (2.5), we may proceed with the following algorithm.

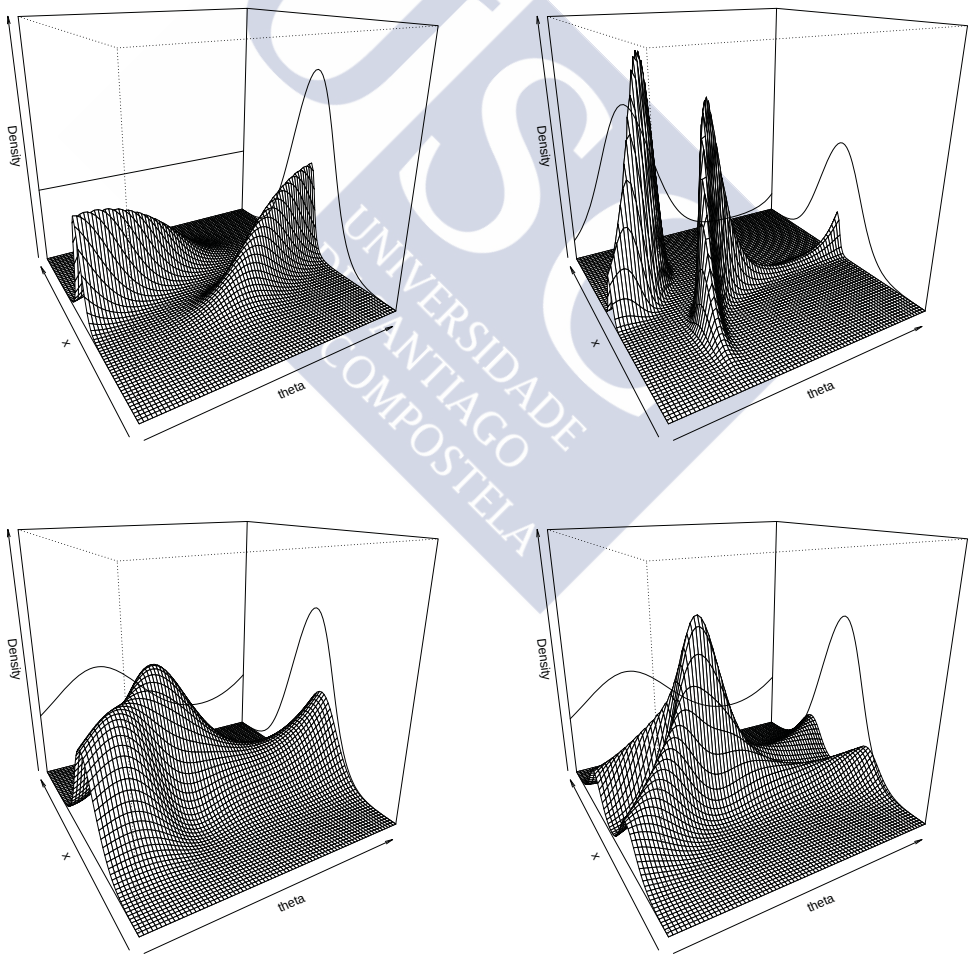


Figure 2.5: Density surfaces for the simulation study. Top left: Example 2.1 with $\mu = \pi$ and $\kappa = 2$. Top right: Example 2.2 with $\mu' = \pi$, $\kappa' = 5$, $\mu_2 = \pi/2$ and $\kappa_2 = 2$. Bottom left: Example 2.3 with $\alpha = (2\pi)^{-1}$, $\mu_3 = \pi/2$ and $\kappa_3 = 0.5$. Bottom right: Example 2.4 with $\alpha = 10$, $\mu_4 = \pi/2$ and $\kappa_4 = 0.5$.

Algorithm 2.2 (Simulation algorithm)**.**

i. Simulate $U, W \sim \mathcal{U}(0,1)$*, where* $\mathcal{U}(0,1)$ *stands for the standard uniform distribution.*

ii. Compute $V = C_u^{-1}(W)$*.*

iii. Obtain $\Theta = \Psi^{-1}(U)$*,* $X = F^{-1}(V)$*.*

In step *i*, two independent and uniformly distributed random variables are simulated. The conditional simulation method for obtaining (U, V) from the circular-linear copula $C_{\Theta,X}$ is performed in step *ii*. Finally, quantile transformations from the marginals are applied, obtaining a sample from (Θ, X) following the joint density (2.3). For Examples 2.1 and 2.2, the conditional distribution C_u in (2.12) is related to a von Mises $\mathrm{vM}(\mu, \kappa)$ and in step *ii*, for each u, $V = (2\pi)^{-1}\ \Psi_{\mathrm{vM}}^{-1}(W; 2\pi u - \mu, \kappa)$. For Example 2.3, for each u, $V = \frac{a+1-\sqrt{(a+1)^2-4aW}}{2a}$, with $a = 2\pi\alpha\cos(2\pi u)$. Example 2.4 is simulated from the mixture of copulas (2.8), using that for the Frank copula, step *ii* becomes $V = -\frac{1}{\alpha}\log\left(1 + \frac{W(1-e^{-\alpha})}{W(e^{-\alpha u}-1)-e^{-\alpha u}}\right)$.

The estimation algorithm proposed will be applied to estimate the densities in the examples. Different versions of the estimation algorithm can be implemented, considering parametric and nonparametric estimation methods. In addition, for the J&W densities (Examples 2.1 and 2.2), the estimation of the circular-linear density can be approached by representation (2.3), in terms of a circular joining density, or by the more general representation (2.5). Summarizing, the following variants of the estimation algorithm will be presented:

i. J&W, parametric (JWP): based on representation (2.3), marginals as well as joining density are parametrically estimated. The results will be used as a benchmark for the J&W models (Examples 2.1 and 2.2).

ii. J&W, semiparametric (JWSP): based on representation (2.3), marginals are estimated parametrically and a nonparametric kernel method is used for the joining density.

iii. J&W, nonparametric (JWNP): based on representation (2.3), marginals and joining density are estimated by kernel methods.

iv. Copula, semiparametric (CSP): based on representation (2.5), parametric estimation is considered for marginals. The copula density is estimated by kernel methods.

v. Copula, nonparametric (CNP): based on representation (2.5), marginals and copula density are estimated by kernel methods.

In the parametric case, density estimators have been obtained by MLE, specifying the von Mises family for the circular distributions and the normal family for the linear marginal. Nonparametric estimation has been carried out using kernel methods. The kernel density estimator in (2.9), with Gaussian kernel and Sheather and Jones (1991) bandwidth, has been used for obtaining $\hat{f}$. For $\hat{\varphi}$ and $\hat{g}$, the circular kernel density (2.10) has been implemented, with exponential kernel and likelihood cross-validatory bandwidth. In the semiparametric approaches (parametric marginals and nonparametric joining density or copula), Maximum Likelihood has been used for obtaining $\hat{\varphi}$ and $\hat{f}$. The circular kernel estimator (2.10) has been considered for $\hat{g}$. The copula density kernel estimator (2.11) has been used for $\hat{c}_{\Theta,X}$, with bivariate Gaussian kernel and restricted bandwidth matrix. Specifically, following the procedure of Duong and Hazelton (2003),

full bandwidth matrices have been tried, but with non significant differences in the values of the principal diagonal. Hence, a restricted bandwidth with two smoothing parameter values is used, considering the same element in the principal diagonal and a second element in the secondary diagonal, regarding for the kernel orientation.

In order to check the performance of the procedure for estimating circular-linear densities, the MISE criterion is considered:

$$\text{MISE} = \int_0^{2\pi} \int_{-\infty}^{\infty} \mathbb{E}\left[(\hat{p}(\theta, x) - p(\theta, x))^2\right] dx\, d\theta.$$

The MISE is approximated by Monte Carlo simulations, taking 1000 replicates. Four sample sizes have been used: $n = 50$, $n = 100$, $n = 500$ and $n = 1000$. In the first example, the set-up parameters are $\mu = \pi$ and $\kappa = 2$. For the second example, we take $\mu' = \pi$, $\kappa' = 5$, $\mu_2 = \pi/2$ and $\kappa_2 = 2$. Both in Examples 2.3 and 2.4 the parameters in the von Mises marginal were set to $\mu_3 = \mu_4 = \pi/2$ and $\kappa_3 = \kappa_4 = 0.5$, and the linear marginal is a standard normal.

		J&W-based			Copula-based		Relative efficiency			
	n	JWP	JWSP	JWNP	CSP	CNP	JWSP	JWNP	CSP	CNP
Example 2.1	50	0.534	0.741	1.395	1.417	1.851	0.721	0.383	0.377	0.289
	100	0.266	0.420	0.851	0.923	1.180	0.632	0.312	0.288	0.225
	200	0.132	0.234	0.485	0.596	0.735	0.563	0.272	0.221	0.179
	500	0.055	0.109	0.237	0.333	0.392	0.504	0.232	0.165	0.140
	1000	0.027	0.062	0.136	0.216	0.244	0.432	0.196	0.124	0.109
Example 2.2	50	4.059	4.671	8.311	7.497	8.239	0.869	0.488	0.541	0.493
	100	2.090	2.516	5.602	4.845	5.376	0.830	0.373	0.431	0.389
	200	1.068	1.362	3.442	3.015	3.362	0.784	0.310	0.354	0.318
	500	0.429	0.610	1.797	1.566	1.746	0.703	0.239	0.274	0.246
	1000	0.211	0.336	1.061	0.941	1.046	0.627	0.199	0.224	0.202

Table 2.1: MISE $\times$ 100 for estimating the circular-linear density in Examples 2.1 and 2.2. Relative efficiencies for JWSP, JWNP, CSP and CNP are taken with respect to JWP.

		J&W-based		Copula-based	
	n	JWSP	JWNP	CSP	CNP
Example 2.3	50	0.881	1.150	0.612	0.826
	100	0.625	0.813	0.372	0.506
	200	0.483	0.595	0.237	0.307
	500	0.389	0.459	0.135	0.171
	1000	0.357	0.403	0.091	0.109
Example 2.4	50	1.648	1.995	1.158	1.443
	100	1.339	1.568	0.747	0.926
	200	1.168	1.322	0.502	0.607
	500	1.058	1.148	0.284	0.333
	1000	1.019	1.075	0.183	0.210

Table 2.2: MISE $\times$ 100 for estimating the circular-linear density in Examples 2.3 and 2.4.

In Figure 2.5, surface plots for the example densities are shown, the top row corresponding to the J&W family (Example 2.1 and Example 2.2). Simulation results for these examples can be seen in Table 2.1. For all the alternatives of the algorithm, the MISE is reduced when increasing the sample size. Example 2.2 presents higher values for the MISE, and it is due to the estimation of a more complex structure in the circular marginal density (circular uniform in Example 2.1 and von Mises in Example 2.2). In both models, the estimation methods providing the information about the J&W structure (that is, based on representation (2.3)) work better, as expected. Nevertheless, the copula based approaches, CSP and CNP, are competitive with the JWNP.

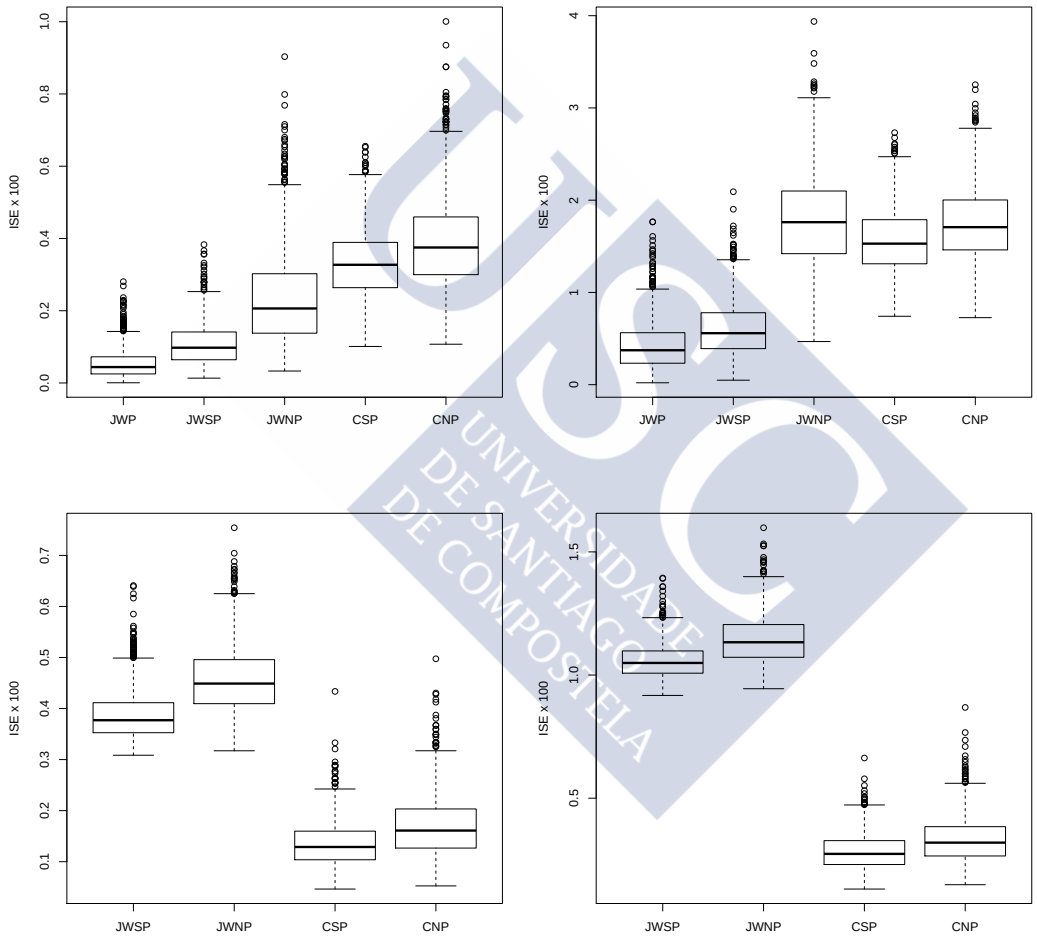


Figure 2.6: Boxplots of the ISE $\times$ 100 for Example 2.1 (top left), Example 2.2 (top right), Example 2.3 (bottom left) and Example 2.4 (bottom right) for $n = 500$ and different estimation procedures.

The parametric method JWP presents the lowest MISE values for all sample sizes in both examples, so it will be taken as a benchmark for computing the relative efficiencies of the nonparametric and semiparametric approaches, both based on the density (2.3) or on the copula density (2.5). Relative efficiencies are obtained as the ratio between the MISE of the parametric method and the MISE of the nonparametric and semiparametric procedures. The relative efficiencies (see Table 2.1) are higher for the semiparametric approach, with better results for Example 2.2. Boxplots for the ISE for sample size $n = 500$ can be seen in Figure 2.6, top row.

The larger variability of the nonparametric methods (JWNP and CNP) can be appreciated for both examples. For Example 2.2 (see Figure 2.6, top right plot), note that JWNP shows the highest values for the ISE.

Obviously, the first three proposals (JWP, JWSP, JWNP) make sense for distributions belonging to the J&W family. However, since the marginals considered in Examples 2.3 and 2.4 also belong to the von Mises (circular) and the normal (linear) classes, as in Examples 2.1 and 2.2, JWSP, JWNP, CSP and CNP variants of the algorithm will be applied in the last two cases. Results are reported in Table 2.2. MISE values decrease with sample size, showing CSP and CNP a similar behaviour. Note also that assuming a representation like (2.3), increases dramatically the MISE: for instance, in Example 2.3, for $n = 500$ or $n = 1000$, the MISE for JWSP or JWNP is four times the one provided by CNP. The effect of this misspecification in the copula structure can be clearly seen in the bottom row of Figure 2.6.

2.4 Application to wind direction and SO_2 concentration

The goal of this work is to explore the relation between wind incidence direction and SO_2 concentration in monitoring station B2 near a power plant (see Figure 2.1 for location of station B2). SO_2 is measured in $\mu g/m^3$ and wind direction as a counterclockwise angle in $[0, 2\pi)$. With this codification, 0, $\frac{\pi}{2}$, π and $\frac{3\pi}{2}$ represent east, north, west and south direction, respectively.

The dataset contains observations recorded minutely in January 2004 and January 2011, but due to technical limitations in the measuring device, SO_2 is only registered when it is higher than $3\,\mu g/m^3$. Concentration values below this threshold are considered as non significant and are recorded as the lower detection limit ($3\,\mu g/m^3$). Data have been hourly averaged, resulting 736 observations for 2004 and 743 for 2011. In order to avoid repeated data, perturbation procedures have been applied to both marginals, and will be detailed below. Afterwards, a Box-Cox transformation for the SO_2 concentration with $\lambda = -0.84$ for 2004 and $\lambda = -7.34$ for 2011 have been applied. For the sake of simplicity, we will refer to these transformed data as SO_2 concentration, but note that figures are shown in the transformed scale.

Measurement devices, both for the wind direction and for SO_2 concentration, did not present a sufficient precision to avoid repeated data, and this problem was inherited also for the hourly averages. The appearance of repeated measurements posed a problem in the application of the procedure, specifically, in the bandwidth computation. Perturbation in the linear variable, the SO_2 concentration, was carried out following Azzalini (1981). A pseudo-sample of SO_2 levels is obtained as follows:

$$\widetilde{X}_i = X_i + b\varepsilon_i,$$

where X_i denote the observed values, $b = 1.3\hat{\sigma}n^{-1/3}$ and ε_i, $i = 1, \ldots, n$, are independent and identically distributed random variables from the Epanechnikov kernel in $(-\sqrt{5}, \sqrt{5})$. $\hat{\sigma}$ is a robust estimator of the variance, which has been computed using the standardized interquartile range. Azzalini (1981) showed that this choice of b for the data perturbation allows for consistent estimation of the distribution function, getting a mean squared error with the same magnitude as the one from the empirical cumulative distribution function.

The problem of repeated measures also occurs for wind direction. In this case, a perturbation procedure similar to the linear variable case can be used, just considering the circular variable

(the angle) in the real line. Then, a pseudo-sample of wind direction was obtained as

$$\widetilde{\theta}_i = \theta_i + d\varepsilon_i,$$

with θ_i denoting the wind direction measurements and ε_i, $i = 1, \ldots, n$, were independently generated from a von Mises distribution with $\mu = 0$ and $\kappa = 1$, with $d = n^{-1/3}$. We have checked by simulations that the applied perturbation did not affect the distribution of the data.

2.4.1 Exploring SO_2 and wind direction in 2004 and 2011

The estimation procedure is applied to data from 2004 and 2011 in station B2, considering the fully nonparametric approach. Specifically, a nonparametric kernel density estimator for the SO_2 concentration was used, with the plug-in rule bandwidth obtained by method of Sheather and Jones (1991) ($h = 1.75 \times 10^{-6}$ for 2011, $h = 0.02$ for 2004). For the wind direction, circular kernel density estimation has been also performed, with likelihood cross-validatory bandwidth ($\nu = 46.41$ for 2011, $\nu = 78.56$ for 2004). In Figure 2.7, the estimation of the circular-linear density surfaces, with the corresponding contour plots, is shown.

For 2011 two modes in the SW and NE directions (see Figure 2.7, right column) can be identified, both with a similar behaviour and with quite low values of SO_2 concentrations. Recall that the scale is Box-Cox transformed (for data in the original scale, see Figure 2.2). A different situation occurs in 2004 (see Figure 2.7, left column) where, in addition to the two modes that appear in 2011 for low values of SO_2, a third mode arises. This mode is related to winds blowing from SW (from the thermal power plant) and to SO_2 values significantly higher than for 2011. This relation suggests that the additional mode represents SO_2 pollutants coming from the power plant, and its disappearance for 2011 illustrates the effectiveness of the control measures applied during 2005–2008 to reduce the SO_2 emissions from the power plant.

2.5 Final comments

A flexible algorithm for estimating circular-linear densities is proposed based on a copula density representation. The method provides a completely nonparametric estimator, but it can be modified to accommodate the classical Johnson and Wehrly family of circular-linear distributions. In the purely nonparametric version of the algorithm, circular and linear kernel density estimators have been used for the marginals, although other nonparametric density estimators could be considered. In addition, the extension of the algorithm for circular-circular density estimation is straightforward.

In our air quality data application the precision of the measurement devices posed some extra problems in the data analysis. The lack of precision resulted in the appearance of repeated values for the wind direction, and a data perturbation procedure was needed in order to apply the algorithm. The perturbation method proposed has been checked empirically, and it is inspired by the results for kernel distribution estimation, but our guess is that similar results could be obtained with just perturbing the data by summing errors from a highly concentrated distribution (e.g. a von Mises distribution with large κ). Nevertheless, data perturbation in the circular setting needs further investigation. Another possible problem that may be encountered in practice, for linear variables, is censoring, that may be due to detection limits or other

phenomena. Under censoring, the observation values are only partially known, and suitable estimation procedures for density estimation with censored data should be applied.

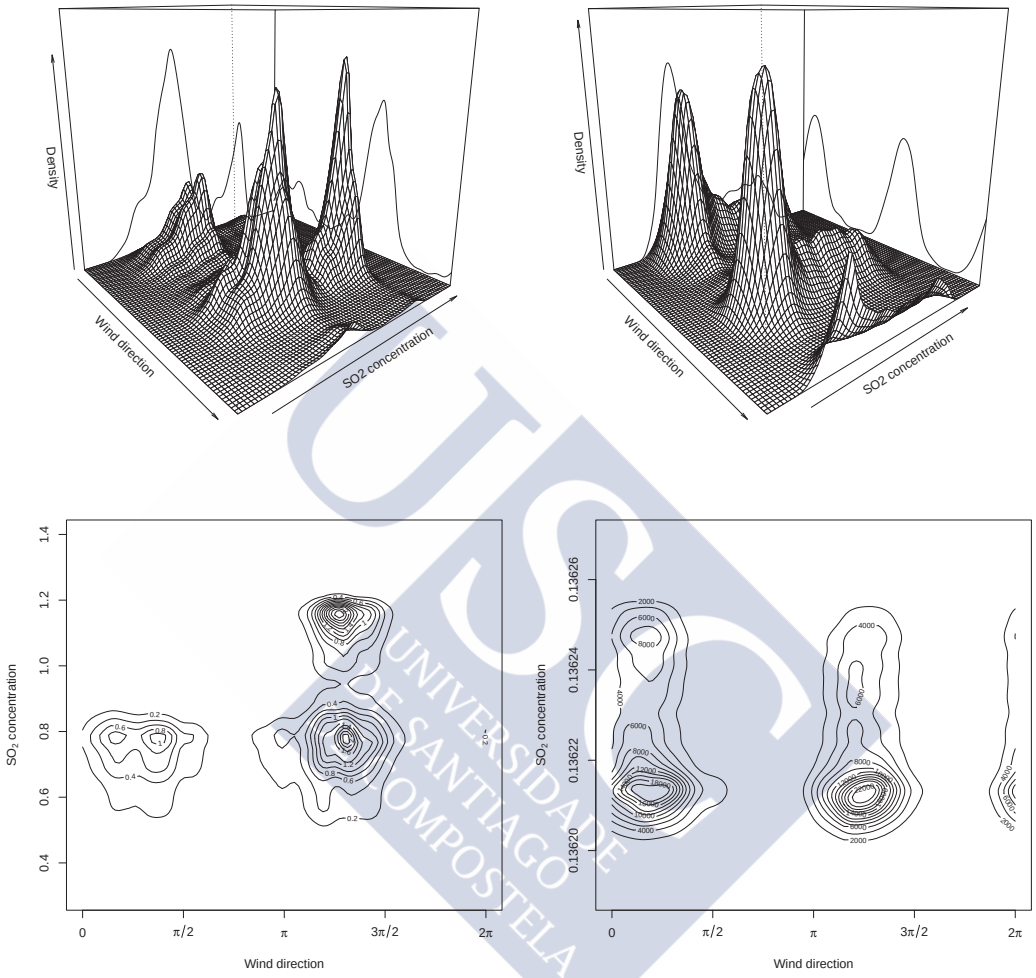


Figure 2.7: Circular-linear density estimator for wind direction and SO_2 concentrations in monitoring station B2. Left column: 2004. Right column: 2011.

Finally, a natural criticism of our analysis of the air quality data is that the measurements were not independent of one another. Temporal dependence could be accounted for in the estimation procedure by using a proper bandwidth selector, such as a k-fold cross-validatory bandwidth for the linear kernel density estimator. However, there are not such alternatives for circular data (up to the authors' knowledge) and the study of bandwidth selection rules for circular dependent data is beyond the scope of this paper.

The simulation study and real data analysis has been carried out in `R` 2.14 (R Development Core Team (2011)), using self-programmed code and package `circular` (Agostinelli and Lund (2011)). For the real data analysis, the computing time for 2004 is 30.77 seconds, taking the computation of the copula estimator 8.58 seconds. The same procedures take 31.57 seconds for the 2011 data. All computations were done on a computer with 1.6 GHz core. This shows that the computational cost of the method is not high and its application is feasible in practice.

Acknowledgements

The authors acknowledge the support of Project MTM2008–03010, from the Spanish Ministry of Science and Innovation and Project 10MDS207015PR from Dirección Xeral de I+D, Xunta de Galicia. Work of E. García-Portugués has been supported by FPU grant AP2010–0957 from the Spanish Ministry of Education. We also thank the referee and the Associate Editor for providing constructive comments and help in improving this paper.

References

Agostinelli, C. and Lund, U. (2011). *R package* `circular`*: circular statistics (version 0.4-3)*.

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, 68(1):326–328.

Bayraktar, H., Turalioğlu, F. S., and Tuncel, G. (2010). Average mass concentrations of TSP, PM10 and PM2.5 in Erzurum urban atmosphere, Turkey. *Stoch. Environ. Res. Risk Assess.*, 24(1):57–65.

Chen, C.-C., Wu, K.-Y., and Chang-Chien, G.-P. (2012). Point source identification using a simple permutation test: a case study of elevated PCDD/F levels in ambient air and soil and their relation to the distance to a local municipal solid waste incinerator. *Stoch. Environ. Res. Risk Assess.*, 26(2):225–233.

Deschepper, E., Thas, O., and Ottoy, J. P. (2008). Tests and diagnostic plots for detecting lack-of-fit for circular-linear regression models. *Biometrics*, 64(3):912–920.

Duong, T. and Hazelton, M. L. (2003). Plug-in bandwidth matrices for bivariate kernel density estimation. *J. Nonparametr. Stat.*, 15(1):17–30.

Fernández-Durán, J. J. (2004). Circular distributions based on nonnegative trigonometric sums. *Biometrics*, 60(2):499–503.

Fernández-Durán, J. J. (2007). Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics*, 63(2):579–585.

Gijbels, I. and Mielniczuk, J. (1990). Estimating the density of a copula function. *Comm. Statist. Theory Methods*, 19(2):445–464.

Hall, P., Watson, G. S., and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762.

Jammalamadaka, S. R. and Lund, U. J. (2006). The effect of wind direction on ozone levels: a case study. *Environ. Ecol. Stat.*, 13(3):287–298.

Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in circular statistics*, volume 5 of *Series on Multivariate Analysis*. World Scientific Publishing, River Edge.

Johnson, M. E. (1987). *Multivariate statistical simulation*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. John Wiley & Sons, New York.

Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *J. Amer. Statist. Assoc.*, 73(363):602–606.

Mardia, K. V. and Jupp, P. E. (2000). *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, second edition.

Nazemi, A. and Elshorbagy, A. (2012). Application of copula modelling to the performance assessment of reconstructed watersheds. *Stoch. Environ. Res. Risk Assess.*, 26(2):189–205.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition.

Omelka, M., Gijbels, I., and Veraverbeke, N. (2009). Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *Ann. Statist.*, 37(5B):3023–3058.

R Development Core Team (2011). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.*, 22(3):1346–1370.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B*, 53(3):683–690.

Shih, D. C.-F. (2008). Wind characterization and potential assessment using spectral analysis. *Stoch. Environ. Res. Risk Assess.*, 22(2):247–256.

Taylor, C. C. (2008). Automatic bandwidth selection for circular density estimation. *Comput. Statist. Data Anal.*, 52(7):3493–3500.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Wehrly, T. E. and Johnson, R. A. (1979). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, 67(1):255–256.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 3

Kernel density estimation for directional-linear data

Abstract

A nonparametric kernel density estimator for directional-linear data is introduced. The proposal is based on a product kernel accounting for the different nature of both (directional and linear) components of the random vector. Expressions for bias, variance and Mean Integrated Squared Error (MISE) are derived, jointly with an asymptotic normality result for the proposed estimator. For some particular distributions, an explicit formula for the MISE is obtained and compared with its asymptotic version, both for directional and directional-linear kernel density estimators. In this same setting a closed expression for the bootstrap MISE is also derived.

Reference

García-Portugués, E., Crujeiras, R. M. and González-Manteiga, W. (2013). Kernel density estimation for directional-linear data. *J. Multivariate Anal.*, 121:152–175.

Contents

3.1 Introduction **42**
3.2 Background on linear and directional kernel density estimation . . **43**
3.2.1 Kernel density estimation for directional data 43
3.2.2 Kernel density estimation for directional-linear data 47
3.3 Main results **47**
3.4 Error measurement and optimal bandwidth **49**
3.4.1 MISE for directional and directional-linear kernel density estimators . . 49
3.4.2 Some exact MISE calculations for mixture distributions 51
3.5 Conclusions **55**
3.A Some technical lemmas **56**
3.B Proofs of the main results **57**
3.C Proofs of the technical lemmas **70**
References **76**

3.1 Introduction

Kernel density estimation, and kernel smoothing methods in general, is a classical topic in nonparametric statistics. Starting from the first papers by Akaike (1954), Rosenblatt (1956) and Parzen (1962), extensions of the kernel density methodology have been brought up in different contexts, dealing with other smoothers, more complex data (censorship, truncation, dependence) or dynamical models (see Müller (2006) for a review). Some comprehensive references in this topic include the books by Silverman (1986), Scott (1992) and Wand and Jones (1995), among others.

Beyond the linear case, kernel density estimation has been also adapted to directional data, that is, data in the q-dimensional sphere (see Jupp and Mardia (1989) for a complete review of the theory of directional statistics). Hall et al. (1987) defined two type of kernel estimators and give asymptotic formulae of bias, variance and square loss. Almost simultaneously, Bai et al. (1988) established the pointwise, uniformly strong consistency and $\mathcal{L}_1$ consistency of a quite similar estimator in the same context. Later, Zhao and Wu (2001) stated a central limit theorem for the integrated squared error of the previous kernel density estimator based on the U-statistic martingale ideas developed by Hall (1984). Some of the results by Hall et al. (1987) were extended by Klemelä (2000), who studied the estimation of the Laplacian of the density and other types of derivatives. All these references consider the data lying on a general q-sphere of arbitrary dimension q, which comprises as particular cases circular data ($q = 1$) and spherical data ($q = 2$). For the particular case of circular data, there are more recent works dealing with the problem of smoothing parameter selection in kernel density estimation, such as Taylor (2008) and Oliveira et al. (2012). Di Marzio et al. (2011) study the kernel density estimator on the q-dimensional torus, and propose some bandwidth selection methods. A more general approach has been followed by Hendriks (1990), who discusses the estimation of the underlying distribution by means of Fourier expansions in a Riemannian manifold. This differential geometry viewpoint has been exploited recently by Pelletier (2005) and Henry and Rodriguez (2009). Nevertheless, the original approach seems to present a good balance between generality and complexity.

The aim of this work is to introduce and derive some basic properties of a joint kernel density estimator for directional-linear data, *i.e.* data with a directional and a linear component. This type of data arise in a variety of applied fields such as meteorology (when analysing the relation between wind direction and wind speed), oceanography (in the study of sea currents) and environmental sciences, among others. As an example, such an estimator has been used by García-Portugués et al. (2013) for studying the relation between pollutants and wind direction in the presence of an emission source. Specifically, the novelty of this work comprises the analysis of asymptotic properties of the directional-linear kernel density estimator, deriving bias, variance and asymptotic normality. As a by-product, the Mean Integrated Squared Error (MISE) follows, as well as the expression for optimal Asymptotic MISE (AMISE) bandwidths. In addition, for a particular class of densities consisting of mixtures of directional von Mises and normals, it is possible to compare the AMISE with the exact MISE. These results have been also obtained for the purely directional case, considering mixtures of von Mises distributions in the q-dimensional sphere, completing the existing results for directional data.

This paper is organized as follows. Section 3.2 presents some background on kernel density estimation for linear data and directional data. The proposed directional-linear kernel density estimator and the main results of this paper are included in Section 3.3, where the bias, variance and asymptotic normality are derived. Section 3.4 is focused in the issue of error measurement and expressions for the AMISE of the estimator and the exact MISE for particular cases of mixtures are obtained, both in the directional and directional-linear contexts. Conclusions and final comments are given in Section 3.5. The proofs of the results and some technical lemmas are given in the Appendix.

3.2 Background on linear and directional kernel density estimation

This section is devoted to a brief introduction on kernel density estimation for linear and directional data. For the sake of simplicity, f will denote the target density in this paper, which may be linear, directional, or directional-linear, depending on the context.

Let Z denote a linear random variable with support $\text{supp}(Z) \subseteq \mathbb{R}$ and density f. Consider $Z_1, \ldots, Z_n$ a random sample of Z, with size n. The linear kernel density estimator introduced by Akaike (1954), Rosenblatt (1956) and Parzen (1962) is defined as

$$\hat{f}_g(z) = \frac{1}{ng} \sum_{i=1}^{n} K\left(\frac{z - Z_i}{g}\right), \quad z \in \mathbb{R}, \tag{3.1}$$

where K denotes the kernel, usually a symmetric density about the origin, and $g > 0$ is the bandwidth parameter, which controls the smoothness of the estimator. Specifically, large values of the bandwidth parameter will produce oversmoothed estimates of f, whereas small values will provide undersmoothed curves. The asymptotic properties of this estimator and its adaptation to different contexts yielded a remarkably prolific field within the statistical literature, as noted in the introduction.

It is well known that under some regularity conditions on the kernel and the target density, the bias of the estimator (3.1) is of order $\mathcal{O}(g^2)$, whereas the variance is $\mathcal{O}((ng)^{-1})$, clearly showing the need of accounting for a trade-off between bias and variance in any bandwidth selection procedure. Specifically, the expected value of the linear kernel estimator at $z \in \mathbb{R}$ is:

$$\mathbb{E}\left[\hat{f}_g(z)\right] = f(z) + \frac{1}{2}\mu_2(K) f''(z) g^2 + o\left(g^2\right),$$

where $\mu_p(K) = \int_{\mathbb{R}} z^p K(z)\, dz$ represents the p-th moment of the kernel K. Similarly, the variance of (3.1) at $z \in \mathbb{R}$ is given by:

$$\mathbb{V}\text{ar}\left[\hat{f}_g(z)\right] = (ng)^{-1} R(K) f(z) + o\left((ng)^{-1}\right),$$

where $R(K) = \int_{\mathbb{R}} K^2(z)\, dz$. Further details on computations for the linear kernel density estimator can be found in Section 2.5 of Wand and Jones (1995).

3.2.1 Kernel density estimation for directional data

As previously mentioned, kernel density estimation has been adapted to different contexts such as directional data, that is, data on a q-dimensional sphere, being circular data ($q = 1$) and

spherical data ($q = 2$) particular cases. Let $\mathbf{X}$ denote a directional random variable with density f. The support of such a variable is the q-dimensional sphere, denoted by $\Omega_q = \{\mathbf{x} \in \mathbb{R}^{q+1} : x_1^2 + \cdots + x_{q+1}^2 = 1\}$. The Lebesgue measure in Ω_q will be denoted by ω_q and, therefore, a directional density satisfies

$$\int_{\Omega_q} f(\mathbf{x})\, \omega_q(d\mathbf{x}) = 1.$$

Remark 3.1. *When there is no possible misunderstanding, ω_q will also denote the surface area of Ω_q:*

$$\omega_q = \omega_q(\Omega_q) = \frac{2\pi^{\frac{q+1}{2}}}{\Gamma\left(\frac{q+1}{2}\right)}, \quad q \geq 1,$$

where Γ represents the Gamma function defined as $\Gamma(p) = \int_0^\infty x^{p-1}e^{-x}\, dx$, for $p > -1$.

The directional kernel density estimator was proposed by Hall et al. (1987) and Bai et al. (1988), following two different perspectives in the treatment of directional data. In this paper, the definition in Bai et al. (1988) will be considered, although it can also be related with one of the proposals in Hall et al. (1987). Given a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$, of a directional variable $\mathbf{X}$ with density f, the directional kernel density estimator is given by:

$$\hat{f}_h(\mathbf{x}) = \frac{c_{h,q}(L)}{n} \sum_{i=1}^{n} L\left(\frac{1 - \mathbf{x}^T\mathbf{X}_i}{h^2}\right), \quad \mathbf{x} \in \Omega_q, \tag{3.2}$$

where L is the directional kernel, $h > 0$ is the bandwidth parameter and $c_{h,q}(L)$ is a normalizing constant depending on the kernel L, the bandwidth h and the dimension q. The scalar product of two vectors, $\mathbf{x}$ and $\mathbf{y}$, is denoted by $\mathbf{x}^T\mathbf{y}$, where ^{T} is the transpose operator.

In this setting, directional kernels are not directional densities but functions of rapid decay. Therefore, to ensure that the resulting estimator is indeed a directional density, the normalizing constant $c_{h,q}(L)$ is needed. Specifically (see Bai et al. (1988)), the inverse of this normalizing constant for any $\mathbf{x} \in \Omega_q$ is given by

$$c_{h,q}(L)^{-1} = \int_{\Omega_q} L\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}\right) \omega_q(d\mathbf{y}) = h^q \lambda_{h,q}(L) \sim h^q \lambda_q(L), \tag{3.3}$$

with $\lambda_{h,q}(L) = \omega_{q-1} \int_0^{2h^{-2}} L(r) r^{\frac{q}{2}-1}(2 - rh^2)^{\frac{q}{2}-1}\, dr$ and $\lambda_q(L) = 2^{\frac{q}{2}-1}\omega_{q-1} \int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr$. The asymptotic behaviour of $\lambda_{h,q}(L)$ is established in Lemma 3.1 and the notation $a_n \sim b_n$ indicates that $\frac{a_n}{b_n} \to 1$ as $n \to \infty$ (see also Bai et al. (1988) and Zhao and Wu (2001)).

Properties of the directional kernel density estimator (3.2) have been analysed by Bai et al. (1988), who proved pointwise, uniform and $\mathcal{L}_1$-norm consistency. A central limit theorem for the integrated squared error of the estimator has been established by Zhao and Wu (2001), as well as the expression for the bias under some regularity conditions, stated below:

D1. Extend f from Ω_q to $\mathbb{R}^{q+1}\setminus\{\mathbf{0}\}$ by defining $f(\mathbf{x}) \equiv f\left(\mathbf{x}/||\mathbf{x}||\right)$ for all $\mathbf{x} \neq \mathbf{0}$, where $||\cdot||$ denotes the Euclidean norm. Assume that the gradient vector $\boldsymbol{\nabla} f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_{q+1}}\right)^T$ and the Hessian matrix $\boldsymbol{\mathcal{H}} f(\mathbf{x}) = \left(\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}\right)_{1 \leq i,j \leq q+1}$ exist, are continuous on $\mathbb{R}^{q+1}\setminus\{\mathbf{0}\}$ and square integrable.

D2. Assume that $L : [0, \infty) \to [0, \infty)$ is a bounded and Riemann integrable function such that

$$0 < \int_0^\infty L^k(r) r^{\frac{q}{2}-1}\, dr < \infty, \quad \forall q \geq 1, \text{ for } k = 1, 2.$$

D3. Assume that $h = h_n$ is a sequence of positive numbers such that $h_n \to 0$ and $nh_n^q \to \infty$ as $n \to \infty$.

Remark 3.2. *L must be a rapidly decreasing function, quite different from the bell-shaped kernels K involved in the linear estimator (3.1). To verify **D2**, L must decrease faster than any power function, since $\int_0^\infty r^\alpha r^{\frac{q}{2}-1}\, dr = \infty$, $\forall \alpha \in \mathbb{R}$, $\forall q \geq 1$.*

Lemma 2 in Zhao and Wu (2001) states that, under the previous conditions **D1**–**D3**, the expected value of the directional kernel density estimator in a point $\mathbf{x} \in \Omega_q$, is

$$\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right] = f(\mathbf{x}) + b_q(L)\Psi(f, \mathbf{x})h^2 + o\left(h^2\right),$$

where

$$\Psi(f, \mathbf{x}) = -\mathbf{x}^T \boldsymbol{\nabla} f(\mathbf{x}) + q^{-1}\left(\nabla^2 f(\mathbf{x}) - \mathbf{x}^T \boldsymbol{\mathcal{H}} f(\mathbf{x})\mathbf{x}\right), \tag{3.4}$$

$$b_q(L) = \int_0^\infty L(r) r^{\frac{q}{2}}\, dr \Big/ \int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr, \tag{3.5}$$

being $\nabla^2 f(\mathbf{x}) = \sum_{i=1}^{q+1} \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_i^2}$ the Laplacian of f. Note that the bias is of order $\mathcal{O}(h^2)$, but in (3.4), apart from the curvature of the target density which is captured by the Hessian matrix, a gradient vector also appears. On the other hand, the scaling constant $b_q(L)$ can be interpreted as a kind of moment of the directional kernel L. Note that, condition **D2** with $k = 1$ is needed for the bias computation. The same condition with $k = 2$ is required for deriving the pointwise variance of the estimator (3.2), which was also given by Hall et al. (1987) and Klemelä (2000).

Proposition 3.1. *Under conditions **D1**–**D3**, the variance of $\hat{f}_h(\mathbf{x})$ at $\mathbf{x} \in \Omega_q$ is given by*

$$\mathbb{V}\mathrm{ar}\left[\hat{f}_h(\mathbf{x})\right] = \frac{c_{h,q}(L)}{n} d_q(L) f(\mathbf{x}) + o\left((nh^q)^{-1}\right),$$

where

$$d_q(L) = \int_0^\infty L^2(r) r^{\frac{q}{2}-1}\, dr \Big/ \int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr.$$

Regarding the normalizing constant expression (3.3), the order of the variance is $\mathcal{O}\left((nh^q)^{-1}\right)$, where q is the dimension of the sphere. This order coincides with the corresponding one for a multivariate kernel density estimator in $\mathbb{R}^q$ (see Scott (1992)).

A popular choice for the directional kernel is $L(r) = e^{-r}$, $r \geq 0$, also known as the von Mises kernel due to its relation with the von Mises-Fisher distribution (see Watson (1983)). In a q-dimensional sphere, the von Mises model $\mathrm{vM}(\boldsymbol{\mu}, \kappa)$ has density

$$f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = C_q(\kappa) \exp\left\{\kappa \mathbf{x}^T \boldsymbol{\mu}\right\}, \quad C_q(\kappa) = \frac{\kappa^{\frac{q-1}{2}}}{(2\pi)^{\frac{q+1}{2}} \mathcal{I}_{\frac{q-1}{2}}(\kappa)}, \tag{3.6}$$

being $\boldsymbol{\mu} \in \Omega_q$ the directional mean and $\kappa \geq 0$ the concentration parameter around the mean. In Figure 3.1 (left plot), the contour plot of a spherical von Mises is shown. $\mathcal{I}_\nu$ is the modified Bessel function of order ν,

$$\mathcal{I}_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\pi^{1/2}\Gamma\left(\nu + \frac{1}{2}\right)} \int_{-1}^{1} (1-t^2)^{\nu - \frac{1}{2}} e^{zt}\, dt.$$

For the particular case of the target density being a q-dimensional von Mises $\text{vM}(\boldsymbol{\mu}, \kappa)$, the term (3.4) in the bias computation becomes:

$$\Psi\left(f_{\text{vM}}(\cdot; \boldsymbol{\mu}, \kappa), \mathbf{x}\right) = \kappa C_q(\kappa) e^{\kappa \mathbf{x}^T \boldsymbol{\mu}} \left(-\mathbf{x}^T \boldsymbol{\mu} + \kappa q^{-1} \left(1 - (\mathbf{x}^T \boldsymbol{\mu})^2\right)\right).$$

As $\kappa \to 0$, which means that the distribution is approaching a uniform model in the sphere, the previous term also tends to zero.

Considering the von Mises kernel in the directional estimator (3.2) allows for its interpretation as a mixture of von Mises-Fisher densities

$$\hat{f}_h(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} f_{\text{vM}}\left(\mathbf{x}; \mathbf{X}_i, 1/h^2\right), \tag{3.7}$$

where, for each von Mises component, the mean value is i-th observation $\mathbf{X}_i$ and the concentration is given by $\frac{1}{h^2}$, involving the smoothing parameter.

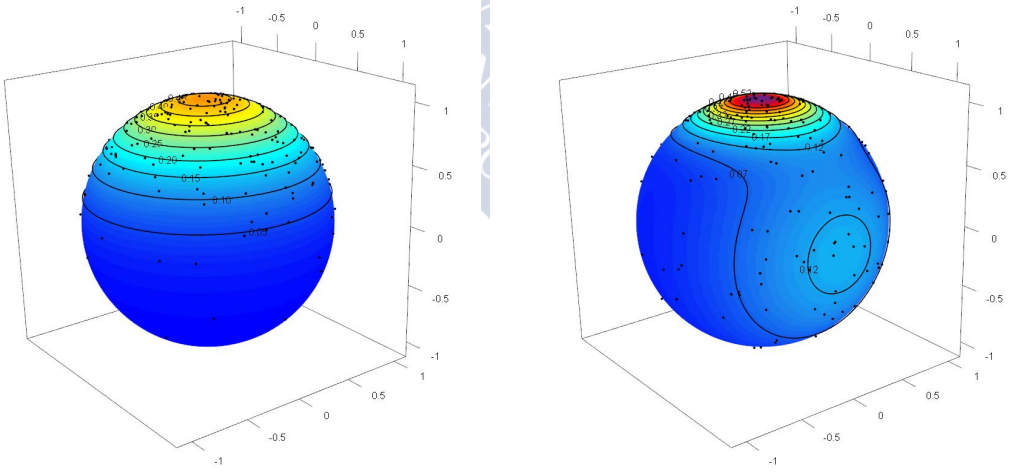


Figure 3.1: Left: contour plot of a von Mises density $\text{vM}(\boldsymbol{\mu}, \kappa)$, with $\boldsymbol{\mu} = (0, 0, 1)$ and $\kappa = 3$. Right: contour plot of the mixture of von Mises densities (3.14).

In addition, the normalizing constant (3.3) appearing in the construction of the directional kernel estimator (3.2) has a simple expression for a von Mises kernel, given by

$$c_{h,q}(L)^{-1} = \frac{2\pi^{\frac{q}{2}}}{\Gamma\left(\frac{q}{2}\right)} \int_{-1}^{1} \exp\left\{\frac{-1+t}{h^2}\right\} (1-t^2)^{\frac{q}{2}-1}\, dt = C_q(1/h^2)^{-1} e^{-1/h^2}. \tag{3.8}$$

For a general kernel, the asymptotic behaviour of $c_{h,q}(L)^{-1}$ was remarked in (3.3) and it can be specified for the von Mises kernel. In this case, (3.8) depends on $C_q(1/h^2)$, which involves

a Bessel function of order $(q-1)/2$. Applying a Taylor expansion for $\mathcal{I}_\nu$, it can be seen that $\mathcal{I}_\nu(z) = e^z\left(\frac{z^{-\frac{1}{2}}}{\sqrt{2\pi}} + \mathcal{O}(z^{-\frac{3}{2}})\right)$, $z \geq 0$ and $c_{h,q}(L)^{-1}$ presents also a simple form:

$$c_{h,q}(L)^{-1} = (2\pi)^{\frac{q+1}{2}}\, e^{-\frac{1}{h^2}} h^{q-1} e^{\frac{1}{h^2}} \left(\frac{h}{\sqrt{2\pi}} + \mathcal{O}\left(h^3\right)\right) = (2\pi)^{\frac{q}{2}}\, h^q + \mathcal{O}\left(h^{q+2}\right).$$

Finally, the other terms involved in bias and variance, namely $b_q(L)$ and $d_q(L)$, become

$$b_q(L) = \frac{q}{2}, \quad d_q(L) = 2^{-\frac{q}{2}+1}, \quad \forall q \geq 1$$

for the von Mises kernel.

3.2.2 Kernel density estimation for directional-linear data

Consider a directional-linear random variable, $(\mathbf{X}, Z)$ with support $\text{supp}(\mathbf{X}, Z) \subseteq \Omega_q \times \mathbb{R}$ and joint density f. For the simple case of circular data ($q = 1$), the support of the variable is the cylinder. Following the ideas in the previous section for the linear and directional cases, given a random sample $(\mathbf{X}_1, Z_1), \ldots, (\mathbf{X}_n, Z_n)$, the directional-linear kernel density estimator can be defined as:

$$\hat{f}_{h,g}(\mathbf{x}, z) = \frac{c_{h,q}(L)}{ng} \sum_{i=1}^{n} LK\left(\frac{1 - \mathbf{x}^T\mathbf{X}_i}{h^2}, \frac{z - Z_i}{g}\right), \quad (\mathbf{x}, z) \in \Omega_q \times \mathbb{R}, \tag{3.9}$$

where LK is a directional-linear kernel, g is the bandwidth parameter for the linear component, h the bandwidth parameter for the directional component and $c_{h,q}(L)$ is the normalizing constant for the directional part, defined in (3.3). For the sake of simplicity, a product kernel $LK(\cdot,\cdot) = L(\cdot) \times K(\cdot)$ will be considered throughout this paper. Although a product kernel formulation has been adopted, the results could be generalized for a directional-linear kernel, with the suitable modifications in the required conditions.

3.3 Main results

Before stating the main results, some notation will be introduced. The target directional-linear density will be denoted by f. The gradient vector and Hessian matrix of f, with respect to both components (directional and linear) are defined in this setting as:

$$\boldsymbol{\nabla} f(\mathbf{x}, z) = \left(\frac{\partial f(\mathbf{x}, z)}{\partial x_1}, \ldots, \frac{\partial f(\mathbf{x}, z)}{\partial x_{q+1}}, \frac{\partial f(\mathbf{x}, z)}{\partial z}\right)^T = \left(\boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}, z), \nabla_z f(\mathbf{x}, z)\right)^T,$$

$$\boldsymbol{\mathcal{H}} f(\mathbf{x}, z) = \left(\begin{array}{ccc|c} \frac{\partial^2 f(\mathbf{x},z)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\mathbf{x},z)}{\partial x_1 \partial x_{q+1}} & \frac{\partial^2 f(\mathbf{x},z)}{\partial x_1 \partial z} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial^2 f(\mathbf{x},z)}{\partial x_{q+1} \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x},z)}{\partial x_{q+1}^2} & \frac{\partial^2 f(\mathbf{x},z)}{\partial x_{q+1} \partial z} \\ \hline \frac{\partial^2 f(\mathbf{x},z)}{\partial z \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x},z)}{\partial z \partial x_{q+1}} & \frac{\partial^2 f(\mathbf{x},z)}{\partial z^2} \end{array}\right) = \left(\begin{array}{c|c} \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z) & \boldsymbol{\mathcal{H}}_{\mathbf{x},z} f(\mathbf{x}, z) \\ \hline \boldsymbol{\mathcal{H}}_{\mathbf{x},z} f(\mathbf{x}, z)^T & \mathcal{H}_z f(\mathbf{x}, z) \end{array}\right),$$

where subscripts $\mathbf{x}$ and z are used to denote the derivatives with respect to the directional and linear components, respectively. The Laplacian of f restricted to the directional component is denoted by $\nabla^2_{\mathbf{x}} f(\mathbf{x}, z) = \sum_{i=1}^{q+1} \frac{\partial^2 f(\mathbf{x},z)}{\partial \mathbf{x}_i^2}$. The following conditions will be required in order to prove the main results:

DL1. Extend f from $\Omega_q \times \mathbb{R}$ to $\mathbb{R}^{q+2}\backslash A$, $A = \{(\mathbf{x}, z) \in \mathbb{R}^{q+2} : \mathbf{x} = \mathbf{0}\}$, by defining $f(\mathbf{x}, z) \equiv f\left(\mathbf{x}/\left|\left|\mathbf{x}\right|\right|, z\right)$ for all $\mathbf{x} \neq \mathbf{0}$ and $z \in \mathbb{R}$, where $||\cdot||$ denotes the Euclidean norm. Assume that $\boldsymbol{\nabla} f(\mathbf{x}, z)$ and $\boldsymbol{\mathcal{H}} f(\mathbf{x}, z)$ exist, are continuous and square integrable.

DL2. Assume that the directional kernel L satisfies condition **D2** and the linear kernel K is a symmetric around zero and bounded linear density function with finite second order moment.

DL3. Assume that $h = h_n$ and $g = g_n$ are sequences of positive numbers such that $h_n \to 0$, $g_n \to 0$ and $nh_n^q g_n \to \infty$ as $n \to \infty$.

The next two results provide the expressions for the bias and the variance of the directional-linear kernel density estimator (3.9).

Proposition 3.2. *Under conditions **DL1**–**DL3**, the expected value of the directional-linear kernel density estimator (3.9) in a point $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$ is given by*

$$\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x}, z)\right] = f(\mathbf{x}, z) + b_q(L)\Psi_{\mathbf{x}}(f, \mathbf{x}, z)h^2 + \frac{1}{2}\mu_2(K)\mathcal{H}_z f(\mathbf{x}, z)g^2 + o\left(h^2 + g^2\right),$$

where

$$\Psi_{\mathbf{x}}(f, \mathbf{x}, z) = -\mathbf{x}^T \boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}, z) + q^{-1}\left(\nabla_{\mathbf{x}}^2 f(\mathbf{x}, z) - \mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z)\mathbf{x}\right).$$

Proposition 3.3. *Under conditions **DL1**–**DL3**, the variance for the directional-linear kernel density estimator (3.9) in a point $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$ is given by*

$$\mathbb{V}\mathrm{ar}\left[\hat{f}_{h,g}(\mathbf{x}, z)\right] = \frac{c_{h,q}(L)}{ng} R(K) d_q(L) f(\mathbf{x}, z) + o\left((nh^q g)^{-1}\right).$$

In view of the previous results, some comments must be done. Firstly, the effects of the directional and linear part can be clearly identified. For the bias, marginal contributions appear as two addends and also the remaining orders from each part are separated. For the variance, the terms corresponding to both parts can be also identified, although turning up in a product form. In addition, the respective orders for bias and variance are analogous to those ones obtained with a $(q+1)$-multivariate estimator in $\mathbb{R}^{q+1}$ (see Scott (1992)).

It can be also proved that the directional-linear kernel density estimator (3.9) is asymptotically normal, under the same conditions as those ones used for deriving the expected value and the variance, and a further smoothness property on the product kernel.

Theorem 3.1. *Under conditions **DL1**–**DL3**, if $\displaystyle\int_0^\infty \int_{\mathbb{R}} LK^{2+\delta}(r, v)\, r^{\frac{q}{2}-1}\, dv\, dr < \infty$ for some $\delta > 0$, then the directional-linear kernel density estimator (3.9) is asymptotically normal:*

$$\sqrt{nh^q g}\left(\hat{f}_{h,g}(\mathbf{x}, z) - f(\mathbf{x}, z) - \mathrm{ABias}\left[\hat{f}_{h,g}(\mathbf{x}, z)\right]\right) \xrightarrow{d} \mathcal{N}\left(0, R(K) d_q(L) f(\mathbf{x}, z)\right),$$

pointwise in $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$, where $\mathrm{ABias}\left[\hat{f}_{h,g}(\mathbf{x}, z)\right] = b_q(L)\Psi_{\mathbf{x}}(f, \mathbf{x}, z)h^2 + \frac{1}{2}\mu_2(K)\mathcal{H}_z f(\mathbf{x}, z)g^2$.

The smoothness condition on the directional-linear kernel is required in order to ensure Lyapunov's condition and obtain the asymptotic normal distribution. Again, the effect of the two parts can be identified in the previous equation, as well as in the rate of convergence of the estimator.

3.4 Error measurement and optimal bandwidth

The analysis of the performance of the kernel density estimator requires the specification of appropriate error criteria. Consider a generic kernel density estimator $\hat{f}$, which can be linear, directional or directional-linear. A global error measurement for quantifying the overall performance of this estimator is given by the MISE:

$$\text{MISE}\left[\hat{f}\right] = \int \mathbb{E}\left[(\hat{f}(u) - f(u))^2\right] du.$$

The MISE can be interpreted as a function of the bandwidth and its minimization yields an optimal bandwidth in the sense of the quadratic loss.

For the linear kernel density estimator (3.1) and under some regularity conditions (see Wand and Jones (1995)), the MISE is given by:

$$\text{MISE}\left[\hat{f}_g\right] = \frac{1}{4}\mu_2(K)^2\, R(f'')g^4 + (ng)^{-1}R(K) + o\left(g^4 + (ng)^{-1}\right).$$

The asymptotic version of the MISE, namely the AMISE, can be used to derive an optimal bandwidth that minimizes this error. This optimal bandwidth is given by

$$g_{\text{AMISE}} = \left[\frac{R(K)}{\mu_2(K)^2 R(f'')n}\right]^{\frac{1}{5}}.$$

Although the previous expression does not provide a bandwidth value in practice, given that it depends on the curvature of the target density $R(f'')$, some interesting issues should be noticed. For instance, the order of the asymptotic optimal bandwidth is $\mathcal{O}(n^{-1/5})$. Also, this result is the starting point of more sophisticated bandwidth selectors such as the ones given by Sheather and Jones (1991) and Cao (1993). A comparison of the performance of different bandwidth selectors can be found in Cao et al. (1994), whereas Jones et al. (1996) provides a review on bandwidth selection methods.

3.4.1 MISE for directional and directional-linear kernel density estimators

In the previous sections, the bias and variance for the directional kernel estimator (see Zhao and Wu (2001) for the bias and Proposition 3.1 for the variance) and for the directional-linear kernel estimator (Propositions 3.2 and 3.3) were obtained. Hence, it is straightforward to get the MISE for these estimators.

Proposition 3.4. *Under conditions **D1**–**D3**, the MISE for the directional kernel density estimator (3.2) is given by*

$$\text{MISE}\left[\hat{f}_h\right] = b_q(L)^2 \int_{\Omega_q} \Psi(f,\mathbf{x})^2\, \omega_q(d\mathbf{x})h^4 + \frac{c_{h,q}(L)}{n} d_q(L) + o\left(h^4 + (nh^q)^{-1}\right).$$

Following Wand and Jones (1995), $\text{MISE}[\hat{f}_h] = \text{AMISE}[\hat{f}_h] + o\left(h^4 + (nh^q)^{-1}\right)$, providing $\text{AMISE}[\hat{f}_h]$ a suitable large sample approximation that allows for the computation of an optimal bandwidth with closed expression, minimizing this asymptotic error criterion.

Corollary 3.1. *The AMISE optimal bandwidth for the directional kernel density estimator (3.2) is given by*

$$h_{\mathrm{AMISE}} = \left[\frac{q d_q(L)}{4 b_q(L)^2 \lambda_q(L) R(\Psi(f,\cdot)) n}\right]^{\frac{1}{4+q}},$$

where $R(\Psi(f,\cdot)) = \int_{\Omega_q} \Psi(f,\mathbf{x})^2\, \omega_q(d\mathbf{x})$ *and* $\lambda_q(L) = 2^{\frac{q}{2}-1}\omega_{q-1}\int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr$.

Expressions for MISE and AMISE can be also derived for the directional-linear estimator. In order to simplify the notation, let denote $I[\phi] = \int_{\Omega_q\times\mathbb{R}} \phi(\mathbf{x},z)\, dz\, \omega_q(d\mathbf{x})$, for a function $\phi : \Omega_q \times \mathbb{R} \to \mathbb{R}$.

Proposition 3.5. *Under conditions* ***DL1***–***DL3****, the MISE for the directional-linear kernel density estimator (3.9) is given by*

$$\begin{aligned}\mathrm{MISE}\left[\hat{f}_{h,g}\right] = {} & b_q(L)^2 I\left[\Psi_{\mathbf{x}}(f,\cdot,\cdot)^2\right] h^4 + \frac{1}{4}\mu_2(K)^2 I\left[\mathcal{H}_z f(\cdot,\cdot)^2\right] g^4 \\ & + b_q(L)\mu_2(K) I\left[\Psi_{\mathbf{x}}(f,\cdot,\cdot)\mathcal{H}_z f(\cdot,\cdot)\right] h^2 g^2 + \frac{c_{h,q}(L)}{ng} d_q(L) R(K) \\ & + o\left(h^4 + g^4 + (nh^q g)^{-1}\right).\end{aligned}$$

Unfortunately, it is not straightforward to derive a full closed expression for the optimal pair of bandwidths $(h,g)_{\mathrm{AMISE}}$, although it is possible to compute them by numerical optimization. However, such a closed expression can be obtained for the particular case $q = 1$, where the circular and linear bandwidths can be considered as proportional.

Corollary 3.2. *Consider the parametrization* $g = \beta h$*. The optimal AMISE pair of bandwidths* $(h,g)_{\mathrm{AMISE}} = (h_{\mathrm{AMISE}}, \beta h_{\mathrm{AMISE}})$ *can be obtained from*

$$h_{\mathrm{AMISE}} = \left[\frac{(q+1) d_q(L) R(K)}{4\beta\lambda_q(L) R\left(b_q(L)\Psi_{\mathbf{x}}(f,\cdot,\cdot) + \frac{\beta^2}{2}\mu_2(K)\mathcal{H}_z f(\cdot,\cdot)\right) n}\right]^{\frac{1}{5+q}},$$

where $R\Big(b_q(L)\Psi_{\mathbf{x}}(f,\cdot,\cdot) + \frac{\beta^2}{2}\mu_2(K)\mathcal{H}_z f(\cdot,\cdot)\Big) = \int_{\Omega_q\times\mathbb{R}} \left(b_q(L)\Psi_{\mathbf{x}}(f,\mathbf{x},z) + \frac{\beta^2}{2}\mu_2(K)\mathcal{H}_z f(\mathbf{x},z)\right)^2 dz\, \omega_q(d\mathbf{x})$ *and* $\lambda_q(L)$ *is defined as in the previous corollary. For the circular-linear data case* $(q = 1)$*, the parameter* β *is given by:*

$$\beta = \left(\frac{\frac{1}{4}\mu_2(K)^2 I\left[\mathcal{H}_z f(\cdot,\cdot)^2\right]}{b_q(L)^2 I\left[\Psi_{\mathbf{x}}(f,\cdot,\cdot)^2\right]}\right)^{\frac{1}{4}}.$$

Despite a formal way for deriving the orders of the AMISE bandwidths has not been derived, a quite plausible conjecture is that for $q > 1$, $(h,g)_{\mathrm{AMISE}} = (\mathcal{O}(n^{-1/(4+q)}), \mathcal{O}(n^{-1/5}))$ or, equivalently, that $\beta = \beta_n = \mathcal{O}(n^{-(q-1)/(5(4+q))})$. Indeed, this is satisfied for $q = 1$.

Finally, it is interesting to note that considering $g = \beta h$, a single bandwidth for the kernel estimator (3.9) is required, having the optimal bandwidth under this formulation order $\mathcal{O}(n^{-1/(5+q)})$. This coincides with the order of the kernel linear estimator in $\mathbb{R}^p$, with $p = \dim(\Omega_q \times \mathbb{R}) = q+1$.

3.4.2 Some exact MISE calculations for mixture distributions

Closed expressions for the MISE for the directional and directional-linear estimators can be obtained for some particular distribution models, and they will be derived in this section. In the linear setting, Marron and Wand (1992) obtained a closed expression for the MISE of (3.1) if the kernel K is a normal density and the underlying model is a mixture of normal distributions. Specifically, the density of an r-mixture of normal distributions with respective means m_j and variances σ_j^2, for $j = 1, \ldots, r$ is given by

$$f_r(z) = \sum_{j=1}^{r} p_j \phi_{\sigma_j}(z - m_j), \quad \sum_{j=1}^{r} p_j = 1, \quad p_j \geq 0,$$

where p_j, $j = 1, \ldots, r$ denote the mixture weights and ϕ_σ is the density of a normal with zero mean and variance σ^2, *i.e.*, $\phi_\sigma(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{z^2}{2\sigma^2}\right\}$. Marron and Wand (1992) showed that the exact MISE of the linear kernel estimator is

$$\text{MISE}\left[\hat{f}_g\right] = \left(2\pi^{\frac{1}{2}} g n\right)^{-1} + \mathbf{p}^T \left[(1 - n^{-1})\mathbf{\Omega_2}(g) - 2\mathbf{\Omega_1}(g) + \mathbf{\Omega_0}(g)\right] \mathbf{p}, \tag{3.10}$$

where $\mathbf{p} = (p_1, \ldots, p_r)^T$ and $\mathbf{\Omega_a}(g)$ are matrices with entries $\mathbf{\Omega_a}(g) = (\phi_{\sigma_a}(m_i - m_j))_{ij}$, $\sigma_a = (ag^2 + \sigma_i^2 + \sigma_j^2)^{\frac{1}{2}}$, for $a = 0, 1, 2$.

Similar results can be obtained for the directional and directional-linear estimators, when considering mixtures of von Mises for the directional case, and mixtures of von Mises and normals for the directional-linear scenario (see Figure 3.2 for some examples). For the directional setting, an r-mixture of von Mises with means $\boldsymbol{\mu}_j$ and concentration parameters κ_j, for $j = 1, \ldots, r$ is given by

$$f_r(\mathbf{x}) = \sum_{j=1}^{r} p_j f_{\text{vM}}(\mathbf{x}; \boldsymbol{\mu}_j, \kappa_j), \quad \sum_{j=1}^{r} p_j = 1, \quad p_j \geq 0. \tag{3.11}$$

Consider a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$, of a directional variable $\mathbf{X}$ with density f_r (see Figure 3.1, right plot). The following result gives a closed expression for the MISE of the directional kernel estimator.

Proposition 3.6. *Let f_r be the density of an r-mixture of directional von Mises (3.11). The exact MISE of the directional kernel estimator (3.2), obtained from a random sample of size n, with von Mises kernel $L(r) = e^{-r}$ is*

$$\text{MISE}\left[\hat{f}_h\right] = (D_q(h)n)^{-1} + \mathbf{p}^T \left[(1 - n^{-1})\mathbf{\Psi_2}(h) - 2\mathbf{\Psi_1}(h) + \mathbf{\Psi_0}(h)\right] \mathbf{p}, \tag{3.12}$$

where $\mathbf{p} = (p_1, \ldots, p_r)^T$ and $D_q(h) = C_q\left(1/h^2\right)^2 C_q\left(2/h^2\right)^{-1}$. The matrices $\mathbf{\Psi_a}(h)$, $a = 0, 1, 2$ have entries:

$$\mathbf{\Psi_0}(h) = \left(\frac{C_q(\kappa_i) C_q(\kappa_j)}{C_q(||\kappa_i \boldsymbol{\mu}_i + \kappa_j \boldsymbol{\mu}_j||)}\right)_{ij},$$

$$\mathbf{\Psi_1}(h) = C_q(1/h^2) \left(C_q(\kappa_i) C_q(\kappa_j) \int_{\Omega_q} \frac{e^{\kappa_j \mathbf{x}^T \boldsymbol{\mu}_j}}{C_q\left(||\mathbf{x}/h^2 + \kappa_i \boldsymbol{\mu}_i||\right)} \omega_q(d\mathbf{x})\right)_{ij},$$

$$\mathbf{\Psi_2}(h) = C_q(1/h^2)^2 \left(C_q(\kappa_i) C_q(\kappa_j) \int_{\Omega_q} \left[C_q(||\mathbf{x}/h^2 + \kappa_i \boldsymbol{\mu}_i||) C_q(||\mathbf{x}/h^2 + \kappa_j \boldsymbol{\mu}_j||) \right]^{-1} \omega_q(d\mathbf{x}) \right)_{ij},$$

where C_q is defined in equation (3.6).

The matrices involved in (3.12) are not as simple as the ones for the linear case, due to the convolution properties of the von Mises density. For practical implementation of the exact MISE, it should be noticed that matrices $\mathbf{\Psi_2}(h)$ and $\mathbf{\Psi_1}(h)$ can be evaluated using numerical integration in q-spherical coordinates. For clarity purposes, constants $C_q(\kappa_i)$ are included inside matrices $\mathbf{\Psi_2}(h)$, $\mathbf{\Psi_1}(h)$ and $\mathbf{\Psi_0}(h)$ but it is computationally more efficient to consider them within the weights, that is, take $\mathbf{p} = (p_1 C_q(\kappa_1), \ldots, p_r C_q(\kappa_r))$.

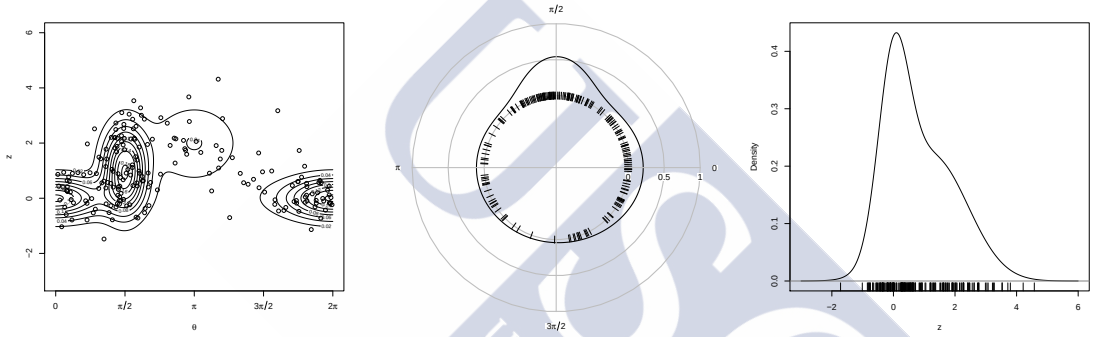


Figure 3.2: From left to right: circular-linear mixture (3.15) and corresponding circular and linear marginal densities, respectively. Random samples of size $n = 200$ are drawn.

From Proposition 3.6, it is easy to derive an analogous result for the case of a r-mixture of directional-linear independent von Mises and normals:

$$f_r(\mathbf{x}, z) = \sum_{j=1}^{r} p_j f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}_j, \kappa_j) \times \phi_{\sigma_j}(z - m_j), \quad \sum_{j=1}^{r} p_j = 1, \quad p_j \geq 0. \tag{3.13}$$

Proposition 3.7. *Let f_r be the density of an r-mixture of directional-linear independent von Mises and normals densities given in (3.13). For a random sample of size n, the exact MISE of the directional-linear kernel density estimator (3.9) with von Mises-normal kernel $LK(r,t) = e^{-r} \times \phi_1(t)$ is*

$$\begin{aligned} \mathrm{MISE}\left[\hat{f}_{h,g}\right] = & \left(D_q(h) 2\pi^{\frac{1}{2}} gn \right)^{-1} \\ & + \mathbf{p}^T \left[(1 - n^{-1}) \mathbf{\Psi_2}(h) \circ \mathbf{\Omega_2}(g) - 2\mathbf{\Psi_1}(h) \circ \mathbf{\Omega_1}(g) + \mathbf{\Psi_0}(h) \circ \mathbf{\Omega_0}(g) \right] \mathbf{p}, \end{aligned}$$

where $\circ$ denotes the Hadamard product between matrices and the involved terms are defined as in Proposition 3.6 and equation (3.10).

Once the exact MISE and the AMISE for mixtures of von Mises and normals are derived, it is possible to compare these two error criteria. To that end, let consider the following directional mixture

$$\frac{2}{5}\mathrm{vM}\left((1, \mathbf{0}_q)\right), 2) + \frac{2}{5}\mathrm{vM}\left((\mathbf{0}_q, 1), 10\right) + \frac{1}{5}\mathrm{vM}\left((-1, \mathbf{0}_q), 2\right), \tag{3.14}$$

where $\mathbf{0}_q$ represents a vector of q zeros, and the directional-linear mixture

$$\frac{2}{5}\mathcal{N}\left(0,\frac{1}{4}\right)\times \mathrm{vM}\left((1,\mathbf{0}_q)\right),2)+\frac{2}{5}\mathcal{N}\left(1,1\right)\times \mathrm{vM}\left((\mathbf{0}_q,1),10\right) + \frac{1}{5}\mathcal{N}\left(2,1\right)\times \mathrm{vM}\left((-1,\mathbf{0}_q),2\right). \tag{3.15}$$

Figure 3.3 shows the comparison between the exact and asymptotic MISE for the linear, circular and spherical case. As first noted by Marron and Wand (1992) for the linear estimator, there exist significative differences between these two errors, being the most remarkable one the rapid growth of the AMISE with respect to the MISE for larger values of the bandwidth. This effect is due to the fact that, for a general bandwidth h, $\lim_{h\to\infty}\mathrm{AMISE}\left[\hat{f}_h\right]=\infty$ since $\mathrm{AMISE}\left[\hat{f}_h\right]$ is proportional to h^4, whereas the MISE level offs at $\lim_{h\to\infty}\mathrm{MISE}\left[\hat{f}_h\right]=\int_{\Omega_q} f(\mathbf{x})^2\,\omega_q(d\mathbf{x})$. Besides, for the directional case, this effect seems to be augmented probably because of a scale effect in the bandwidths, in the sense that the support of the directional variables is bounded, which is not the case for the linear ones considered. However, although the AMISE and MISE curves differ significantly, the corresponding optimal bandwidths get closer for increasing sample sizes.

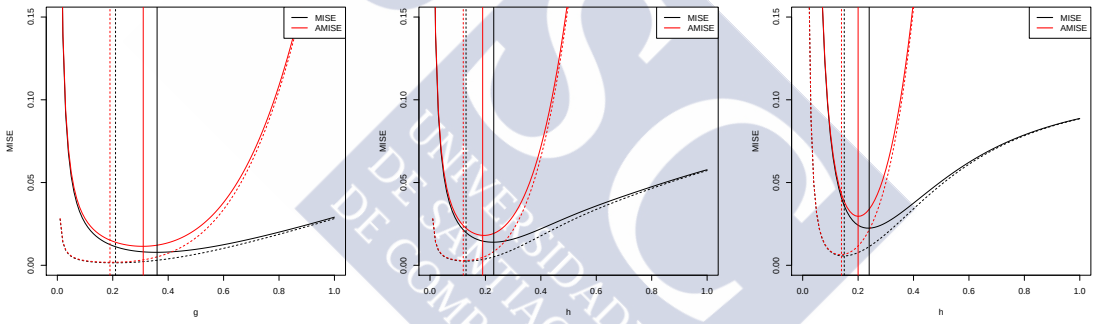


Figure 3.3: From left to right: exact MISE and AMISE for the linear mixture $\frac{2}{5}\mathcal{N}\left(0,\frac{1}{4}\right)+\frac{2}{5}\mathcal{N}\left(1,1\right)+\frac{1}{5}\mathcal{N}\left(2,1\right)$ and the circular and spherical mixtures (3.14), for a range of bandwidths between 0 and 1. The black curves are for the MISE, whereas the red ones are for the AMISE. Solid curves correspond to $n=100$ and dotted to $n=1000$. Vertical lines represent the bandwidth values minimizing each curve.

Figure 3.4 contains the contourplots of the exact and asymptotic MISE for the circular-linear and spherical-linear cases. The conclusions are more or less the same as for Figure 3.3: the asymptotic MISE grows rather quickly than the exact MISE for large values of h or g. On the other hand, the contour lines of both surfaces are quite close for small values of the bandwidths and the optimal bandwidths also get closer for larger sample sizes.

As an immediate application of Propositions 3.6 and 3.7, a bootstrap version of the MISE for the directional and directional-linear estimators can be derived. The bootstrap MISE is an estimator of the true MISE obtained by considering a smooth bootstrap resampling scheme, which will be briefly detailed. In the linear case, the bootstrap MISE is given by

$$\mathrm{MISE}^*_{g_P}\left[\hat{f}_g\right]=\mathbb{E}^*\left[\int_{\mathbb{R}}\left(\hat{f}^*_g(z)-\hat{f}_{g_P}(z)\right)^2 dz\right],$$

where $\hat{f}^*_g(z)=\frac{1}{ng}\sum_{i=1}^n K\left(\frac{z-Z^*_i}{g}\right)$, being the sample $Z^*_1,\ldots,Z^*_n$ distributed as $\hat{f}_{g_P}$. In this case, g_P is a pilot bandwidth and the expectation $\mathbb{E}^*$ is taken with respect to the density estimator $\hat{f}_{g_P}$.

For the linear case, Cao (1993) derived an exact closed expression for $\mathrm{MISE}^*_{g_P}[\hat{f}_g]$ that actually avoids the needing of resampling and obtained a bandwidth that minimizes the bootstrap MISE by previously computing a suitable pilot bandwidth g_P.

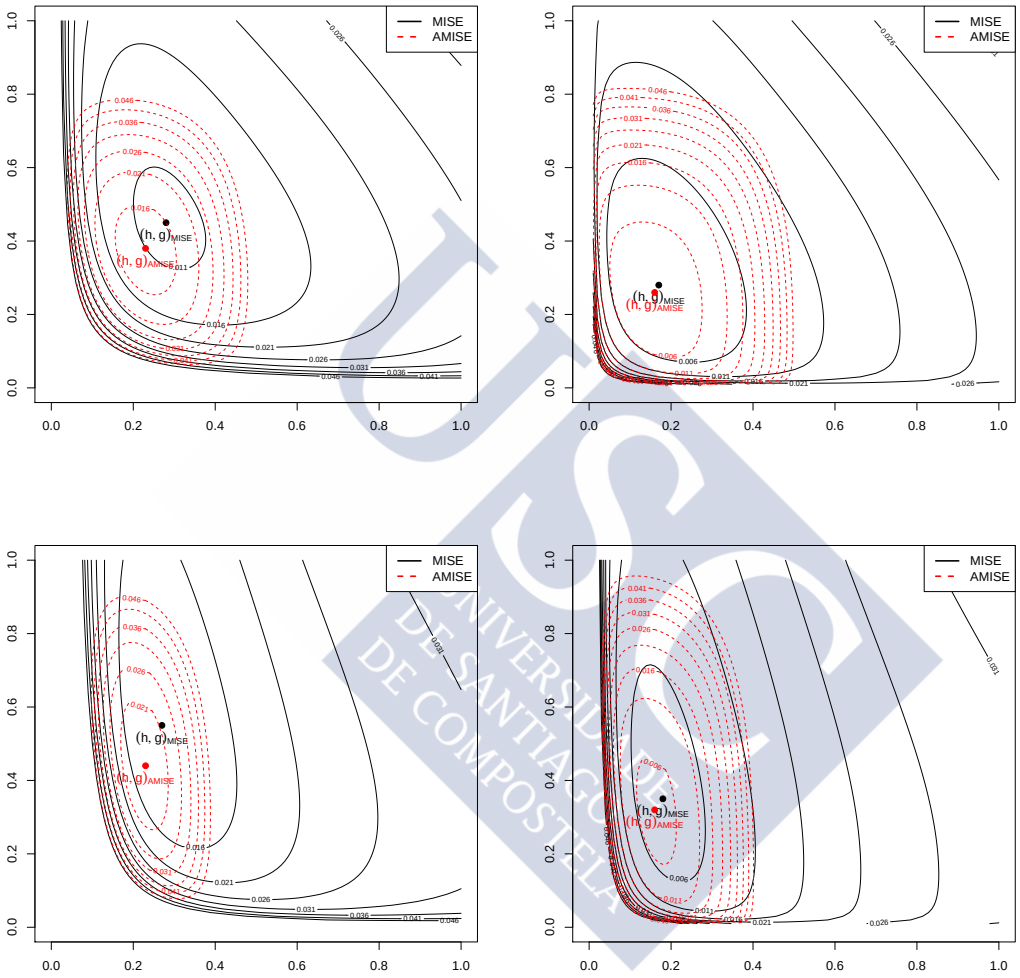


Figure 3.4: Upper plot, from left to right: exact MISE versus AMISE for the circular-linear mixture (3.15) for $n = 100$ and $n = 1000$. Lower plot, from left to right: spherical-linear mixture (3.15) for $n = 100$ and $n = 1000$. The solid curves are for the MISE, where the dashed ones are for the AMISE. The pairs of bandwidths that minimizes each surface error are denoted by $(h, g)_{\mathrm{MISE}}$ and by $(h, g)_{\mathrm{AMISE}}$.

The following two results show the bootstrap MISE expressions for the estimators (3.2) and (3.9) in the case where the kernels are von Mises and normals. As in the linear case, no resampling is needed for computing the bootstrap MISE. These bootstrap versions of the error provide an overall summary of the estimator behaviour, with no restriction on the underlying densities, as long as von Mises and normal kernels are considered. In addition, the following results could be used to derive a bandwidth selector, but it will depend on the selection of pilot bandwidths for both components, which is not an easy problem.

Corollary 3.3. *The bootstrap MISE for directional data, given a sample of length n, the von Mises kernel $L(r) = e^{-r}$ and a pilot bandwidth h_P, is:*

$$\mathrm{MISE}^*_{h_P}\left[\hat{f}_h\right] = (D_q(h)n)^{-1} + n^{-2}\mathbf{1}^T\left[(1-n^{-1})\mathbf{\Psi}^*_{\mathbf{2}}(h) - 2\mathbf{\Psi}^*_{\mathbf{1}}(h) + \mathbf{\Psi}^*_{\mathbf{0}}(h)\right]\mathbf{1},$$

where the matrices $\mathbf{\Psi}^*_{\mathbf{a}}(h)$, $a = 0, 1, 2$ have the same entries as $\mathbf{\Psi}_{\mathbf{a}}(h)$ but with $\kappa_i = 1/h_P^2$ and $\boldsymbol{\mu}_i = \mathbf{X}_i$ for $i = 1, \ldots, n$.

Remark 3.3. *The particular case where $q = 1$ and $h_P = h$, Corollary 3.3 corresponds to the expression of the bootstrap MISE given in Di Marzio et al. (2011).*

Corollary 3.4. *The bootstrap MISE for directional-linear data, given a sample of length n, the von Mises-normal kernel $LK(r,t) = e^{-r} \times \phi_1(t)$ and a pair of pilot bandwidths (h_P, g_P), is:*

$$\begin{aligned}\mathrm{MISE}^*_{h_P,g_P}\left[\hat{f}_{h,g}\right] = &\left(D_q(h)2\pi^{\frac{1}{2}}gn\right)^{-1} \\ &+ n^{-2}\mathbf{1}^T\left[(1-n^{-1})\mathbf{\Psi}^*_{\mathbf{2}}(h)\circ\mathbf{\Omega}^*_{\mathbf{2}}(g) - 2\mathbf{\Psi}^*_{\mathbf{1}}(h)\circ\mathbf{\Omega}^*_{\mathbf{1}}(g) + \mathbf{\Psi}_{\mathbf{0}}{}^*(h)\circ\mathbf{\Omega}^*_{\mathbf{0}}(g)\right]\mathbf{1},\end{aligned}$$

where the matrices $\mathbf{\Psi}^*_{\mathbf{a}}(h)$ and $\mathbf{\Omega}^*_{\mathbf{a}}(g)$, $a = 0, 1, 2$ have the same entries as $\mathbf{\Psi}_{\mathbf{a}}(h)$ and $\mathbf{\Omega}_{\mathbf{a}}(g)$ but with $\kappa_i = 1/h_P^2$, $\boldsymbol{\mu}_i = \mathbf{X}_i$, $m_i = Z_i$ and $\sigma_i = g_P$ for $i = 1, \ldots, n$.

3.5 Conclusions

A kernel density estimator for directional-linear data is proposed. Bias, variance and asymptotic normality of the estimator are derived, as well as expressions for the MISE and AMISE. For the particular case of mixtures of von Mises, for directional data, and mixtures of von Mises and normals, in the directional-linear case, the exact expressions for the MISE are obtained, which enables the comparison with their asymptotic versions.

Undoubtedly, one of the main issues in kernel estimation is the appropriate selection of the bandwidth parameter. Although an optimal pair of bandwidths in the AMISE sense has been derived, further research must be done in order to obtain a bandwidth selection method that could be applied in practice. This problem extends somehow to the directional setting, where (likelihood and least squares) cross-validation methods seem to be the available procedures. However, the exact MISE computations open a route to develop bandwidth selectors, for instance, following the ideas in Oliveira et al. (2012). In fact, a bootstrap version for the MISE when assuming that the underlying mode is a mixture allows for the derivation of bootstrap bandwidths, as in Cao (1993) for the linear case.

A straightforward extension of the proposed estimator can be found in the directional-multidimensional setting, considering a multidimensional random variable. In this case, the linear part of the estimator should be properly adapted including a multidimensional kernel and possibly a bandwidth matrix.

Acknowledgements

The authors acknowledge the support of Project MTM2008–03010, from the Spanish Ministry of Science and Innovation, Project 10MDS207015PR from Dirección Xeral de I+D, Xunta de

Galicia and IAP network StUDyS, from Belgian Science Policy. Work of E. García-Portugués has been supported by FPU grant AP2010–0957 from the Spanish Ministry of Education. The authors also acknowledge the suggestions by two anonymous referees that helped improving this paper.

3.A Some technical lemmas

Some technical lemmas that will be used along the proofs of the main results are introduced in this section. To begin with, Lemma 3.1 establishes the asymptotic behaviour of $\lambda_{h,q}(L)$ in (3.3). With the aim of clarifying the computation of the integrals in the proofs of the main results, Lemma 3.2 details a change of variables in Ω_q, whereas Lemma 3.3 is used to simplify integrals in Ω_q. Lemma 3.4 shows some of the constants introduced along the work for the case where the kernel is von Mises and, finally, Lemma 3.5 states the Lemma 2 of Zhao and Wu (2001).

Detailed proofs of these lemmas can be found in Appendix 3.C. This appendix also includes a rebuild of the proof of the Lemma 3.5, using the same techniques as for the other results, which presents some differences from the original proof.

Lemma 3.1. *Under condition **D2**, the limit of* $\lambda_{h,q}(L) = \omega_{q-1}\int_0^{2h^{-2}} L(r)r^{\frac{q}{2}-1}(2-rh^2)^{\frac{q}{2}-1}\,dr$, *when* $h \to 0$, *is*

$$\lim_{h\to 0}\lambda_{h,q}(L) = \lambda_q(L) = 2^{\frac{q}{2}-1}\omega_{q-1}\int_0^{\infty} L(r)r^{\frac{q}{2}-1}\,dr, \tag{3.16}$$

where ω_q *is the surface area of* Ω_q, *for* $q \geq 1$.

Lemma 3.2 (A change of variables in Ω_q)**.** *Let* f *be a function defined in* Ω_q *and* $\mathbf{y} \in \Omega_q$ *a fixed point. The integral* $\int_{\Omega_q} f(\mathbf{x})\,\omega_q(d\mathbf{x})$ *can be expressed in one of the following equivalent integrals:*

$$\int_{\Omega_q} f(\mathbf{x})\,\omega_q(d\mathbf{x}) = \int_{-1}^{1}\int_{\Omega_{q-1}} f\left(t,(1-t^2)^{\frac{1}{2}}\boldsymbol{\xi}\right)(1-t^2)^{\frac{q}{2}-1}\,\omega_{q-1}(d\boldsymbol{\xi})\,dt \tag{3.17}$$

$$= \int_{-1}^{1}\int_{\Omega_{q-1}} f\left(t\mathbf{y}+(1-t^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{y}}\boldsymbol{\xi}\right)(1-t^2)^{\frac{q}{2}-1}\,\omega_{q-1}(d\boldsymbol{\xi})\,dt, \tag{3.18}$$

where $\mathbf{B}_{\mathbf{y}} = (\mathbf{b}_1,\ldots,\mathbf{b}_q)_{(q+1)\times q}$ *is the semi-orthonormal matrix (*$\mathbf{B}_{\mathbf{y}}^T\mathbf{B}_{\mathbf{y}} = \mathbf{I}_q$ *and* $\mathbf{B}_{\mathbf{y}}\mathbf{B}_{\mathbf{y}}^T = \mathbf{I}_{q+1} - \mathbf{y}\mathbf{y}^T$*) resulting from the completion of* $\mathbf{y}$ *to the orthonormal basis* $\{\mathbf{y},\mathbf{b}_1,\ldots,\mathbf{b}_q\}$.

Lemma 3.3. *Consider* $\mathbf{x} \in \Omega_q$, *a point in the* q*-dimensional sphere with entries* $(x_1,\ldots,x_{q+1})$. *For all* $i,j = 1,\ldots,q+1$, *it holds that*

$$\int_{\Omega_q} x_i\,\omega_q(d\mathbf{x}) = 0, \quad \int_{\Omega_q} x_i x_j\,\omega_q(d\mathbf{x}) = \begin{cases} 0, & i \neq j,\\ \frac{\omega_q}{q+1}, & i = j,\end{cases}$$

where ω_q *is the surface area of* Ω_q, *for* $q \geq 1$.

Lemma 3.4. *For the von Mises kernel, i.e.,* $L(r) = e^{-r}$, $r \geq 0$,

$$c_{h,q}(L) = e^{1/h^2}h^{q-1}(2\pi)^{\frac{q+1}{2}}\mathcal{I}_{\frac{q-1}{2}}(1/h^2), \quad \lambda_q(L) = (2\pi)^{\frac{q}{2}}, \quad b_q(L) = \frac{q}{2}, \quad d_q(L) = 2^{-\frac{q}{2}}.$$

Lemma 3.5 (Lemma 2 in Zhao and Wu (2001))**.** *Under the conditions **D1**–**D3**, the expected value of the directional kernel density estimator in a point* $\mathbf{x} \in \Omega_q$, *is*

$$\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right] = f(\mathbf{x}) + b_q(L)\Psi(f,\mathbf{x})h^2 + o\left(h^2\right),$$

where $\Psi(f,\mathbf{x})$ *and* $b_q(L)$ *are given in (3.4) and (3.5), respectively.*

3.B Proofs of the main results

Proof of Proposition 3.1. The variance can be decomposed in two terms as follows:

$$\mathbb{V}\mathrm{ar}\left[\hat{f}_h(\mathbf{x})\right] = \frac{c_{h,q}(L)^2}{n}\mathbb{E}\left[L^2\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right] - n^{-1}\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]^2, \tag{3.19}$$

where the calculus of the first term is quite similar to the calculus of the bias given in Lemma 3.5 and the second is given by the same result.

Therefore, analogously to the equation (3.47) of Lemma 3.5, the first addend can be expressed as

$$\frac{c_{h,q}(L)^2}{n}h^q\int_0^{2h^{-2}} L^2(r)r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}\int_{\Omega_{q-1}} f\left(\mathbf{x}+\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right)\,\omega_{q-1}(d\boldsymbol{\xi})\,dr, \tag{3.20}$$

just replacing the kernel L by the squared kernel L^2 and where $\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}} = -rh^2\mathbf{x} + h\left[r(2-h^2r)\right]^{\frac{1}{2}}$ $\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi} \in \Omega_q$ with $\mathbf{B}_{\mathbf{x}}$ defined as in Lemma 3.2. By condition **D1**, the Taylor expansion of f at $\mathbf{x}$ is

$$f(\mathbf{x}+\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}) - f(\mathbf{x}) = \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T\boldsymbol{\nabla}f(\mathbf{x}) + \frac{1}{2}\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}} + o\left(\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right).$$

Hence,

$$\begin{aligned}
(3.20) = \frac{c_{h,q}(L)^2}{n}h^q\int_0^{2h^{-2}} L^2(r)r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}\Bigg\{& f(\mathbf{x}) - rh^2\omega_{q-1}\mathbf{x}^T\boldsymbol{\nabla}f(\mathbf{x}) \\
&+ \frac{r^2h^4\omega_{q-1}}{2}\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x} + \frac{h^2r(2-h^2r)\omega_{q-1}}{2q}\left(\nabla^2 f(\mathbf{x}) - \mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\right) + r\omega_{q-1}o\left(h^2\right)\Bigg\}dr \\
= \frac{c_{h,q}(L)}{n}\Bigg\{&\omega_{q-1}\left[\int_0^{2h^{-2}} c_{h,q}(L)h^qL^2(r)r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}\,dr\right]f(\mathbf{x}) \\
&- h^2\omega_{q-1}\left[\int_0^{2h^{-2}} c_{h,q}(L)h^qL^2(r)r^{\frac{q}{2}}(2-h^2r)^{\frac{q}{2}-1}\,dr\right]\mathbf{x}^T\boldsymbol{\nabla}f(\mathbf{x}) \\
&+ \frac{h^4\omega_{q-1}}{2}\left[\int_0^{2h^{-2}} c_{h,q}(L)h^qL^2(r)r^{\frac{q}{2}+1}(2-h^2r)^{\frac{q}{2}-1}\,dr\right]\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x} \\
&+ \frac{h^2\omega_{q-1}}{2}\left[\int_0^{2h^{-2}} c_{h,q}(L)h^qL^2(r)r^{\frac{q}{2}}(2-h^2r)^{\frac{q}{2}}\,dr\right]q^{-1}\left(\nabla^2 f(\mathbf{x}) - \mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\right) \\
&+\omega_{q-1}\left[\int_0^{2h^{-2}} c_{h,q}(L)h^qL^2(r)r^{\frac{q}{2}}(2-h^2r)^{\frac{q}{2}-1}\,dr\right]o\left(h^2\right)\Bigg\}.
\end{aligned} \tag{3.21}$$

The integrals in (3.21) can be simplified. For that purpose, define for $h > 0$ and indices $i = -1, 0, 1$, $j = 0, 1$ the following function:

$$\phi_{h,i,j}(r) = c_{h,q}(L)h^qL^2(r)r^{\frac{q}{2}+i}(2-h^2r)^{\frac{q}{2}-j}\mathbf{1}_{[0,2h^{-2})}(r), \quad r \in [0,\infty).$$

As $n \to \infty$, the bandwidth $h \to 0$ and the limit of $\phi_{h,i,j}$ is given by

$$\phi_{i,j}(r) = \lim_{h\to 0}\phi_{h,i,j}(r) = \lambda_q(L)^{-1}L^2(r)r^{\frac{q}{2}+i}2^{\frac{q}{2}-j}\mathbf{1}_{[0,\infty)}(r).$$

Applying the Dominated Convergence Theorem (DCT) and the same techniques of the proof of Lemma 3.1 (see Remark 3.5), it can be seen that:

$$\lim_{h\to 0}\int_0^\infty \phi_{h,i,j}(r)\,dr = \lambda_q(L)^{-1}2^{\frac{q}{2}-j}\int_0^\infty L^2(r)r^{\frac{q}{2}+i}\,dr \overset{(3.16)}{=} \begin{cases} \frac{2^{1-j}}{\omega_{q-1}}d_q(L), & i=-1,\\ \frac{2^{1-j}}{\omega_{q-1}}e_q(L), & i=0,\\ \frac{2^{1-j}}{\omega_{q-1}}\frac{\int_0^\infty L^2(r)r^{\frac{q}{2}+1}\,dr}{\int_0^\infty L(r)r^{\frac{q}{2}-1}\,dr}, & i=1, \end{cases}$$

where $e_q(L)=\int_0^\infty L^2(r)r^{\frac{q}{2}}\,dr/\int_0^\infty L(r)r^{\frac{q}{2}-1}\,dr$. Then, taking into account that $\int_0^\infty \varphi_{h,i,j}(r)\,dr = \int_0^\infty \varphi_{i,j}(r)\,dr\,(1+o(1))$ the integrals in brackets of (3.21) can be replaced, obtaining that

$$(3.21) = \frac{c_{h,q}(L)}{n}\left[d_q(L)f(\mathbf{x}) + e_q(L)h^2\Psi(f,\mathbf{x})\right] + o\left((nh^q)^{-1}\right). \tag{3.22}$$

The second term in (3.19) is given by

$$\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]^2 = \left[f(\mathbf{x}) + b_q(L)h^2\Psi(f,\mathbf{x})\right]^2 + o\left(h^2\right). \tag{3.23}$$

The result holds from (3.22) and (3.23):

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\hat{f}_h(\mathbf{x})\right] &= \frac{c_{h,q}(L)}{n}\left[d_q(L)f(\mathbf{x}) + e_q(L)h^2\Psi(f,\mathbf{x})\right]\\ &\quad - \frac{1}{n}\left[f(\mathbf{x}) + b_q(L)h^2\Psi(f,\mathbf{x})\right]^2 + o\left((nh^q)^{-1}\right),\end{aligned}$$

which can be simplified into

$$\mathbb{V}\mathrm{ar}\left[\hat{f}_h(\mathbf{x})\right] = \frac{c_{h,q}(L)}{n}d_q(L)f(\mathbf{x}) + o\left((nh^q)^{-1}\right).$$

□

Proof of Proposition 3.2. Denote by $\mathrm{Bias}[\hat{f}_{h,g}(\mathbf{x},z)] = \mathbb{E}[\hat{f}_{h,g}(\mathbf{x},z)] - f(\mathbf{x},z)$ the bias of the kernel estimator. Applying the change of variables stated in Lemma 3.2 and then an ordinary change of variables given by $r=\frac{1-t}{h^2}$, the bias results in:

$$\begin{aligned}\mathrm{Bias}\left[\hat{f}_{h,g}(\mathbf{x},z)\right] &= \frac{c_{h,q}(L)}{g}\mathbb{E}\left[LK\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2},\frac{z-Z}{g}\right)\right] - f(\mathbf{x},z)\\
&= \frac{c_{h,q}(L)}{g}\int_{\Omega_q}\int_{\mathbb{R}} LK\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2},\frac{z-t}{g}\right)(f(\mathbf{y},t)-f(\mathbf{x},z))\;dt\,\omega_q(d\mathbf{y})\\
&= c_{h,q}(L)\int_{\Omega_q}\int_{\mathbb{R}} LK\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2},v\right)(f(\mathbf{y},z-gv)-f(\mathbf{x},z))\;dv\,\omega_q(d\mathbf{y})\\
&= c_{h,q}(L)\int_{-1}^1\int_{\Omega_{q-1}}\int_{\mathbb{R}} LK\left(\frac{1-u}{h^2},v\right)\left(f\left(u\mathbf{x}+(1-u^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi},z-gv\right)-f(\mathbf{x},z)\right)\\
&\quad\times(1-u^2)^{\frac{q}{2}-1}\,dv\,\omega_{q-1}(d\boldsymbol{\xi})\,du\\
&= c_{h,q}(L)h^q\int_0^{2h^{-2}}\int_{\Omega_{q-1}}\int_{\mathbb{R}} LK(r,v)\left(f\left((\mathbf{x},z)+\boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}}\right)-f(\mathbf{x},z)\right)\,dv\,\omega_{q-1}(d\boldsymbol{\xi})\\
&\quad\times r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}\,dr\end{aligned}$$

$$= c_{h,q}(L) h^q \int_0^{2h^{-2}} L(r)\, r^{\frac{q}{2}-1} (2-h^2 r)^{\frac{q}{2}-1} \int_{\mathbb{R}} K(v) \times \int_{\Omega_{q-1}} \left(f\left((\mathbf{x},z)+\boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}}\right) - f(\mathbf{x},z)\right) \omega_{q-1}(d\boldsymbol{\xi})\, dv\, dr, \tag{3.24}$$

where $\boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}} = \left(-rh^2\mathbf{x} + h\left[r(2-h^2r)\right]^{\frac{1}{2}} \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}, -gv\right) \in \Omega_q \times \mathbb{R}$. The computation of the last integral in (3.24) is achieved using the multivariate Taylor expansion of f at $(\mathbf{x},z)$, in virtue of condition **DL1**:

$$f((\mathbf{x},z)+\boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}}) - f(\mathbf{x},z) = \boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}}^T \boldsymbol{\nabla} f(\mathbf{x},z) + \frac{1}{2}\boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}}^T \boldsymbol{\mathcal{H}} f(\mathbf{x},z) \boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}} + o\left(\boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}}^T \boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}}\right).$$

Let denote by $\boldsymbol{\gamma}_{\mathbf{x},\boldsymbol{\xi}} = -rh^2\mathbf{x} + h\left[r(2-h^2r)\right]^{\frac{1}{2}} \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}$. Bearing in mind the directional and linear components of the gradient $\boldsymbol{\nabla} f(\mathbf{x},z)$ and the Hessian matrix $\boldsymbol{\mathcal{H}} f(\mathbf{x},z)$, it follows

$$\begin{aligned} f((\mathbf{x},z)+\boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}}) - f(\mathbf{x},z) = &\left[\boldsymbol{\gamma}_{\mathbf{x},\boldsymbol{\xi}}^T \boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x},z) - gv\nabla_z f(\mathbf{x},z)\right] \\ &+ \frac{1}{2}\left[\boldsymbol{\gamma}_{\mathbf{x},\boldsymbol{\xi}}^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z) \boldsymbol{\gamma}_{\mathbf{x},\boldsymbol{\xi}} - 2gv\boldsymbol{\gamma}_{\mathbf{x},\boldsymbol{\xi}}^T \boldsymbol{\mathcal{H}}_{\mathbf{x},z} f(\mathbf{x},z) + g^2v^2 \mathcal{H}_z f(\mathbf{x},z)\right] \\ &+ o\left(\boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}}^T \boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}}\right). \end{aligned}$$

Then, the calculus of the integral $\int_{\Omega_{q-1}} \left(f\left((\mathbf{x},z)+\boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}}\right) - f(\mathbf{x},z)\right) \omega_{q-1}(d\boldsymbol{\xi})$ can be split into six addends. Second and sixth terms are computed straightforward:

$$\int_{\Omega_{q-1}} -gv\nabla_z f(\mathbf{x},z)\, \omega_{q-1}(d\boldsymbol{\xi}) = -\omega_{q-1}\, gv\nabla_z f(\mathbf{x},z), \tag{3.25}$$

$$\int_{\Omega_{q-1}} g^2v^2 \mathcal{H}_z f(\mathbf{x},z)\, \omega_{q-1}(d\boldsymbol{\xi}) = \omega_{q-1}\, g^2v^2 \mathcal{H}_z f(\mathbf{x},z). \tag{3.26}$$

For the first and fourth addends, by Lemma 3.3, the integration of ξ_i with respect to $\boldsymbol{\xi}$ is zero:

$$\int_{\Omega_{q-1}} \boldsymbol{\gamma}_{\boldsymbol{\xi},z}^T \boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x},z)\, \omega_{q-1}(d\boldsymbol{\xi}) = -\omega_{q-1} h^2 r \mathbf{x}^T \boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x},z), \tag{3.27}$$

$$\int_{\Omega_{q-1}} -2gv\boldsymbol{\gamma}_{\mathbf{x},\boldsymbol{\xi}} \boldsymbol{\mathcal{H}}_{\mathbf{x},z} f(\mathbf{x},z)\, \omega_{q-1}(d\boldsymbol{\xi}) = 2gv\omega_{q-1} h^2 r \mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x},z} f(\mathbf{x},z). \tag{3.28}$$

Finally, in the fifth term, the integrand can be decomposed as follows:

$$\begin{aligned} \boldsymbol{\gamma}_{\mathbf{x},\boldsymbol{\xi}}^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z) \boldsymbol{\gamma}_{\mathbf{x},\boldsymbol{\xi}} = &\, h^4 r^2 \mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z) \mathbf{x} + h^2 r(2-h^2r) \sum_{i,j=1}^{q} \xi_i \xi_j \mathbf{b}_i^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z) \mathbf{b}_j \\ &- 2h^3 r^{\frac{3}{2}} (2-h^2r)^{\frac{1}{2}} \sum_{i=1}^{q} \xi_i \mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z) \mathbf{b}_i. \end{aligned}$$

In virtue of Lemma 3.3, the third addend vanishes as well as the second, except for the diagonal terms. Next, as $\{\mathbf{x}, \mathbf{b}_1, \ldots, \mathbf{b}_q\}$ is an orthonormal basis in $\mathbb{R}^{q+1}$, the sum of the diagonal terms can be computed by simple algebra:

$$\sum_{i=1}^{q} \mathbf{b}_i^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z) \mathbf{b}_i = \operatorname{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z) \sum_{i=1}^{q} \mathbf{b}_i \mathbf{b}_i^T\right] = \operatorname{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z) \left(\mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T\right)\right]$$

$$= \nabla^2_{\mathbf{x}} f(\mathbf{x}, z) - \mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z)\mathbf{x},$$

where $\nabla^2_{\mathbf{x}} f(\mathbf{x}, z)$ is the Laplacian of f restricted to the directional component $\mathbf{x}$, $\mathbf{I}_{q+1}$ is the identity matrix of order $q+1$ and tr is the trace operator. By Lemma 3.3 and the previous calculus, the fifth term is

$$\begin{aligned}
&\int_{\Omega_{q-1}} \boldsymbol{\gamma}^T_{\boldsymbol{\xi},z} \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z) \boldsymbol{\gamma}_{\boldsymbol{\xi},z}\, \omega_{q-1}(d\boldsymbol{\xi}) \\
&\quad = \omega_{q-1} h^4 r^2 \mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z)\mathbf{x} + \omega_{q-1} h^2 r (2 - h^2 r) q^{-1} \left[\nabla^2_{\mathbf{x}} f(\mathbf{x}, z) - \mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z)\mathbf{x} \right]. \quad (3.29)
\end{aligned}$$

Note also that the order of $\boldsymbol{\alpha}^T_{\mathbf{x},z,\boldsymbol{\xi}} \boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}}$ is easily computed:

$$o\left(\boldsymbol{\alpha}^T_{\mathbf{x},z,\boldsymbol{\xi}} \boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}} \right) = r\, o\left(h^2\right) + v^2 o\left(g^2\right). \quad (3.30)$$

Combining (3.25)–(3.30), and using condition **DL2** on the kernel K:

$$\begin{aligned}
(3.24) &= c_{h,q}(L) h^q \int_0^{2h^{-2}} L(r)\, r^{\frac{q}{2}-1} (2 - h^2 r)^{\frac{q}{2}-1} \int_{\mathbb{R}} K(v) \Big\{ \int_{\Omega_{q-1}} \left[\boldsymbol{\gamma}_{\boldsymbol{\xi},\mathbf{x}} \boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}, z) - g v \nabla_z f(\mathbf{x}, z) \right] \\
&\quad + \frac{1}{2} \left[\boldsymbol{\gamma}^T_{\boldsymbol{\xi},\mathbf{x}} \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z) \boldsymbol{\gamma}_{\boldsymbol{\xi},\mathbf{x}} - 2 g v \boldsymbol{\gamma}^T_{\boldsymbol{\xi},\mathbf{x}} \boldsymbol{\mathcal{H}}_{\mathbf{x},z} f(\mathbf{x}, z) + g^2 v^2 \mathcal{H}_z f(\mathbf{x}, z) \right] \\
&\quad + r\, o\left(h^2\right) + v^2 o\left(g^2\right) \omega_{q-1}(d\boldsymbol{\xi}) \Big\}\, dv\, dr \\
&= \omega_{q-1} c_{h,q}(L) h^q \int_0^{2h^{-2}} L(r)\, r^{\frac{q}{2}-1} (2 - h^2 r)^{\frac{q}{2}-1} \int_{\mathbb{R}} K(v) \Big\{ - h^2 r \mathbf{x}^T \boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}, z) \\
&\quad - g v \nabla_z f(\mathbf{x}, z) + \frac{1}{2} \left[h^4 r^2 \mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z) + h^2 r (2 - h^2 r) q^{-1} \left(\nabla^2_{\mathbf{x}} f(\mathbf{x}, z) - \mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z)\mathbf{x} \right) \right] \\
&\quad + g v h^2 r \mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x},z} f(\mathbf{x}, z) + \frac{g^2 v^2}{2} \mathcal{H}_z f(\mathbf{x}, z) + r\, o\left(h^2\right) + v^2 o\left(g^2\right) \Big\}\, dv\, dr \\
&= \omega_{q-1} c_{h,q}(L) h^q \int_0^{2h^{-2}} L(r)\, r^{\frac{q}{2}-1} (2 - h^2 r)^{\frac{q}{2}-1} \Big\{ - h^2 r \mathbf{x}^T \boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}, z) \\
&\quad + \frac{1}{2} \Big[h^4 r^2 \mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z) + h^2 r (2 - h^2 r) q^{-1} \left(\nabla^2_{\mathbf{x}} f(\mathbf{x}, z) - \mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z)\mathbf{x} \right) \\
&\quad + g^2 \boldsymbol{\mathcal{H}}_z f(\mathbf{x}, z) \mu_2(K) \Big] + r\, o\left(h^2\right) + \mu_2(K) o\left(g^2\right) \Big\}\, dr. \quad (3.31)
\end{aligned}$$

For $h > 0$, $i = -1, 0, 1$, $j = 0, 1$, consider the following functions

$$\varphi_{h,i,j}(r) = c_{h,q}(L) h^q L(r) r^{\frac{q}{2}+i} (2 - h^2 r)^{\frac{q}{2}-j} \mathbf{1}_{[0, 2h^{-2})}(r), \quad r \in [0, \infty).$$

When $n \to \infty$, $h \to 0$ and the limit of $\varphi_{h,i,j}$ is given by

$$\varphi_{i,j}(r) = \lim_{h \to 0} \varphi_{h,i,j}(r) = \lambda_q(L)^{-1} L(r) r^{\frac{q}{2}+i} 2^{\frac{q}{2}-j} \mathbf{1}_{[0,\infty)}(r).$$

Applying Remark 3.5 of Lemma 3.1,

$$\lim_{h \to 0} \int_0^\infty \varphi_{i,j,h}(r)\, dr = \lambda_q(L)^{-1} 2^{\frac{q}{2}-j} \int_0^\infty L(r) r^{\frac{q}{2}+i}\, dr \overset{(3.16)}{=} \begin{cases} \frac{2^{1-j}}{\omega_{q-1}}, & i = -1, \\ \frac{2^{1-j}}{\omega_{q-1}} b_q(L), & i = 0, \\ \frac{2^{1-j}}{\omega_{q-1}} \frac{\int_0^\infty L(r) r^{\frac{q}{2}+1}\, dr}{\int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr}, & i = 1. \end{cases}$$

Then, the six integrals in (3.31) can be written using $\int_0^\infty \varphi_{i,j,h}(r)\,dr = \int_0^\infty \varphi_{i,j}(r)\,dr\,(1+o(1))$. Replacing this in (3.31) leads to

$$\begin{aligned}
(3.31) =& -h^2\omega_{q-1}\left[\frac{b_q(L)}{\omega_{q-1}}+o(1)\right]\mathbf{x}^T\boldsymbol{\nabla}_{\mathbf{x}}f(\mathbf{x},z)\\
&+\frac{h^4\omega_{q-1}}{2}\left[\frac{b_q(L)}{\omega_{q-1}}\frac{\int_0^\infty L(r)r^{\frac{q}{2}+1}\,dr}{\int_0^\infty L(r)r^{\frac{q}{2}}\,dr}+o(1)\right]\mathbf{x}^T\boldsymbol{\mathcal{H}}_{\mathbf{x}}f(\mathbf{x},z)\mathbf{x}\\
&+\frac{h^2\omega_{q-1}}{2}\left[\frac{b_q(L)}{\omega_{q-1}}+o(1)\right]q^{-1}\left(\nabla^2_{\mathbf{x}}f(\mathbf{x},z)-\mathbf{x}^T\boldsymbol{\mathcal{H}}_{\mathbf{x}}f(\mathbf{x},z)\mathbf{x}\right)\\
&+\frac{\omega_{q-1}}{2}\left[\frac{1}{\omega_{q-1}}+o(1)\right]g^2\mathcal{H}_zf(\mathbf{x},z)\mu_2(K)\\
&+\omega_{q-1}\left[\frac{b_q(L)}{\omega_{q-1}}+o(1)\right]o\left(h^2\right)+\omega_{q-1}\left[\frac{1}{\omega_{q-1}}+o(1)\right]o\left(g^2\right)\\
=&\,h^2b_q(L)\left[-\mathbf{x}^T\boldsymbol{\nabla}_{\mathbf{x}}f(\mathbf{x},z)+q^{-1}\left(\nabla^2_{\mathbf{x}}f(\mathbf{x})-\mathbf{x}^T\boldsymbol{\mathcal{H}}_{\mathbf{x}}f(\mathbf{x},z)\mathbf{x}\right)\right]+g^2\mathcal{H}_zf(\mathbf{x},z)\mu_2(K)\\
&+\mathcal{O}\left(h^4\right)+o\left(h^2\right)+o\left(g^2\right)\\
=&\,h^2b_q(L)\Psi_{\mathbf{x}}(f,\mathbf{x},z)+\frac{g^2}{2}\mathcal{H}_zf(\mathbf{x},z)\mu_2(K)+o\left(h^2+g^2\right).
\end{aligned}$$

□

Proof of Proposition 3.3. The variance can be decomposed as

$$\mathbb{V}\mathrm{ar}\left[\hat{f}_{h,g}(\mathbf{x},z)\right]=\frac{c_{h,q}(L)^2}{ng^2}\mathbb{E}\left[LK^2\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2},\frac{z-Z}{g}\right)\right]-n^{-1}\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x},z)\right]^2, \tag{3.32}$$

where the calculus of the first term is quite similar to the calculus of the bias and the second is given in the previous result.

Analogous to (3.24),

$$\begin{aligned}
\frac{c_{h,q}(L)^2}{ng^2}\mathbb{E}\left[LK^2\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2},\frac{z-Z}{g}\right)\right]=&\frac{c_{h,q}(L)^2}{ng}h^q\int_0^{2h^{-2}}L^2(r)r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}\int_{\mathbb{R}}K^2(v)\\
&\times\int_{\Omega_{q-1}}f((\mathbf{x},z)+\boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}})\,\omega_{q-1}(d\boldsymbol{\xi})\,dv\,dr,
\end{aligned} \tag{3.33}$$

just replacing LK by LK^2. Then, using that K^2 is a symmetric function around zero:

$$\int_{\mathbb{R}}K^2(v)\,dv=R(K),\ \int_{\mathbb{R}}vK^2(v)\,dv=0,\ \int_{\mathbb{R}}v^2K^2(v)\,dv=\mu_2\left(K^2\right), \tag{3.34}$$

Applying the multivariate Taylor expansion of f at $(\mathbf{x},z)$ and by (3.34), equation (3.33) results in

$$\begin{aligned}
(3.33) =&\,\omega_{q-1}\frac{c_{h,q}(L)^2}{ng}h^q\int_0^{2h^{-2}}L^2(r)r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}\int_{\mathbb{R}}K^2(v)\Big\{f(\mathbf{x},z)-h^2r\mathbf{x}^T\boldsymbol{\nabla}_{\mathbf{x}}f(\mathbf{x},z)\\
&-gv\nabla_zf(\mathbf{x},z)+\frac{1}{2}\Big[h^4r^2\mathbf{x}^T\boldsymbol{\mathcal{H}}_{\mathbf{x}}f(\mathbf{x},z)+h^2r(2-h^2r)q^{-1}\left(\nabla^2_{\mathbf{x}}f(\mathbf{x},z)-\mathbf{x}^T\boldsymbol{\mathcal{H}}_{\mathbf{x}}f(\mathbf{x},z)\mathbf{x}\right)\Big]
\end{aligned}$$

$$+ gvh^2 r\mathbf{x}^T \boldsymbol{\mathcal{H}}_{\mathbf{x},z} f(\mathbf{x},z) + \frac{g^2v^2}{2}\mathcal{H}_z f(\mathbf{x},z) + r\,o\left(h^2\right) + v^2 o\left(g^2\right)\Big\}\, dv\, dr$$

$$\overset{(3.34)}{=} \omega_{q-1}\frac{c_{h,q}(L)^2}{ng}h^q \int_0^{2h^{-2}} L^2(r) r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}\Big\{R(K)f(\mathbf{x},z) - R(K)h^2 r\mathbf{x}^T\boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x},z)$$

$$+ \frac{R(K)}{2}\left[h^4r^2\mathbf{x}^T\boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z) + h^2r(2-h^2r)q^{-1}\left(\nabla^2_{\mathbf{x}} f(\mathbf{x},z) - \mathbf{x}^T\boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z)\mathbf{x}\right)\right]$$

$$+ \mu_2\left(K^2\right)\frac{g^2}{2}\mathcal{H}_z f(\mathbf{x},z) + r\,o\left(h^2\right) + \mu_2(K^2)o\left(g^2\right)\Big\}\, dr. \tag{3.35}$$

Define the following functions, for $h>0$, $i=-1,0,1$ and $j=0,1$:

$$\phi_{h,i,j}(r) = c_{h,q}(L)h^q L^2(r) r^{\frac{q}{2}+i}(2-h^2r)^{\frac{q}{2}-j}\mathbf{1}_{[0,2h^{-2})}(r), \quad r\in[0,\infty).$$

When $n\to\infty$, $h\to 0$ and the limit of $\phi_{h,i,j}$ is given by

$$\phi_{i,j}(r) = \lim_{h\to 0}\phi_{h,i,j}(r) = \lambda_q(L)^{-1}L^2(r)r^{\frac{q}{2}+i}2^{\frac{q}{2}-j}\mathbf{1}_{[0,\infty)}(r).$$

Applying the same techniques of the proof of Lemma 3.1 to the functions $\phi_{h,i,j}$ with the different values of i,j and L^2 instead of L, and using the relation (3.3), it follows:

$$\lim_{h\to 0}\int_0^\infty \phi_{h,i,j}(r)\,dr = \lambda_q(L)^{-1}2^{\frac{q}{2}-j}\int_0^\infty L^2(r)r^{\frac{q}{2}+i}\,dr \overset{(3.16)}{=} \begin{cases} \frac{2^{1-j}}{\omega_{q-1}}d_q(L), & i=-1,\\ \frac{2^{1-j}}{\omega_{q-1}}e_q(L), & i=0,\\ \frac{2^{1-j}}{\omega_{q-1}}\frac{\int_0^\infty L^2(r)r^{\frac{q}{2}+1}\,dr}{\int_0^\infty L(r)r^{\frac{q}{2}-1}\,dr}, & i=1,\end{cases}$$

where $e_q(L) = \int_0^\infty L^2(r)r^{\frac{q}{2}}\,dr / \int_0^\infty L(r)r^{\frac{q}{2}-1}\,dr$. So, for the terms between square brackets of (3.35), $\int_0^\infty \phi_{h,i,j}(r)\,dr = \int_0^\infty \phi_{i,j}(r)\,dr\,(1+o(1))$. Replacing this leads to

$$(3.35) = \frac{c_{h,q}(L)}{ng}\Big\{R(K)\omega_{q-1}\left[\frac{d_q(L)}{\omega_{q-1}}+o(1)\right]f(\mathbf{x},z) - R(K)h^2\omega_{q-1}\left[\frac{e_q(L)}{\omega_{q-1}}+o(1)\right]\mathbf{x}^T\boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x},z)$$

$$+ \frac{R(K)h^4\omega_{q-1}}{2}\left[\frac{1}{\omega_{q-1}}\frac{\int_0^\infty L^2(r)r^{\frac{q}{2}+1}\,dr}{\int_0^\infty L(r)r^{\frac{q}{2}-1}\,dr}+o(1)\right]\mathbf{x}^T\boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z)$$

$$+ \frac{R(K)h^2\omega_{q-1}}{2}\left[\frac{2e_q(L)}{\omega_{q-1}}+o(1)\right]q^{-1}\left(\nabla^2_{\mathbf{x}} f(\mathbf{x},z) - \mathbf{x}^T\boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z)\mathbf{x}\right)$$

$$+ \frac{\mu_2\left(K^2\right)g^2\omega_{q-1}}{2}\left[\frac{d_q(L)}{\omega_{q-1}}+o(1)\right]\mathcal{H}_z f(\mathbf{x},z)$$

$$+ \omega_{q-1}\left[\frac{e_q(L)}{\omega_{q-1}}+o(1)\right]o\left(h^2\right) + \omega_{q-1}\left[\frac{d_q(L)}{\omega_{q-1}}+o(1)\right]o\left(g^2\right)\Big\}$$

$$= \frac{c_{h,q}(L)}{ng}\left[R(K)d_q(L)f(\mathbf{x},z) + R(K)e_q(L)h^2\Psi_{\mathbf{x}} f(\mathbf{x},z) + \mu_2\left(K^2\right)d_q(L)\frac{g^2}{2}\mathcal{H}_z f(\mathbf{x},z)\right]$$

$$+ o\left((nh^qg)^{-1}\right). \tag{3.36}$$

The second term of (3.32) is

$$\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x},z)\right]^2 = \left[f(\mathbf{x},z) + b_q(L)h^2\Psi_{\mathbf{x}}(f,\mathbf{x},z) + \frac{g^2}{2}\mu_2(K)\mathcal{H}_z f(\mathbf{x},z)\right]^2 + o\left(h^2+g^2\right). \tag{3.37}$$

Joining (3.36) and (3.37),

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\hat{f}_{h,g}(\mathbf{x},z)\right] = \frac{c_{h,q}(L)}{ng}\Big[&R(K)d_q(L)f(\mathbf{x},z) + R(K)e_q(L)h^2\Psi_{\mathbf{x}}f(\mathbf{x},z)\\ &+ \mu_2\left(K^2\right)d_q(L)\frac{g^2}{2}\mathcal{H}_z f(\mathbf{x},z)\Big]\\ &- \frac{1}{n}\left[f(\mathbf{x},z) + b_q(L)h^2\Psi_{\mathbf{x}}(f,\mathbf{x},z) + \frac{g^2}{2}\mu_2(K)\mathcal{H}_z f(\mathbf{x},z)\right]^2\\ &+ o\left((nh^q g)^{-1}\right) + o\left(n^{-1}(h^2+g^2)\right),\end{aligned}$$

which can be simplified into

$$\mathbb{V}\mathrm{ar}\left[\hat{f}_{h,g}(\mathbf{x},z)\right] = \frac{c_{h,q}(L)}{ng}R(K)d_q(L)f(\mathbf{x},z) + o\left((nh^q g)^{-1}\right).$$

□

Proof of Theorem 3.1. Let $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$ be a random sample from the directional-linear random variable $(\mathbf{X}, Z)$, whose support is contained in $\Omega_q \times \mathbb{R}$. The directional kernel estimator in a fixed point $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$ can be written as

$$\hat{f}_{h_n,g_n}(\mathbf{x},z) = \frac{1}{n}\sum_{i=1}^n V_{n,i}, \quad V_{n,i} = \frac{c_{h,q}(L)}{g}LK\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h_n^2}, \frac{z-Z_i}{g_n}\right),$$

where notation h_n and g_n for the bandwidths remarks their dependence on the sample size n given by condition **DL3**.

As $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$ is a collection of independent and identically distributed (iid) copies of $(\mathbf{X}, Z)$, then $\{V_{n,i}\}_{i=1}^n$ is also an iid collection of copies of the random variable $V_n = LK(\frac{1-\mathbf{x}^T\mathbf{X}}{h_n^2}, \frac{z-Z}{g_n})$. Then, the Lyapunov's condition ensures that, if for some $\delta > 0$ the next condition holds:

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[|V_n - \mathbb{E}[V_n]|^{2+\delta}\right]}{n^{\frac{\delta}{2}}\mathbb{V}\mathrm{ar}[V_n]^{1+\frac{\delta}{2}}} = 0,$$

then the following central limit theorem is valid:

$$\sqrt{n}\frac{\bar{V}_n - \mathbb{E}[V_n]}{\sqrt{\mathbb{V}\mathrm{ar}[V_n]}} \xrightarrow{d} \mathcal{N}(0,1),$$

where $\bar{V}_n = \frac{1}{n}\sum_{i=1}^n V_{n,i}$. This condition will be proved for $V_n = LK(\frac{1-\mathbf{x}^T\mathbf{X}}{h_n^2}, \frac{z-Z}{g_n})$.

First of all, the order of $\mathbb{E}\left[\,|V_n|^{2+\delta}\,\right]$ is

$$\begin{aligned}\mathbb{E}\left[|V_n|^{2+\delta}\right] &= \int_{\Omega_q\times\mathbb{R}}\left|\frac{c_{h_n,q}(L)}{g_n}LK\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h_n^2}, \frac{z-t}{g_n}\right)\right|^{2+\delta} f(\mathbf{y},t)\,dt\,\omega_q(d\mathbf{y})\\ &= \left(\frac{c_{h_n,q}(L)}{g_n}\right)^{2+\delta}\int_{\Omega_q\times\mathbb{R}} LK^{2+\delta}\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h_n^2}, \frac{z-t}{g_n}\right) f(\mathbf{y},t)\,dt\,\omega_q(d\mathbf{y})\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{c_{h_n,q}(L)}{g_n}\right)^{2+\delta} g_n h_n^q \int_0^{2h_n^{-2}} \int_{\Omega_{q-1}} \int_{\mathbb{R}} LK^{2+\delta}(r,v)\, f((\mathbf{x},z)+\boldsymbol{\alpha}_{\mathbf{x},z,\boldsymbol{\xi}})\, dv\, \omega_{q-1}(d\boldsymbol{\xi}) \\
&\quad \times r^{\frac{q}{2}-1}\left(2-h_n^2 r\right)^{\frac{q}{2}-1} dr \\
&= \left(\frac{c_{h_n,q}(L)}{g_n}\right)^{2+\delta} g_n h_n^q \int_0^{2h_n^{-2}} \int_{\Omega_{q-1}} \int_{\mathbb{R}} LK^{2+\delta}(r,v)\; dv\, \omega_{q-1}(d\boldsymbol{\xi})\, r^{\frac{q}{2}-1}\left(2-h_n^2 r\right)^{\frac{q}{2}-1} dr \\
&\quad \times \left[f(\mathbf{x},z)+o\left(h_n^2+g_n^2\right)\right] \\
&\sim \left(\frac{c_{h_n,q}(L)}{g_n}\right)^{2+\delta} g_n h_n^q 2^{\frac{q}{2}-1}\omega_{q-1} \int_0^{\infty} \int_{\mathbb{R}} LK^{2+\delta}(r,v)\, r^{\frac{q}{2}-1}\, dv\, dr \times f(\mathbf{x},z) \\
&\sim \left(\frac{\lambda_q(L)^{-1} h_n^{-q}}{g_n}\right)^{2+\delta} g_n h_n^q 2^{\frac{q}{2}-1}\omega_{q-1} \int_0^{\infty} \int_{\mathbb{R}} LK^{2+\delta}(r,v)\, r^{\frac{q}{2}-1}\, dv\, dr \times f(\mathbf{x},z) \\
&= (h_n^q g_n)^{-(1+\delta)} \times \frac{\int_0^{\infty}\int_{\mathbb{R}} LK^{2+\delta}(r,v)\, r^{\frac{q}{2}-1}\, dv\, dr \times f(\mathbf{x},z)}{\left(2^{\frac{q}{2}-1}\omega_{q-1}\right)^{1+\delta}\left(\int_0^{\infty} L(r) r^{\frac{q}{2}-1}\, dr\right)^{2+\delta}} \\
&= \mathcal{O}\left((h_n^q g_n)^{-(1+\delta)}\right).
\end{aligned}
$$

On the other hand, by Proposition 3.3, the variance of V_n has order

$$
\mathbb{V}\mathrm{ar}\left[V_n\right] = \frac{c_{h,q}(L)}{g_n} R(K) d_q(L) f(\mathbf{x},z) + o\left((h_n^q g_n)^{-1}\right) \sim \frac{R(K) d_q(L) f(\mathbf{x},z)}{\lambda_q(L)} \frac{1}{h_n^q g_n} = \mathcal{O}\left((h_n^q g_n)^{-1}\right).
$$

Using that $\mathbb{E}\big[\,|V_n - \mathbb{E}\left[V_n\right]|^{2+\delta}\,\big] = \mathcal{O}(\mathbb{E}\big[\,|V_n|^{2+\delta}\,\big])$ (see Remark 3.4) and by condition **DL3**, it follows that the Lyapunov's condition is satisfied:

$$
\frac{\mathbb{E}\left[|V_n - \mathbb{E}\left[V_n\right]|^{2+\delta}\right]}{n^{\frac{\delta}{2}}\mathbb{V}\mathrm{ar}\left[V_n\right]^{1+\frac{\delta}{2}}} = \mathcal{O}\left(\frac{(h_n^q g_n)^{-(1+\delta)}}{n^{\frac{\delta}{2}}(h_n^q g_n)^{-(1+\frac{\delta}{2})}}\right) = \mathcal{O}\left((n h_n^q g_n)^{-\frac{\delta}{2}}\right) \longrightarrow 0,
$$

as $n \to \infty$. Therefore,

$$
\frac{\hat{f}_{h_n,g_n}(\mathbf{x},z) - \mathbb{E}\left[\hat{f}_{h_n,g_n}(\mathbf{x},z)\right]}{\sqrt{\mathbb{V}\mathrm{ar}\left[\hat{f}_{h_n,g_n}(\mathbf{x},z)\right]}} \xrightarrow{d} \mathcal{N}(0,1),
$$

pointwise for every $(\mathbf{x},z) \in \Omega_q \times \mathbb{R}$ (note that $\sqrt{n}$ is included in the variance term). Plugging-in the asymptotic expressions for the bias and the variance results

$$
\sqrt{n h_n^q g_n}\left(\hat{f}_{h_n,g_n}(\mathbf{x},z) - f(\mathbf{x},z) - \mathrm{ABias}\left[\hat{f}_{h_n,g_n}(\mathbf{x},z)\right]\right) \xrightarrow{d} \mathcal{N}\left(0, R(K) d_q(L) f(\mathbf{x},z)\right).
$$

Remark 3.4. *The proof of* $\mathbb{E}\big[\,|V_n - \mathbb{E}\left[V_n\right]|^{2+\delta}\,\big] = \mathcal{O}(\mathbb{E}\big[\,|V_n|^{2+\delta}\,\big])$ *is simple. For example, using the Newton Binomial: for any* $r \in \mathbb{R}$, $(x+y)^r = \sum_{k=0}^{\infty}\binom{r}{k} x^{r-k}y^k$, *with* $\binom{r}{k} := \frac{r(r-1)\cdots(r-k+1)}{k!}$. *As* $2+\delta > 1$, *by the triangular inequality and the Newton Binomial,*

$$
\begin{aligned}
\mathbb{E}\left[|V_n - \mathbb{E}\left[V_n\right]|^{2+\delta}\right] &\leq \mathbb{E}\left[\left(|V_n| - |\mathbb{E}\left[V_n\right]|\right)^{2+\delta}\right] \\
&= \mathbb{E}\left[\sum_{k=0}^{\infty}\binom{2+\delta}{k}|V_n|^{2+\delta-k}\,|\mathbb{E}\left[V_n\right]|^k\right]
\end{aligned}
$$

$$= \sum_{k=0}^{\infty} \binom{2+\delta}{k} \mathbb{E}\left[|V_n|^{2+\delta-k}\right] |\mathbb{E}[V_n]|^k .$$

Now, as $\mathbb{E}[V_n] = f(\mathbf{x}, z) + \mathcal{O}(h_n^2 + g_n^2) = \mathcal{O}(1)$ *by Proposition 3.2 and by condition* ***DL3****, the terms* $\mathbb{E}[V_n]$ *are constants asymptotically. Also, as* $\mathbb{E}[|V_n|^r] \leq \mathbb{E}[|V_n|^s]$ *for* $0 \leq r \leq s$*, it follows*

$$\mathbb{E}\left[|V_n - \mathbb{E}[V_n]|^{2+\delta}\right] = \mathcal{O}\left(\sum_{k=0}^{\infty} \binom{2+\delta}{k} \mathbb{E}\left[|V_n|^{2+\delta-k}\right]\right) = \mathcal{O}\left(\mathbb{E}\left[|V_n|^{2+\delta}\right]\right).$$

□

Proof of Proposition 3.4. It is straightforward from Proposition 3.1 and Lemma 3.5. For a point $\mathbf{x}$ in Ω_q:

$$\begin{aligned}
\mathrm{MSE}\left[\hat{f}_h(\mathbf{x})\right] &= \left[\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right] - f(\mathbf{x})\right]^2 + \mathbb{V}\mathrm{ar}\left[\hat{f}_h(\mathbf{x})\right] \\
&= \left[b_q(L)\Psi(f,\mathbf{x})h^2 + o\left(h^2\right)\right]^2 + \frac{c_{h,q}(L)}{n} d_q(L) f(\mathbf{x}) + o\left((nh^q)^{-1}\right) \\
&= b_q(L)^2 \Psi(f,\mathbf{x})^2 h^4 + \frac{c_{h,q}(L)}{n} d_q(L) f(\mathbf{x}) + o\left(h^4 + (nh^q)^{-1}\right).
\end{aligned}$$

Integrating over Ω_q in the previous equation,

$$\mathrm{MISE}\left[\hat{f}_h\right] = b_q(L)^2 \int_{\Omega_q} \Psi(f,\mathbf{x})^2 \omega_q(d\mathbf{x}) h^4 + \frac{c_{h,q}(L)}{n} d_q(L) + o\left(h^4 + (nh^q)^{-1}\right).$$

□

Proof of Corollary 3.1. To obtain the bandwidth that minimizes AMISE consider (3.3) in the previous equation and derive it with respect to h:

$$\frac{d}{dh}\mathrm{AMISE}\left[\hat{f}_h\right] = 4b_q(L)^2 R\left(\Psi(f,\cdot)\right) h^3 - q\lambda_q(L)^{-1} h^{-(q+1)} d_q(L) n^{-1} = 0.$$

The solution of this equation results in

$$h_{\mathrm{AMISE}} = \left[\frac{q d_q(L)}{4b_q(L)^2 \lambda_q(L) R(\Psi(f,\cdot)) n}\right]^{\frac{1}{4+q}}.$$

□

Proof of Proposition 3.5. It is straightforward from Propositions 3.2 and 3.3:

$$\begin{aligned}
\mathrm{MSE}\left[\hat{f}_{h,g}(\mathbf{x}, z)\right] &= \left[\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x}, z)\right] - f(\mathbf{x}, z)\right]^2 + \mathbb{V}\mathrm{ar}\left[\hat{f}_{h,g}(\mathbf{x}, z)\right] \\
&= \left[h^2 b_q(L) \Psi_{\mathbf{x}}(f,\mathbf{x},z) + \frac{g^2}{2} \mathcal{H}_z f(\mathbf{x}, z) \mu_2(K) + o\left(h^2\right) + o\left(g^2\right)\right]^2 \\
&\quad + \frac{c_{h,q}(L)}{ng} R(K) d_q(L) f(\mathbf{x}, z) + o\left((nh^q g)^{-1}\right) \\
&= h^4 b_q(L)^2 \Psi_{\mathbf{x}}(f,\mathbf{x},z)^2 + \frac{g^4}{4} \mu_2(K)^2 \mathcal{H}_z f(\mathbf{x}, z)^2
\end{aligned}$$

$$+ h^2 g^2 b_q(L)\mu_2(K)\mathcal{H}_z f(\mathbf{x},z)\Psi_{\mathbf{x}}(f,\mathbf{x},z)$$
$$+ \frac{c_{h,q}(L)}{ng} R(K) d_q(L) f(\mathbf{x},z) + o\left(h^4 + g^4 + (nh^q g)^{-1}\right).$$

Integrating the previous equation and denoting by $I[\phi] = \int_{\Omega_q \times \mathbb{R}} \phi(\mathbf{x},z)\, dz\, \omega_q(d\mathbf{x})$ for a function $\phi : \Omega_q \times \mathbb{R} \to \mathbb{R}$,

$$\begin{aligned}\mathrm{MISE}\left[\hat{f}_{h,g}\right] =& b_q(L)^2 I\left[\Psi_{\mathbf{x}}(f,\cdot,\cdot)^2\right] h^4 + \frac{g^4}{4}\mu_2(K)^2 I\left[\mathcal{H}_z f(\cdot,\cdot)^2\right] \\ &+ h^2 g^2 b_q(L)\mu_2(K) I\left[\Psi_{\mathbf{x}}(f,\cdot,\cdot)\mathcal{H}_z f(\cdot,\cdot)\right] + \frac{c_{h,q}(L)}{ng} d_q(L) R(K) \\ &+ o\left(h^4 + g^4 + (nh^q g)^{-1}\right).\end{aligned}$$

□

Proof of Corollary 3.2. Suppose that $g = \beta h$ in the previous equation. Again, use that $c_{h,q}(L) \sim \lambda_q(L)^{-1} h^{-q}$ and derive with respect to h to obtain

$$\frac{d}{dh}\mathrm{AMISE}\left[\hat{f}_{h,\beta h}\right] = 4c_1 h^3 + 4c_2 h^3 + 4c_3 h^3 - (q+1)c_4 h^{-(q+2)} = 0,$$

where

$$\begin{aligned} c_1 =& b_q(L)^2 I\left[\Psi_{\mathbf{x}}(f,\cdot,\cdot)^2\right], \quad c_2 = \frac{1}{4}\mu_2(K)^2 I\left[\mathcal{H}_z f(\cdot,\cdot)^2\right]\beta^4, \\ c_3 =& b_q(L)\mu_2(K) I\left[\Psi_{\mathbf{x}}(f,\cdot,\cdot)\mathcal{H}_z f(\cdot,\cdot)\right]\beta^2, \quad c_4 = \frac{d_q(L)R(K)}{\lambda_q(L) n\beta}. \end{aligned}$$

It follows immediately that

$$h_{\mathrm{AMISE}} = \left[\frac{(q+1)c_4}{4(c_1+c_2+c_3)}\right]^{\frac{1}{5+q}}.$$

Given that $R(b_q(L)\Psi_{\mathbf{x}}(f,\cdot,\cdot) + \frac{\beta^2}{2}\mu_2(K)\mathcal{H}_z f(\cdot,\cdot)) = c_1 + c_2 + c_3$, the desired expression is obtained. In the case where $q = 1$ it is possible to derive the form of β by solving $\frac{\partial}{\partial h}\mathrm{AMISE}[\hat{f}_{h,g}] = 0$ and $\frac{\partial}{\partial g}\mathrm{AMISE}[\hat{f}_{h,g}] = 0$. For this case, β has the closed form

$$\beta = \left(\frac{\frac{1}{4}\mu_2(K)^2 I\left[\mathcal{H}_z f(\cdot,\cdot)^2\right]}{b_q(L)^2 I\left[\Psi_{\mathbf{x}}(f,\cdot,\cdot)^2\right]}\right)^{\frac{1}{4}}.$$

□

Proof of Proposition 3.6. Consider the r-mixture of directional von Mises densities given in (3.11). Then:

$$\begin{aligned}\mathrm{MISE}\left[\hat{f}_h\right] =& \mathbb{E}\left[\int_{\Omega_q}\left(\hat{f}_h(\mathbf{x}) - f_r(\mathbf{x})\right)^2 \omega_q(d\mathbf{x})\right] \\ =& \mathbb{E}\left[\int_{\Omega_q} \hat{f}_h(\mathbf{x})^2 - 2\hat{f}_h(\mathbf{x}) f_r(\mathbf{x}) + f_r(\mathbf{x})^2\, \omega_q(d\mathbf{x})\right]\end{aligned}$$

$$
\begin{aligned}
&= \frac{c_{h,q}(L)^2}{n} \int_{\Omega_q}\int_{\Omega_q} L^2\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) f_r(\mathbf{y})\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y}) \\
&\quad + \frac{c_{h,q}(L)^2(n-1)}{n} \int_{\Omega_q}\int_{\Omega_q}\int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) L\left(\frac{1-\mathbf{x}^T\mathbf{z}}{h^2}\right) f_r(\mathbf{y}) f_r(\mathbf{z}) \\
&\quad \times \omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y})\,\omega_q(d\mathbf{z}) \\
&\quad - 2c_{h,q}(L) \int_{\Omega_q}\int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) f_r(\mathbf{x}) f_r(\mathbf{y})\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y}) \\
&\quad + \int_{\Omega_q} f_r(\mathbf{x})^2\,\omega_q(d\mathbf{x}) \\
&= (38) + (39) - (40) + (41).
\end{aligned}
$$

The four terms of the previous equation will be computed separately. The first one is

$$
\begin{aligned}
(38) &= \frac{c_{h,q}(L)^2}{n} \int_{\Omega_q}\int_{\Omega_q} L^2\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) f_r(\mathbf{y})\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y}) \\
&= \sum_{j=1}^{n} p_j \frac{c_{h,q}(L)^2}{n} \int_{\Omega_q}\int_{\Omega_q} e^{-2\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}} C_q(\kappa_j) e^{\kappa_j \mathbf{y}^T\boldsymbol{\mu}_j}\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y}) \\
&= \sum_{j=1}^{n} p_j \frac{c_{h,q}(L)^2}{n} \int_{\Omega_q}\int_{\Omega_q} e^{-\frac{1-\mathbf{x}^T\mathbf{y}}{(h/\sqrt{2})^2}}\,\omega_q(d\mathbf{x}) C_q(\kappa_j) e^{\kappa_j \mathbf{y}^T\boldsymbol{\mu}_j}\,\omega_q(d\mathbf{y}) \\
&= \sum_{j=1}^{n} p_j \frac{c_{h,q}(L)^2}{c_{h/\sqrt{2},q}(L)n} \int_{\Omega_q} C_q(\kappa_j) e^{\kappa_j \mathbf{y}^T\boldsymbol{\mu}_j}\,\omega_q(d\mathbf{y}) \\
&= \sum_{j=1}^{n} p_j \frac{c_{h,q}(L)^2}{c_{h/\sqrt{2},q}(L)n} \\
&= (D_q(h)n)^{-1}.
\end{aligned}
$$

The second one is

$$
\begin{aligned}
(39) &= \frac{c_{h,q}(L)^2(n-1)}{n} \int_{\Omega_q}\int_{\Omega_q}\int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) L\left(\frac{1-\mathbf{x}^T\mathbf{z}}{h^2}\right) f_r(\mathbf{y}) f_r(\mathbf{z})\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y})\,\omega_q(d\mathbf{z}) \\
&= \frac{c_{h,q}(L)^2(n-1)}{n} \int_{\Omega_q}\int_{\Omega_q}\int_{\Omega_q} e^{-2/h^2} e^{\mathbf{x}^T\mathbf{y}/h^2} e^{\mathbf{x}^T\mathbf{z}/h^2} \sum_{j=1}^{r}\sum_{l=1}^{r} p_j p_l C_q(\kappa_j) C_q(\kappa_l) e^{\kappa_j \mathbf{y}^T\boldsymbol{\mu}_j} e^{\kappa_l \mathbf{z}^T\boldsymbol{\mu}_l} \\
&\quad \times\ \omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y})\,\omega_q(d\mathbf{z}) \\
&= \frac{c_{h,q}(L)^2(n-1)}{n} e^{-2/h^2} \sum_{j=1}^{r}\sum_{l=1}^{r} p_j p_l C_q(\kappa_j) C_q(\kappa_l) \\
&\quad \times \int_{\Omega_q}\int_{\Omega_q}\int_{\Omega_q} e^{\mathbf{x}^T\mathbf{y}/h^2} e^{\mathbf{x}^T\mathbf{z}/h^2} e^{\kappa_j \mathbf{y}^T\boldsymbol{\mu}_j} e^{\kappa_l \mathbf{z}^T\boldsymbol{\mu}_l}\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y})\,\omega_q(d\mathbf{z}) \\
&= \frac{(n-1)}{n} \left((2\pi)^{\frac{q+1}{2}} h^{q-1} \mathcal{I}_{\frac{q-1}{2}}(1/h^2)\right)^{-1} \sum_{j=1}^{r}\sum_{l=1}^{r} p_j p_l C_q(\kappa_j) C_q(\kappa_l) \\
&\quad \times \int_{\Omega_q} \left[\int_{\Omega_q} e^{\mathbf{x}^T\mathbf{y}/h^2 + \kappa_j \mathbf{y}^T\boldsymbol{\mu}_j}\,\omega_q(d\mathbf{y}) \int_{\Omega_q} e^{\mathbf{x}^T\mathbf{z}/h^2 + \kappa_l \mathbf{z}^T\boldsymbol{\mu}_l}\,\omega_q(d\mathbf{z})\right] \omega_q(d\mathbf{x})
\end{aligned}
$$

$$
\begin{aligned}
&= (1-n^{-1}) C_q(1/h^2) \sum_{j=1}^{r}\sum_{l=1}^{r} p_j p_l C_q(\kappa_j) C_q(\kappa_l) \int_{\Omega_q} \left[\int_{\Omega_q} e^{||\mathbf{x}/h^2+\kappa_j\boldsymbol{\mu}_j||\mathbf{y}^T\left(\frac{\mathbf{x}/h^2+\kappa_j\boldsymbol{\mu}_j}{||\mathbf{x}/h^2+\kappa_j\boldsymbol{\mu}_j||}\right)} \omega_q(d\mathbf{y}) \right. \\
&\quad \left. \times \int_{\Omega_q} e^{||\mathbf{x}/h^2+\kappa_l\boldsymbol{\mu}_l||\mathbf{z}^T\left(\frac{\mathbf{x}/h^2+\kappa_l\boldsymbol{\mu}_l}{||\mathbf{x}/h^2+\kappa_l\boldsymbol{\mu}_l||}\right)} \omega_q(d\mathbf{z}) \right] \omega_q(d\mathbf{x}) \\
&= (1-n^{-1}) C_q(1/h^2) \sum_{j=1}^{r}\sum_{l=1}^{r} p_j p_l \int_{\Omega_q} \frac{C_q(\kappa_j) C_q(\kappa_l)}{C_q(||\mathbf{x}/h^2+\kappa_j\boldsymbol{\mu}_j||) C_q(||\mathbf{x}/h^2+\kappa_l\boldsymbol{\mu}_l||)} \omega_q(d\mathbf{x}) \\
&= (1-n^{-1}) \mathbf{p}^T \boldsymbol{\Psi}_{\mathbf{2}}(h) \mathbf{p},
\end{aligned}
$$

where $\boldsymbol{\Psi}_{\mathbf{2}}(h)_{r\times r}$ is the matrix with ij-th entry $C_q\left(1/h^2\right) \int_{\Omega_q} \frac{C_q(\kappa_j)C_q(\kappa_l)}{C_q(||\mathbf{x}/h^2+\kappa_j\boldsymbol{\mu}_j||)C_q(||\mathbf{x}/h^2+\kappa_l\boldsymbol{\mu}_l||)} \omega_q(d\mathbf{x})$. The third one results in:

$$
\begin{aligned}
(40) &= c_{h,q}(L) \int_{\Omega_q}\int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) f_r(\mathbf{x}) f_r(\mathbf{y}) \, \omega_q(d\mathbf{x}) \, \omega_q(d\mathbf{y}) \\
&= c_{h,q}(L) \sum_{j=1}^{r}\sum_{l=1}^{r} p_j p_l \int_{\Omega_q}\int_{\Omega_q} e^{-\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}} C_q(\kappa_j) C_q(\kappa_l) e^{\kappa_j \mathbf{x}^T\boldsymbol{\mu}_j} e^{\kappa_l \mathbf{y}^T\boldsymbol{\mu}_l} \, \omega_q(d\mathbf{x}) \, \omega_q(d\mathbf{y}) \\
&= c_{h,q}(L) e^{-1/h^2} \sum_{j=1}^{r}\sum_{l=1}^{r} p_j p_l C_q(\kappa_j) C_q(\kappa_l) \int_{\Omega_q}\int_{\Omega_q} e^{||\mathbf{y}/h^2+\kappa_j\boldsymbol{\mu}_j||\mathbf{x}^T\left(\frac{\mathbf{y}/h^2+\kappa_j\boldsymbol{\mu}_j}{||\mathbf{y}/h^2+\kappa_j\boldsymbol{\mu}_j||}\right)} \\
&\quad \times \, \omega_q(d\mathbf{x}) e^{\kappa_l \mathbf{y}^T\boldsymbol{\mu}_l} \, \omega_q(d\mathbf{y}) \\
&= C_q(1/h^2) \sum_{j=1}^{r}\sum_{l=1}^{r} p_j p_l C_q(\kappa_j) C_q(\kappa_l) \int_{\Omega_q} \frac{e^{\kappa_l \mathbf{y}^T\boldsymbol{\mu}_l}}{C_q(||\mathbf{y}/h^2+\kappa_j\boldsymbol{\mu}_j||)} \omega_q(d\mathbf{y}) \\
&= \mathbf{p}^T \boldsymbol{\Psi}_{\mathbf{1}}(h) \mathbf{p},
\end{aligned}
$$

where the matrix $\boldsymbol{\Psi}_{\mathbf{1}}(h)_{r\times r}$ has ij-th entry $C_q\left(1/h^2\right) C_q(\kappa_j) C_q(\kappa_l) \int_{\Omega_q} \frac{e^{\kappa_l \mathbf{y}^T\boldsymbol{\mu}_l}}{C_q(||\mathbf{y}/h^2+\kappa_j\boldsymbol{\mu}_j||)} \omega_q(d\mathbf{y})$. Finally, the fourth term is:

$$
\begin{aligned}
(41) &= \int_{\Omega_q} \left(\sum_{j=1}^{r} p_j f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}_j, \kappa_j) \right)^2 \omega_q(d\mathbf{x}) \\
&= \int_{\Omega_q} \sum_{j=1}^{r}\sum_{l=1}^{r} p_j p_l f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}_j, \kappa_j) f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}_l, \kappa_l) \, \omega_q(d\mathbf{x}) \\
&= \sum_{j=1}^{r}\sum_{l=1}^{r} p_j p_l C_q(\kappa_j) C_q(\kappa_l) \int_{\Omega_q} e^{\kappa_j \mathbf{x}^T\boldsymbol{\mu}_j} e^{\kappa_l \mathbf{x}^T\boldsymbol{\mu}_l} \, \omega_q(d\mathbf{x}) \\
&= \sum_{j=1}^{r}\sum_{l=1}^{r} p_j p_l C_q(\kappa_j) C_q(\kappa_l) \int_{\Omega_q} e^{||\kappa_j\boldsymbol{\mu}_j+\kappa_l\boldsymbol{\mu}_l||\mathbf{x}^T\left(\frac{\kappa_j\boldsymbol{\mu}_j+\kappa_l\boldsymbol{\mu}_l}{||\kappa_j\boldsymbol{\mu}_j+\kappa_l\boldsymbol{\mu}_l||}\right)} \omega_q(d\mathbf{x}) \\
&= \sum_{j=1}^{r}\sum_{l=1}^{r} p_j p_l \frac{C_q(\kappa_j) C_q(\kappa_l)}{C_q(||\kappa_j\boldsymbol{\mu}_j+\kappa_l\boldsymbol{\mu}_l||)} \\
&= \mathbf{p}^T \boldsymbol{\Psi}_{\mathbf{0}}(h) \mathbf{p},
\end{aligned}
$$

where $\boldsymbol{\Psi}_{\mathbf{2}}(h)_{r\times r}$ represents the matrix with ij-th entry $C_q\left(1/h^2\right) \int_{\Omega_q} \frac{C_q(\kappa_j)C_q(\kappa_l)}{C_q(||\mathbf{y}/h^2+\kappa_j\boldsymbol{\mu}_j||)} \omega_q(d\mathbf{y})$. Note that if $\kappa_j\boldsymbol{\mu}_j + \kappa_l\boldsymbol{\mu}_l = 0$, then $\int_{\Omega_q} \omega_q(d\mathbf{x}) = \frac{1}{C_q(0)} = \omega_q$ so the result is consistent in this situation. □

Proof of Proposition 3.7. Consider the r-mixture of directional-linear independent von Mises and normals $f_r(\mathbf{x}, z) = \sum_{j=1}^{r} p_j f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}_j, \kappa_j) \times \phi_{\sigma_j}(z - m_j)$. Hence:

$$
\begin{aligned}
\mathrm{MISE}\left[\hat{f}_{h,g}\right] &= \mathbb{E}\left[\int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - f_r(\mathbf{x}, z)\right)^2 dz\, \omega_q(d\mathbf{x})\right] \\
&= \mathbb{E}\left[\int_{\Omega_q \times \mathbb{R}} \hat{f}_{h,g}(\mathbf{x}, z)^2 - 2\hat{f}_{h,g}(\mathbf{x}, z) f_r(\mathbf{x}, z) + f_r(\mathbf{x}, z)^2\, dz\, \omega_q(d\mathbf{x})\right] \\
&= \frac{c_{h,q}(L)^2}{ng^2} \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} LK^2\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}, \frac{z - t}{g}\right) f_r(\mathbf{y}, t)\, dz\, \omega_q(d\mathbf{x})\, dt\, \omega_q(d\mathbf{y}) \\
&\quad + \frac{c_{h,q}(L)^2(n-1)}{ng} \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} LK\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}, \frac{z - t}{g^2}\right) \\
&\quad \times LK\left(\frac{1 - \mathbf{x}^T\mathbf{u}}{h^2}, \frac{z - s}{g}\right) f_r(\mathbf{y}, t) f_r(\mathbf{u}, s)\, dz\, \omega_q(d\mathbf{x})\, dt\, \omega_q(d\mathbf{y})\, ds\, \omega_q(d\mathbf{u}) \\
&\quad - 2\frac{c_{h,q}(L)}{g} \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} LK\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}, \frac{z - t}{g}\right) f_r(\mathbf{x}, z) f_r(\mathbf{y}, t) \\
&\quad \times\ dz\, \omega_q(d\mathbf{x})\, dt\, \omega_q(d\mathbf{y}) \\
&\quad + \int_{\Omega_q \times \mathbb{R}} f_r(\mathbf{x}, z)^2\, dz\, \omega_q(d\mathbf{x}). \qquad (3.42)
\end{aligned}
$$

As the directional-kernel is a product kernel and the mixtures are independent the directional and linear parts can be easily disentangled:

$$
\begin{aligned}
(3.42) &= n^{-1} \sum_{j=1}^{n} p_j \left[c_{h,q}(L)^2 \int_{\Omega_q} \int_{\Omega_q} L^2\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}\right) f_{\mathrm{vM}}(\mathbf{y}; \boldsymbol{\mu}_j, \kappa_j)\, \omega_q(d\mathbf{x})\, \omega_q(d\mathbf{y})\right] \\
&\quad \times \left[\frac{1}{g^2} \int_{\mathbb{R}} \int_{\mathbb{R}} K^2\left(\frac{z - t}{g}\right) \phi_{\sigma_j}(t - m_j)\, dz\, dt\right] \\
&\quad + (1 - n^{-1}) \sum_{j=1}^{n} \sum_{l=1}^{n} p_j p_l \left[c_{h,q}(L)^2 \int_{\Omega_q} \int_{\Omega_q} \int_{\Omega_q} L\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}\right) L\left(\frac{1 - \mathbf{x}^T\mathbf{u}}{h^2}\right)\right. \\
&\quad \left. \times f_{\mathrm{vM}}(\mathbf{y}; \boldsymbol{\mu}_j, \kappa_j) f_{\mathrm{vM}}(\mathbf{u}; \boldsymbol{\mu}_l, \kappa_l)\, \omega_q(d\mathbf{x})\, \omega_q(d\mathbf{y})\, \omega_q(d\mathbf{u})\right] \\
&\quad \times \left[\frac{1}{g} \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} K\left(\frac{z - t}{g}\right) K\left(\frac{z - s}{g}\right) \phi_{\sigma_j}(t - m_j) \phi_{\sigma_l}(s - m_l)\, dz\, dt\, ds\right] \\
&\quad - 2 \sum_{j=1}^{n} \sum_{l=1}^{n} p_j p_l \left[c_{h,q}(L) \int_{\Omega_q} \int_{\Omega_q} L\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}\right) f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}_j, \kappa_j) f_{\mathrm{vM}}(\mathbf{y}; \boldsymbol{\mu}_l, \kappa_l)\right. \\
&\quad \left. \times\ \omega_q(d\mathbf{x})\, \omega_q(d\mathbf{y})\right] \times \left[\frac{1}{g} \int_{\mathbb{R}} \int_{\mathbb{R}} K\left(\frac{z - t}{g}\right) \phi_{\sigma_j}(z - m_j) \phi_{\sigma_l}(z - m_l)\, dz\, dt\right] \\
&\quad + \sum_{j=1}^{n} \sum_{l=1}^{n} p_j p_l \left[\int_{\mathbb{R}} \phi_{\sigma_j}(z - m_j)\, \phi_{\sigma_l}(z - m_l)\, dz\right] \times \left[\int_{\Omega_q} f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}_j, \kappa_j)\right. \\
&\quad \left. \times f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}_l, \kappa_l)\, \omega_q(d\mathbf{x})\right]. \qquad (3.43)
\end{aligned}
$$

The directional parts were calculated in the previous theorem and the linear ones were studied in Marron and Wand (1992) (see also Wand and Jones (1995), page 26). The combination of these two results yields

$$\begin{aligned}(3.43) = &\left(D_q(h)2\pi^{\frac{1}{2}}ng\right)^{-1} + (1-n^{-1})\mathbf{p}^T\left[\mathbf{\Psi_2}(h)\circ\mathbf{\Omega_2}(g)\right]\mathbf{p} + \mathbf{p}^T\left[\mathbf{\Psi_1}(h)\circ\mathbf{\Omega_1}(g)\right]\mathbf{p}\\ &+ \mathbf{p}^T\left[\mathbf{\Psi_0}(h)\circ\mathbf{\Omega_0}(g)\right]\mathbf{p},\end{aligned}$$

where the $r\times r$ matrices $\mathbf{\Omega_a}(g)$ have the ij-th entry equal to $\phi_{\sigma_a}(m_i-m_j)$, $\sigma_a=\left(ag^2+\sigma_i^2+\sigma_j^2\right)^{\frac{1}{2}}$ for $a=0,1,2$ and $\mathbf{\Psi_a}(h)$ are the matrices of Proposition 3.6. The notation $\circ$ denotes the Hadamard product between matrices, *i.e.*, if $(\mathbf{A})_{ij}=a_{ij}$, $(\mathbf{B})_{ij}=b_{ij}$, then $(\mathbf{A}\circ\mathbf{B})_{ij}=a_{ij}b_{ij}$. □

Proof of Corollary 3.3. In virtue of equation (3.7), if the kernel of the density estimator (3.2) is $L(r)=e^{-r}$, $r\geq 0$, then the kernel estimator is the n-mixture of von Mises with means $\mathbf{X}_i$, $i=1,\ldots,n$, and common concentrations $1/h_P^2$ given by (3.7), where h_P is the pilot bandwidth parameter □

Proof of Corollary 3.4. It follows immediately from the previous proposition and corollary. □

3.C Proofs of the technical lemmas

Proof of Lemma 3.1. Consider the functions

$$\begin{aligned}\varphi_h(r) &= L(r)r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}\mathbf{1}_{[0,2h^{-2})}(r),\\ \varphi(r) &= \lim_{h\to 0}\varphi_h(r) = L(r)r^{\frac{q}{2}-1}2^{\frac{q}{2}-1}\mathbf{1}_{[0,\infty)}(r).\end{aligned}$$

Then, proving $\lim_{h\to 0}\lambda_{h,q}(L)=\lambda_q(L)$ is equivalent to proving $\lim_{h\to 0}\int_0^\infty\varphi_h(r)\,dr=\int_0^\infty\varphi(r)\,dr$.

Consider first the case $q\geq 2$. As $\frac{q}{2}-1\geq 0$, then $(2-h^2r)^{\frac{q}{2}-1}\leq 2^{\frac{q}{2}-1}$, $\forall h>0$, $\forall r\in[0,2h^{-2})$. Then:

$$|\varphi_h(r)|\leq L(r)r^{\frac{q}{2}-1}2^{\frac{q}{2}-1}\mathbf{1}_{[0,2h^{-2})}(r)\leq\varphi(r),\quad\forall r\in[0,\infty),\forall h>0.$$

Because $\int_0^\infty\varphi(r)\,dr<\infty$ by condition **D2** on the kernel L, then by the DCT it follows that $\lim_{h\to 0}\int_0^\infty\varphi_h(r)\,dr=\int_0^\infty\varphi(r)\,dr$.

For the case $q=1$, $\varphi_h(r)=L(r)r^{-\frac{1}{2}}(2-h^2r)^{-\frac{1}{2}}$. Consider now the following decomposition:

$$\int_0^\infty\varphi_h(r)\,dr=\int_0^\infty L(r)r^{-\frac{1}{2}}(2-h^2r)^{-\frac{1}{2}}\mathbf{1}_{[0,h^{-2})}(r)\,dr+\int_0^\infty L(r)r^{-\frac{1}{2}}(2-h^2r)^{-\frac{1}{2}}\mathbf{1}_{[h^{-2},2h^{-2})}(r)\,dr.$$

The limit of the first integral can be derived analogously with the DCT. As $(2-h^2r)^{-\frac{1}{2}}$ is monotone increasing, then $(2-h^2r)^{-\frac{1}{2}}\leq 1$, $\forall r\in[0,h^{-2})$, $\forall h>0$. Therefore:

$$\left|L(r)r^{-\frac{1}{2}}(2-h^2r)^{-\frac{1}{2}}\mathbf{1}_{[0,h^{-2})}(r)\right|\leq L(r)r^{-\frac{1}{2}}\mathbf{1}_{[0,h^{-2})}(r)\leq\varphi(r),\quad\forall r\in[0,\infty),\,\forall h>0.$$

Then, as $\lim_{h\to 0}L(r)r^{-\frac{1}{2}}(2-h^2r)^{-\frac{1}{2}}\mathbf{1}_{[0,h^{-2})}(r)=\varphi(r)$ and $\int_0^\infty\varphi(r)\,dr<\infty$ by condition **D2**, DCT guarantees that $\lim_{h\to 0}\int_0^\infty L(r)r^{-\frac{1}{2}}(2-h^2r)^{-\frac{1}{2}}\mathbf{1}_{[0,h^{-2})}(r)\,dr=\int_0^\infty\varphi(r)\,dr$.

For the second integral, as a consequence of **D2** and Remark 3.2, L must decrease faster than any power function. In particular, for some fixed $h_0 > 0$, $L(r) \leq r^{-1}$, $\forall r \in [h^{-2}, 2h^{-2})$, $\forall h \in (0, h_0)$. Using this, it results in:

$$\lim_{h\to 0} \int_{h^{-2}}^{2h^{-2}} L(r) r^{-\frac{1}{2}} (2 - h^2 r)^{-\frac{1}{2}} \, dr \leq \lim_{h\to 0} \int_{h^{-2}}^{2h^{-2}} r^{-\frac{3}{2}} (2 - h^2 r)^{-\frac{1}{2}} \, dr = \lim_{h\to 0} h = 0.$$

This completes the proof. □

Remark 3.5. *It is possible to apply the same techniques to prove the result with the functions*

$$\varphi_{h,i,j,k}(r) = L^k(r) r^{\frac{q}{2}+i} (2 - h^2 r)^{\frac{q}{2}-j} \mathbf{1}_{[0,2h^{-2})}(r),$$
$$\varphi_{i,j,k}(r) = \lim_{h\to 0} \varphi_{h,i,j,k}(r) = L^k(r) r^{\frac{q}{2}+i} 2^{\frac{q}{2}-j} \mathbf{1}_{[0,\infty)}(r),$$

with $i = -1, 0, 1$, $j = 0, 1$ and $k = 1, 2$. For the cases where $\frac{q}{2} - j \geq 0$, use DCT. For the other cases, subdivide the integral over $[0, 2h^{-2})$ into the intervals $[0, h^{-2})$ and $[h^{-2}, 2h^{-2})$. Then apply DCT in the former and use a suitable power function to make the latter tend to zero in the same way as described previously.

Proof of Lemma 3.2. Following Blumenson (1960), if $\mathbf{x}$ is a vector of norm r with components x_j, $j = 1, \ldots, n$, with respect to an orthonormal basis in $\mathbb{R}^n$, then the n-dimensional spherical coordinates of $\mathbf{x}$ are given by

$$\begin{cases} x_1 = r \cos \phi_1, \\ x_j = r \cos \phi_j \prod_{k=1}^{j-1} \sin \phi_k, \quad j = 2, \ldots, n-2, \\ x_{n-1} = r \sin \theta \prod_{k=1}^{n-2} \sin \phi_k, \\ x_n = r \cos \theta \prod_{k=1}^{n-2} \sin \phi_k, \end{cases} \qquad J = r^{n-1} \prod_{k=1}^{n-2} \sin^k \phi_{n-1-k}. \tag{3.44}$$

where $0 \leq \phi_j \leq \pi$, $j = 1, \ldots, n-2$, $0 \leq \theta < 2\pi$ and $0 \leq r < \infty$. J denotes the Jacobian of the transformation. Special cases of this parametrization are the polar coordinates ($n = 2$),

$$\begin{cases} x_1 = r \cos \theta, \\ x_2 = r \sin \theta, \end{cases} \qquad J = r,$$

and the spherical coordinates ($n = 3$),

$$\begin{cases} x_1 = r \cos \phi, \\ x_2 = r \sin \theta \sin \phi, \\ x_3 = r \cos \theta \sin \phi, \end{cases} \qquad J = r^2 \sin \phi.$$

Note that sometimes this parametrization appears with the roles of x_1 and x_3 swapped.

To continue with the previous notation, let denote $q = n - 1$. Using the spherical coordinates ($r = 1$, as the integration is on Ω_{n-1}) and then applying the change of variables

$$t = \cos \phi_1, \quad d\phi_1 = -(1 - t^2)^{-\frac{1}{2}} \, dt, \tag{3.45}$$

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA USC

it follows that

$$
\begin{aligned}
&\int_{\Omega_{n-1}} f(\mathbf{x})\,\omega_{n-1}(d\mathbf{x}) \\
&\quad= \int_{\Omega_{n-1}} f(x_1,\ldots,x_n)\,d(x_1,\ldots,x_n) \\
&\quad\overset{(3.44)}{=} \int_0^{2\pi}\int_0^{\pi} \times \overset{(n-2)}{\cdots} \times \int_0^{\pi} f\Big(\cos\phi_1, \cos\phi_2\sin\phi_1,\ldots,\cos\theta\prod_{k=1}^{n-2}\sin\phi_k\Big) \\
&\qquad\times \prod_{k=1}^{n-2}\sin^k\phi_{n-1-k}\prod_{j=n-2}^{1} d\phi_j\;d\theta \\
&\quad\overset{(3.45)}{=} \int_0^{2\pi}\int_{-1}^{1}\int_0^{\pi} \times \overset{(n-1)}{\cdots} \times \int_0^{\pi} f\Big(t,\cos\phi_2(1-t^2)^{\frac{1}{2}},\ldots,\cos\theta\prod_{k=2}^{n-2}\sin\phi_k(1-t^2)^{\frac{1}{2}}\Big) \\
&\qquad\times \prod_{k=1}^{n-3}\sin^k\phi_{n-1-k}(1-t^2)^{\frac{n-2}{2}}(1-t^2)^{-\frac{1}{2}}\prod_{j=n-2}^{2} d\phi_j\,dt\,d\theta \\
&\quad= \int_{-1}^{1}\int_0^{2\pi}\int_0^{\pi} \times \overset{(n-1)}{\cdots} \times \int_0^{\pi} f\Big(t,\cos\phi_2(1-t^2)^{\frac{1}{2}},\ldots,\cos\theta\prod_{k=2}^{n-2}\sin\phi_k(1-t^2)^{\frac{1}{2}}\Big) \\
&\qquad\times \prod_{k=1}^{n-3}\sin^k\phi_{n-1-k}(1-t^2)^{\frac{n-3}{2}}\prod_{j=n-2}^{2} d\phi_j\,d\theta\,dt \\
&\quad\overset{(3.44)}{=} \int_{-1}^{1}\int_{\Omega_{n-2}} f\left(t,(1-t^2)^{\frac{1}{2}}\xi_1,\ldots,(1-t^2)^{\frac{1}{2}}\xi_{n-1}\right)(1-t^2)^{\frac{n-3}{2}} \\
&\qquad\times\, d(\xi_1,\ldots,\xi_{n-1})\,dt \\
&\quad= \int_{-1}^{1}\int_{\Omega_{n-2}} f\left(t,(1-t^2)^{\frac{1}{2}}\boldsymbol{\xi}\right)(1-t^2)^{\frac{n-3}{2}}\,\omega_{n-2}(d\boldsymbol{\xi})\,dt.
\end{aligned}
$$

So, for the q-dimensional sphere Ω_q, equation (3.17) follows. Note that as the parametrization (3.44) is invariant to coordinates permutations and t can be placed in any argument of the function. The rest of the arguments will remain having the entries $(1-t^2)^{\frac{n-3}{2}}\boldsymbol{\xi}$.

This expression can be improved using an adequate basis representation. From a fixed point $\mathbf{y}\in\Omega_q$, it is possible to complete an orthonormal basis of $\mathbb{R}^{q+1}$, say $\{\mathbf{y},\mathbf{b}_1,\ldots,\mathbf{b}_q\}$. So an element $\mathbf{x}\in\Omega_q$ will be expressed as:

$$
\mathbf{x} = \langle\mathbf{x},\mathbf{y}\rangle\mathbf{y} + \sum_{i=1}^{q}\langle\mathbf{x},\mathbf{b}_i\rangle\mathbf{b}_i = t\mathbf{y} + (1-t^2)^{\frac{1}{2}}\boldsymbol{\xi},
$$

where $t=\langle\mathbf{x},\mathbf{y}\rangle\in[-1,1]$ and $\boldsymbol{\xi}\in T_{\mathbf{y}}=\{\boldsymbol{\eta}\in\Omega_q:\boldsymbol{\eta}\perp\mathbf{y}\}$. Related to the basis $\{\mathbf{y},\mathbf{b}_1,\ldots,\mathbf{b}_q\}$, there are the orthogonal matrix $\mathbf{B}=(\mathbf{y},\mathbf{b}_1,\ldots,\mathbf{b}_q)_{(q+1)\times(q+1)}$ and the semi-orthogonal matrix $\mathbf{B}_{\mathbf{y}}=(\mathbf{b}_1,\ldots,\mathbf{b}_q)_{(q+1)\times q}$. Using the fact that $\mathbf{B}$ is an orthonormal matrix, is possible to make the change $\mathbf{x}=\mathbf{B}\mathbf{z}$, with $\det\mathbf{B}=1$ and $\mathbf{B}^{-1}\Omega_q=\mathbf{B}^T\Omega_q=\Omega_q$ (as $\mathbf{B}$ preserves distances). Then, the relation (3.18) holds:

$$
\int_{\Omega_q} f(\mathbf{x})\,\omega_q(d\mathbf{x}) = \int_{\mathbf{B}^{-1}\Omega_q} f(\mathbf{B}\mathbf{z})\det\mathbf{B}\,\omega_q(d\mathbf{z})
$$

$$
\begin{aligned}
&= \int_{\Omega_q} f(\mathbf{B}\mathbf{z})\, \omega_q(d\mathbf{z}) \\
&\overset{(3.17)}{=} \int_{-1}^{1} \int_{\Omega_{q-1}} f\Big(\mathbf{B}(t, (1-t^2)^{\frac{1}{2}}\boldsymbol{\xi})^T\Big)(1-t^2)^{\frac{q}{2}-1}\, \omega_{q-1}(d\boldsymbol{\xi})\, dt \\
&= \int_{-1}^{1} \int_{\Omega_{q-1}} f\left(t\mathbf{y} + (1-t^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{y}}\boldsymbol{\xi}\right)(1-t^2)^{\frac{q}{2}-1}\, \omega_{q-1}(d\boldsymbol{\xi})\, dt.
\end{aligned}
$$

□

Proof of Lemma 3.3. Without loss of generality, assume that, by the q-spherical coordinates (3.44), $x_i = \cos\phi_1$ and $x_j = \cos\phi_2 \sin\phi_1$. Using this, the calculus are straightforward for the integrands x_i and $x_i x_j$ (it is assumed that only the terms with positive index are taken into account in the products):

$$
\begin{aligned}
\int_{\Omega_q} x_i\, \omega_q(d\mathbf{x}) &= \int_0^{2\pi} \int_0^{\pi} \overset{(q-1)}{\cdots} \int_0^{\pi} \cos\phi_1 \prod_{k=1}^{q-2} \sin^k \phi_{q-k} \sin^{q-1}\phi_1 \prod_{j=q-1}^{1} d\phi_j\, d\theta \\
&= \int_0^{2\pi} \int_0^{\pi} \overset{(q-2)}{\cdots} \int_0^{\pi} \prod_{k=1}^{q-2} \sin^k \phi_{q-k} \prod_{j=q-1}^{2} d\phi_j\, d\theta \times \int_0^{\pi} \cos\phi_1 \sin^{q-1}\phi_1\, d\phi_1 \\
&= \omega_{q-1} \times 0 = 0, \\
\int_{\Omega_q} x_i x_j\, \omega_q(d\mathbf{x}) &= \int_0^{2\pi} \int_0^{\pi} \overset{(q-1)}{\cdots} \int_0^{\pi} \cos\phi_1 \cos\phi_2 \sin\phi_1 \prod_{k=1}^{q-3} \sin^k \phi_{q-k} \sin^{q-2}\phi_2 \sin^{q-1}\phi_1 \\
&\quad \times \prod_{j=q-1}^{1} d\phi_j\, d\theta \\
&= \int_0^{2\pi} \int_0^{\pi} \overset{(q-3)}{\cdots} \int_0^{\pi} \prod_{k=1}^{q-3} \sin^k \phi_{q-k} \prod_{j=q-1}^{3} d\phi_j\, d\theta \\
&\quad \times \int_0^{\pi} \cos\phi_1 \sin^q \phi_1\, d\phi_1 \int_0^{\pi} \cos\phi_2 \sin^{q-2}\phi_2\, d\phi_2 \\
&= \omega_{q-2} \times 0 \times 0 = 0.
\end{aligned}
$$

The integrand x_i^2 is even simpler, using the fact that the integration is over Ω_q:

$$
\int_{\Omega_q} x_i^2\, \omega_q(d\mathbf{x}) = \frac{1}{q+1} \sum_{k=1}^{q+1} \int_{\Omega_q} x_k^2\, \omega_q(d\mathbf{x}) = \frac{1}{q+1} \int_{\Omega_q} \sum_{k=1}^{q+1} x_k^2\, \omega_q(d\mathbf{x}) = \frac{\omega_q}{q+1}.
$$

□

Proof of Lemma 3.4. For $a = 1, 2$, $p = 0, 1$ and $q \geq 1$, the properties of the Gamma function ensure that

$$
\int_0^{\infty} L^a(r) r^{\frac{q}{2}-p}\, dr = \int_0^{\infty} e^{-ar} r^{\frac{q}{2}-p}\, dr = \frac{\Gamma\left(\frac{q}{2} - p + 1\right)}{a^{\frac{q}{2}-p+1}}.
$$

Therefore:

$$
\lambda_q(L) = 2^{\frac{q}{2}-1} \frac{2\pi^{\frac{q}{2}}}{\Gamma\left(\frac{q}{2}\right)} \Gamma\left(\frac{q}{2}\right) = (2\pi)^{\frac{q}{2}},\ b_q(L) = \Gamma\left(\frac{q}{2}\right) \frac{q}{2} \Big/ \Gamma\left(\frac{q}{2}\right) = \frac{q}{2},\ d_q(L) = \frac{\Gamma\left(\frac{q}{2}\right)}{2^{\frac{q}{2}}} \Big/ \Gamma\left(\frac{q}{2}\right) = 2^{-\frac{q}{2}}.
$$

The expression for $c_{h,q}(L)$ arises from the fact that $c_{h,q}(L) = C_q\left(1/h^2\right) e^{1/h^2}$. □

Proof of Lemma 3.5. This proof is a rebuild of the one given in Zhao and Wu (2001) and is included for the aim of completeness of this work. Furthermore, many techniques used in this proof are also helpful for the proofs of other results in this paper.

Let denote $\text{Bias}\big[\hat{f}_h(\mathbf{x})\big] = \mathbb{E}\big[\hat{f}_h(\mathbf{x})\big] - f(\mathbf{x})$. To compute the bias, use Lemma 3.2 for the change of variables with the orthonormal and semi-orthonormal matrices $\mathbf{B} = (\mathbf{x}, \mathbf{b}_1, \ldots, \mathbf{b}_q)$ and $\mathbf{B}_\mathbf{x} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)$, and then apply the ordinary change of variables

$$r = \frac{1-t}{h^2}, \quad dr = -h^{-2}\, dt. \tag{3.46}$$

This results in:

$$\begin{aligned}
\text{Bias}\left[\hat{f}_h(\mathbf{x})\right] &= c_{h,q}(L)\mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right] - f(\mathbf{x}) \\
&= c_{h,q}(L)\int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) f(\mathbf{y})\,\omega_q(d\mathbf{y}) - c_{h,q}(L)\int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right)\omega_q(d\mathbf{y}) f(\mathbf{x}) \\
&= c_{h,q}(L)\int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right)(f(\mathbf{y}) - f(\mathbf{x}))\,\omega_q(d\mathbf{y}) \\
&= c_{h,q}(L)\int_{-1}^{1}\int_{\Omega_{q-1}} L\left(\frac{1-t}{h^2}\right)\left(f\left(t\mathbf{x} + (1-t^2)^{\frac{1}{2}}\mathbf{B}_\mathbf{x}\boldsymbol{\xi}\right) - f(\mathbf{x})\right) \\
&\quad \times (1-t^2)^{\frac{q}{2}-1}\,\omega_{q-1}(d\boldsymbol{\xi})\,dt \\
&\overset{(3.46)}{=} c_{h,q}(L)h^q\int_0^{2h^{-2}}\int_{\Omega_{q-1}} L(r)\left(f\left(\mathbf{x} + \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right) - f(\mathbf{x})\right) r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}\,\omega_{q-1}(d\boldsymbol{\xi})\,dr \\
&= c_{h,q}(L)h^q\int_0^{2h^{-2}} L(r) r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}\int_{\Omega_{q-1}}\left(f\left(\mathbf{x} + \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right) - f(\mathbf{x})\right) \\
&\quad \times\, \omega_{q-1}(d\boldsymbol{\xi})\,dr,
\end{aligned} \tag{3.47}$$

where $\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}} = -rh^2\mathbf{x} + h\left[r(2-h^2r)\right]^{\frac{1}{2}}\mathbf{B}_\mathbf{x}\boldsymbol{\xi} \in \Omega_q$. By condition **D1**, the Taylor expansion of f at $\mathbf{x}$ is

$$f(\mathbf{x} + \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}) - f(\mathbf{x}) = \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T\boldsymbol{\nabla} f(\mathbf{x}) + \frac{1}{2}\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T\boldsymbol{\mathcal{H}} f(\mathbf{x})\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}} + o\left(\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right),$$

so the calculus of (3.47) can be split in three parts. For the first use that the integration of ξ_i vanishes by Lemma 3.3:

$$\begin{aligned}
\int_{\Omega_{q-1}} \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T\boldsymbol{\nabla} f(\mathbf{x})\,\omega_{q-1}(d\boldsymbol{\xi}) &= -rh^2\int_{\Omega_{q-1}} \mathbf{x}^T\boldsymbol{\nabla} f(\mathbf{x})\,\omega_{q-1}(d\boldsymbol{\xi}) \\
&\quad + h\left[r(2-h^2r)\right]^{\frac{1}{2}}\int_{\Omega_{q-1}} \boldsymbol{\xi}^T\mathbf{B}_\mathbf{x}^T\boldsymbol{\nabla} f(\mathbf{x})\,\omega_{q-1}(d\boldsymbol{\xi}) \\
&= -rh^2\omega_{q-1}\mathbf{x}^T\boldsymbol{\nabla} f(\mathbf{x})
\end{aligned} \tag{3.48}$$

In the second, by the results of Lemma 3.3,

$$\int_{\Omega_{q-1}} \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T\boldsymbol{\mathcal{H}} f(\mathbf{x})\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\,\omega_{q-1}(d\boldsymbol{\xi}) = r^2h^4\int_{\Omega_{q-1}} \mathbf{x}^T\boldsymbol{\mathcal{H}} f(\mathbf{x})\mathbf{x}\,\omega_{q-1}(d\boldsymbol{\xi})$$

$$
\begin{aligned}
&- 2rh^3\left[r(2-h^2r)\right]^{\frac{1}{2}}\int_{\Omega_{q-1}}\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\,\omega_{q-1}(d\boldsymbol{\xi})\\
&+ h^2r(2-h^2r)\int_{\Omega_{q-1}}\boldsymbol{\xi}^T\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\,\omega_{q-1}(d\boldsymbol{\xi})\\
=&\, r^2h^4\omega_{q-1}\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\\
&+ h^2r(2-h^2r)\int_{\Omega_{q-1}}\sum_{i,j=1}^{q}\mathbf{b}_i^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{b}_j\xi_i\xi_j\,\omega_{q-1}(d\boldsymbol{\xi})\\
=&\, r^2h^4\omega_{q-1}\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\\
&+ h^2r(2-h^2r)\sum_{i=1}^{q}\mathbf{b}_i^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{b}_i\int_{\Omega_{q-1}}\xi_i^2\,\omega_{q-1}(d\boldsymbol{\xi})\\
=&\, r^2h^4\omega_{q-1}\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\\
&+ h^2r(2-h^2r)\omega_{q-1}q^{-1}\left[\nabla^2f(\mathbf{x})-\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\right]. \qquad (3.49)
\end{aligned}
$$

In the last step it is used that by $\sum_{i=1}^q \mathbf{b}_i\mathbf{b}_i^T + \mathbf{x}\mathbf{x}^T = \mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T = \mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T$,

$$
\sum_{i=1}^{q}\mathbf{b}_i^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{b}_i = \operatorname{tr}\left[\boldsymbol{\mathcal{H}}f(\mathbf{x})\sum_{i=1}^{q}\mathbf{b}_i\mathbf{b}_i^T\right] = \operatorname{tr}\left[\boldsymbol{\mathcal{H}}f(\mathbf{x})\left(\mathbf{I}_{q+1}-\mathbf{x}\mathbf{x}^T\right)\right] = \nabla^2f(\mathbf{x}) - \mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}.
$$

Apart from this, the order of the Taylor expansion is

$$
o\left(\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right) = o\left(r^2h^4 + h^2r(2-h^2r)\right) = o\left(r^2h^4 + 2h^2r - h^4r^2\right) = r\,o\left(h^2\right). \qquad (3.50)
$$

Adding (3.48)–(3.50),

$$
\begin{aligned}
(3.47) =&\, \omega_{q-1}c_{h,q}(L)h^q\int_0^{2h^{-2}} L(r)r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}\Bigg\{-rh^2\mathbf{x}^T\boldsymbol{\nabla}f(\mathbf{x}) + \frac{r^2h^4}{2}\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\\
&+ \frac{h^2r(2-h^2r)}{2q}\left(\nabla^2f(\mathbf{x})-\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\right) + r\,o\left(h^2\right)\Bigg\}dr\\
=&\, -h^2\omega_{q-1}\left[\int_0^{2h^{-2}} c_{h,q}(L)h^qL(r)r^{\frac{q}{2}}(2-h^2r)^{\frac{q}{2}-1}\,dr\right]\mathbf{x}^T\boldsymbol{\nabla}f(\mathbf{x})\\
&+ \frac{h^4\omega_{q-1}}{2}\left[\int_0^{2h^{-2}} c_{h,q}(L)h^qL(r)r^{\frac{q}{2}+1}(2-h^2r)^{\frac{q}{2}-1}\,dr\right]\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\\
&+ \frac{h^2\omega_{q-1}}{2}\left[\int_0^{2h^{-2}} c_{h,q}(L)h^qL(r)r^{\frac{q}{2}}(2-h^2r)^{\frac{q}{2}}\,dr\right]q^{-1}\left(\nabla^2f(\mathbf{x})-\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\right)\\
&+ \omega_{q-1}\left[\int_0^{2h^{-2}} c_{h,q}(L)h^qL(r)r^{\frac{q}{2}}(2-h^2r)^{\frac{q}{2}-1}\,dr\right]o\left(h^2\right). \qquad (3.51)
\end{aligned}
$$

Consider the following functions for $h>0$ and $i,j=0,1$:

$$
\varphi_{h,i,j}(r) = c_{h,q}(L)h^qL(r)r^{\frac{q}{2}+i}(2-h^2r)^{\frac{q}{2}-j}\mathbf{1}_{[0,2h^{-2})}(r), \quad r\in[0,\infty).
$$

When $n\to\infty$, $h\to 0$ and the limit of $\varphi_{h,i,j}$ is given by

$$
\varphi_{i,j}(r) = \lim_{h\to 0}\varphi_{h,i,j}(r) = \lambda_q(L)^{-1}L(r)r^{\frac{q}{2}+i}2^{\frac{q}{2}-j}\mathbf{1}_{[0,\infty)}(r).
$$

Then, by Remark 3.5 and Lemma 3.1:

$$\lim_{h\to 0}\int_0^\infty \varphi_h(r)\,dr = \lambda_q(L)^{-1}2^{\frac{q}{2}-j}\int_0^\infty L(r)r^{\frac{q}{2}-i}\,dr \overset{(3.16)}{=} \begin{cases} \frac{2^{1-j}}{\omega_{q-1}}b_q(L), & i=0,\\ \frac{2^{1-j}}{\omega_{q-1}}\frac{\int_0^\infty L(r)r^{\frac{q}{2}+1}\,dr}{\int_0^\infty L(r)r^{\frac{q}{2}-1}\,dr}, & i=1. \end{cases}$$

So, for the terms between square brackets of (3.51), $\int_0^\infty \varphi_h(r)\,dr = \int_0^\infty \varphi(r)\,dr\,(1+o(1))$. Replacing this in (3.51) leads to

$$\begin{aligned}(3.51) =& -h^2\omega_{q-1}\left[\frac{b_q(L)}{\omega_{q-1}}+o(1)\right]\mathbf{x}^T\boldsymbol{\nabla}f(\mathbf{x})\\ &+\frac{h^4\omega_{q-1}}{2}\left[\frac{b_q(L)}{\omega_{q-1}}\frac{\int_0^\infty L(r)r^{\frac{q}{2}+1}\,dr}{\int_0^\infty L(r)r^{\frac{q}{2}}\,dr}+o(1)\right]\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\\ &+\frac{h^2\omega_{q-1}}{2}\left[\frac{b_q(L)}{\omega_{q-1}}+o(1)\right]q^{-1}\left(\nabla^2 f(\mathbf{x})-\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\right)+\omega_{q-1}\left[\frac{b_q(L)}{\omega_{q-1}}+o(1)\right]o\left(h^2\right)\\ =&\,h^2b_q(L)\left[-\mathbf{x}^T\boldsymbol{\nabla}f(\mathbf{x})+q^{-1}\left(\nabla^2 f(\mathbf{x})-\mathbf{x}^T\boldsymbol{\mathcal{H}}f(\mathbf{x})\mathbf{x}\right)\right]+\mathcal{O}\left(h^4\right)+o\left(h^2\right)\\ =&\,h^2b_q(L)\Psi(f,\mathbf{x})+o\left(h^2\right).\end{aligned}$$

□

References

Akaike, H. (1954). An approximation to the density function. *Ann. Inst. Statist. Math.*, 6(2):127–132.

Bai, Z. D., Rao, C. R., and Zhao, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.*, 27(1):24–39.

Blumenson, L. E. (1960). Classroom notes: a derivation of n-dimensional spherical coordinates. *Amer. Math. Monthly*, 67(1):63–66.

Cao, R. (1993). Bootstrapping the mean integrated squared error. *J. Multivariate Anal.*, 45(1):137–160.

Cao, R., Cuevas, A., and Gonzalez Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Comput. Statist. Data Anal.*, 17(2):153–176.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2011). Kernel density estimation on the torus. *J. Statist. Plann. Inference*, 141(6):2156–2173.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013). Exploring wind direction and SO_2 concentration by circular-linear density estimation. *Stoch. Environ. Res. Risk Assess.*, 27(5):1055–1067.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.*, 14(1):1–16.

Hall, P., Watson, G. S., and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762.

Hendriks, H. (1990). Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *Ann. Statist.*, 18(2):832–849.

Henry, G. and Rodriguez, D. (2009). Kernel density estimation on Riemannian manifolds: asymptotic results. *J. Math. Imaging Vision*, 34(3):235–239.

Jones, C., Marron, J. S., and Sheather, S. J. (1996). Progress in data-based bandwidth selection for kernel density estimation. *Computation. Stat.*, (11):337–381.

Jupp, P. E. and Mardia, K. V. (1989). A unified view of the theory of directional statistics, 1975-1988. *Int. Stat. Rev.*, 57(3):261–294.

Klemelä, J. (2000). Estimation of densities and derivatives of densities with directional data. *J. Multivariate Anal.*, 73(1):18–40.

Marron, J. S. and Wand, M. P. (1992). Exact mean integrated squared error. *Ann. Statist.*, 20(2):712–736.

Müller, H.-G. (2006). Density estimation-II. In Kotz, S., Balakrishnan, N., Read, C., and Vidakovic, B., editors, *Encyclopedia of statistical sciences*, volume 2, pages 1611–1626. John Wiley & Sons, Hoboken, second edition.

Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2012). A plug-in rule for bandwidth selection in circular density estimation. *Comput. Statist. Data Anal.*, 56(12):3898–3908.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33(3):1065–1076.

Pelletier, B. (2005). Kernel density estimation on Riemannian manifolds. *Statist. Probab. Lett.*, 73(3):297–304.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27(3):832–837.

Scott, D. W. (1992). *Multivariate density estimation*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. John Wiley & Sons, New York.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B*, 53(3):683–690.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Taylor, C. C. (2008). Automatic bandwidth selection for circular density estimation. *Comput. Statist. Data Anal.*, 52(7):3493–3500.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Watson, G. S. (1983). *Statistics on spheres*, volume 6 of *University of Arkansas Lecture Notes in the Mathematical Sciences*. John Wiley & Sons, New York.

Zhao, L. and Wu, C. (2001). Central limit theorem for integrated square error of kernel estimators of spherical density. *Sci. China Ser. A*, 44(4):474–483.

Chapter 4

Bandwidth selectors for kernel density estimation with directional data

Abstract

New bandwidth selectors for kernel density estimation with directional data are presented in this work. These selectors are based on asymptotic and exact error expressions for the kernel density estimator combined with mixtures of von Mises distributions. The performance of the proposed selectors is investigated in a simulation study and compared with other existing rules for a large variety of directional scenarios, sample sizes and dimensions. The selector based on the exact error expression turns out to have the best behaviour of the studied selectors for almost all the situations. This selector is illustrated with real data for the circular and spherical cases.

Reference

García-Portugués (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron. J. Stat.*, 7(2013):1655–1685.

Contents

4.1 Introduction **80**
4.2 Kernel density estimation with directional data **81**
4.2.1 Available bandwidth selectors 84
4.3 A new rule of thumb selector **85**
4.4 Selectors based on mixtures **86**
4.4.1 Mixtures fitting and selection of the number of components 89
4.5 Comparative study **90**
4.5.1 Directional models 90
4.5.2 Circular case 92
4.5.3 Spherical case 94
4.5.4 The effect of dimension 96
4.6 Data application **97**
4.6.1 Wind direction 97
4.6.2 Position of stars 97

4.7 Conclusions **98**
4.A Proofs **99**
4.B Models for the simulation study **99**
4.C Extended tables for the simulation study **100**
References **104**

4.1 Introduction

Bandwidth selection is a key issue in kernel density estimation that has deserved considerable attention during the last decades. The problem of selecting the most suitable bandwidth for the nonparametric kernel density estimator introduced by Rosenblatt (1956) and Parzen (1962) is the main topic of the reviews of Cao et al. (1994), Jones et al. (1996) and Chiu (1996), among others. Comprehensive references on kernel smoothing and bandwidth selection include the books by Silverman (1986), Scott (1992) and Wand and Jones (1995). Bandwidth selection is still an active research field in density estimation, with some recent contributions like Horová et al. (2013) and Chacón and Duong (2013) in the last years.

Kernel density estimation has been also adapted to directional data, that is, data in the unit hypersphere of dimension q. Due to the particular nature of directional data (periodicity for $q = 1$ and manifold structure for any q), the usual multivariate techniques are not appropriate and specific methodology that accounts for their characteristics has to be considered. The classical references for the theory of directional statistics are the complete review of Jupp and Mardia (1989) and the book by Mardia and Jupp (2000). The kernel density estimation with directional data was firstly proposed by Hall et al. (1987), studying the properties of two types of kernel density estimators and providing cross-validatory bandwidth selectors. Almost simultaneously, Bai et al. (1988) provided a similar definition of kernel estimator, establishing its pointwise and $\mathcal{L}_1$ consistency. Some of the results by Hall et al. (1987) were extended by Klemelä (2000), who studied the estimation of the Laplacian of the density and other types of derivatives. Whereas the framework for all these references is the general q-sphere, which comprises as particular case the circle ($q = 1$), there exists a remarkable collection of works devoted to kernel density estimation and bandwidth selection for the circular scenario. Specifically, Taylor (2008) presented the first plug-in bandwidth selector in this context and Oliveira et al. (2012) derived a selector based on mixtures and on the results of Di Marzio et al. (2009) for the circular Asymptotic Mean Integrated Squared Error (AMISE). Recently, Di Marzio et al. (2011) proposed a product kernel density estimator on the q-dimensional torus and cross-validatory bandwidth selection methods for that situation. Another nonparametric approximation for density estimation with circular data was given in Fernández-Durán (2004) and Fernández-Durán and Gregorio-Domínguez (2010). In the general setting of spherical random fields Durastanti et al. (2013) derived an estimation method based on a needlet basis representation.

Directional data arise in many applied fields. For the circular case ($q = 1$) a typical example is wind direction, studied among others in Jammalamadaka and Lund (2006), Fernández-Durán (2007) and García-Portugués et al. (2013a). The spherical case ($q = 2$) poses challenging applications in astronomy, for example in the study of stars position in the celestial sphere or in the study of the cosmic microwave background radiation (Cabella and Marinucci, 2009). Finally, a novel field where directional data is present for large q is text mining (Banerjee et al., 2005),

where documents are usually codified as high dimensional unit vectors. For all these situations, a reliable method for choosing the bandwidth parameter seems necessary to trust the density estimate.

The aim of this work is to introduce new bandwidth selectors for the kernel density estimator for directional data. The first one is a rule of thumb which assumes that the underlying density is a von Mises and it is intended to be the directional analogue of the rule of thumb proposed by Silverman (1986) for data in the real line. This selector uses the AMISE expression that can be seen, among others, in García-Portugués et al. (2013b). The novelty of the selector is that it is more general and robust than the previous proposal by Taylor (2008), although both rules exhibit an unsatisfactory behaviour when the reference density spreads off from the von Mises. To overcome this problem, two new selectors based on the use of mixtures of von Mises for the reference density are proposed. One of them uses the aforementioned AMISE expression, whereas the other one uses the exact MISE computation for mixtures of von Mises densities given in García-Portugués et al. (2013b). Both of them use the Expectation-Maximization algorithm of Banerjee et al. (2005) to fit the mixtures and, to select the number of components, the BIC criteria is employed. These selectors based on mixtures are inspired by the earlier ideas of Ćwik and Koronacki (1997), for the multivariate setting, and Oliveira et al. (2012) for the circular scenario.

This paper is organized as follows. Section 4.2 presents some background on kernel density estimation for directional data and the available bandwidth selectors. The rule of thumb selector is introduced in Section 4.3 and the two selectors based on mixtures of von Mises are presented in Section 4.4. Section 4.5 contains a simulation study comparing the proposed selectors with the ones available in the literature. Finally, Section 4.6 illustrates a real data application and some conclusions are given in Section 4.7. Supplementary materials with proofs, simulated models and extended tables are given in the appendix.

4.2 Kernel density estimation with directional data

Denote by $\mathbf{X}$ a directional random variable with density f. The support of such variable is the q-dimensional sphere, namely $\Omega_q = \{\mathbf{x} \in \mathbb{R}^{q+1} : x_1^2 + \cdots + x_{q+1}^2 = 1\}$, endowed with the Lebesgue measure in Ω_q, that will be denoted by ω_q. Then, a directional density is a nonnegative function that satisfies $\int_{\Omega_q} f(\mathbf{x})\, \omega_q(d\mathbf{x}) = 1$. Also, when there is no possible confusion, the area of Ω_q will be denoted by

$$\omega_q = \omega_q(\Omega_q) = \frac{2\pi^{\frac{q+1}{2}}}{\Gamma\left(\frac{q+1}{2}\right)}, \quad q \geq 1,$$

where Γ represents the Gamma function defined as $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x}\, dx$, $p > -1$.

Among the directional distributions, the von Mises-Fisher distribution (see Watson (1983)) is perhaps the most widely used. The von Mises density, denoted by $\mathrm{vM}(\boldsymbol{\mu}, \kappa)$, is given by

$$f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = C_q(\kappa) \exp\left\{\kappa \mathbf{x}^T \boldsymbol{\mu}\right\}, \quad C_q(\kappa) = \frac{\kappa^{\frac{q-1}{2}}}{(2\pi)^{\frac{q+1}{2}} \mathcal{I}_{\frac{q-1}{2}}(\kappa)}, \tag{4.1}$$

where $\boldsymbol{\mu} \in \Omega_q$ is the directional mean, $\kappa \geq 0$ the concentration parameter around the mean, ^{T} stands for the transpose operator and $\mathcal{I}_p$ is the modified Bessel function of order ν,

$$\mathcal{I}_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\pi^{1/2}\Gamma\left(\nu+\frac{1}{2}\right)} \int_{-1}^{1} (1-t^2)^{\nu-\frac{1}{2}} e^{zt}\, dt.$$

This distribution is the main reference for directional models and, in that sense, plays the role of the normal distribution for directional data (is also a multivariate normal $\mathcal{N}(\boldsymbol{\mu}, \kappa^{-1}\mathbf{I}_{q+1})$ conditioned on Ω_q; see Mardia and Jupp (2000)). A particular case of this density sets $\kappa = 0$, which corresponds to the uniform density that assigns probability ω_q^{-1} to any direction in Ω_q.

Given a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from the directional random variable $\mathbf{X}$, the proposal of Bai et al. (1988) for the directional kernel density estimator at a point $\mathbf{x} \in \Omega_q$ is

$$\hat{f}_h(\mathbf{x}) = \frac{c_{h,q}(L)}{n} \sum_{i=1}^{n} L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right), \tag{4.2}$$

where L is a directional kernel (a rapidly decaying function with nonnegative values and defined in $[0,\infty)$), $h > 0$ is the bandwidth parameter and $c_{h,q}(L)$ is a normalizing constant. This constant is needed in order to ensure that the estimator is indeed a density and satisfies that

$$c_{h,q}(L)^{-1} = \int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) \omega_q(d\mathbf{x}) = \mathcal{O}\left(h^q\right).$$

As usual in kernel smoothing, the selection of the bandwidth is a crucial step that affects notably the final estimation: large values of h result in a uniform density in the sphere, whereas small values of h provide an undersmoothed estimator with high concentrations around the sample observations. On the other hand, the choice of the kernel is not seen as important for practical purposes and the most common choice is the so called von Mises kernel $L(r) = e^{-r}$. Its name is due to the fact that the kernel estimator can be viewed as a mixture of von Mises-Fisher densities as follows:

$$\hat{f}_h(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} f_{\mathrm{vM}}\left(\mathbf{x}; \mathbf{X}_i, 1/h^2\right),$$

where, for each von Mises component, the mean value is the i-th observation $\mathbf{X}_i$ and the common concentration parameter is given by $1/h^2$.

The classical error measurement in kernel density estimation is the $\mathcal{L}_2$ distance between the estimator $\hat{f}_h$ and the target density f, the so called Integrated Squared Error (ISE). As this is a random quantity depending on the sample, its expected value, the Mean Integrated Squared Error (MISE), is usually considered:

$$\mathrm{MISE}(h) = \mathbb{E}\left[\mathrm{ISE}\left[\hat{f}_h\right]\right] = \mathbb{E}\left[\int_{\Omega_q} \left(\hat{f}_h(\mathbf{x}) - f(\mathbf{x})\right)^2 \omega_q(d\mathbf{x})\right],$$

which depends on the bandwidth h, the kernel L, the sample size n and the target density f. Whereas the two last elements are fixed when estimating a density from a random sample, the bandwidth has to be chosen (also the kernel, although this does not present a big impact in the

performance of the estimator). Then, a possibility is to search for the bandwidth that minimizes the MISE:

$$h_{\text{MISE}} = \arg\min_{h>0} \text{MISE}(h).$$

To derive an easier form for the MISE that allows to obtain h_{MISE}, the following conditions on the elements of the estimator (4.2) are required:

D1. Extend f from Ω_q to $\mathbb{R}^{q+1}\setminus\{\mathbf{0}\}$ by $f(\mathbf{x}) \equiv f(\mathbf{x}/||\mathbf{x}||)$ for all $\mathbf{x} \in \mathbb{R}^{q+1}\setminus\{\mathbf{0}\}$, where $||\cdot||$ denotes the Euclidean norm. Assume that the gradient vector $\boldsymbol{\nabla} f(\mathbf{x})$ and the Hessian matrix $\boldsymbol{\mathcal{H}} f(\mathbf{x})$ exist, are continuous and square integrable.

D2. Assume that $L : [0,\infty) \to [0,\infty)$ is a bounded and integrable function such that $0 < \int_0^\infty L^k(r) r^{\frac{q}{2}-1}\, dr < \infty$, $\forall q \geq 1$, for $k = 1, 2$.

D3. Assume that $h = h_n$ is a positive sequence such that $h_n \to 0$ and $nh_n^q \to \infty$ as $n \to \infty$.

The following result, available from García-Portugués et al. (2013b), provides the MISE expansion for the estimator (4.2). It is worth mentioning that, under similar conditions, Hall et al. (1987) and Klemelä (2000) also derived analogous expressions.

Proposition 4.1 (García-Portugués et al. (2013b))**.** *Under conditions **D1**–**D3**, the MISE for the directional kernel density estimator (4.2) is given by*

$$\text{MISE}(h) = b_q(L)^2 R(\Psi(f,\cdot)) h^4 + \frac{c_{h,q}(L)}{n} d_q(L) + o\left(h^4 + (nh^q)^{-1}\right),$$

where $R(\Psi(f,\cdot)) = \int_{\Omega_q} \Psi(f,\mathbf{x})^2\, \omega_q(d\mathbf{x})$, $b_q(L) = \frac{\int_0^\infty L(r) r^{\frac{q}{2}}\, dr}{\int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr}$, $d_q(L) = \frac{\int_0^\infty L^2(r) r^{\frac{q}{2}}\, dr}{\int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr}$ *and*

$$\Psi(f,\mathbf{x}) = -\mathbf{x}^T \boldsymbol{\nabla} f(\mathbf{x}) + q^{-1}\left(\nabla^2 f(\mathbf{x}) - \mathbf{x}^T \boldsymbol{\mathcal{H}} f(\mathbf{x}) \mathbf{x}\right). \tag{4.3}$$

This results leads to the decomposition $\text{MISE}(h) = \text{AMISE}(h) + o\left(h^4 + (nh^q)^{-1}\right)$, where AMISE stands for the Asymptotic MISE. It is possible to derive an optimal bandwidth for the AMISE in this sense, $h_{\text{AMISE}} = \arg\min_{h>0} \text{AMISE}(h)$, that will be close to h_{MISE} when $h^4 + (nh^q)^{-1}$ is small enough.

Corollary 4.1 (García-Portugués et al. (2013b))**.** *The AMISE optimal bandwidth for the directional kernel density estimator (4.2) is given by*

$$h_{\text{AMISE}} = \left[\frac{q d_q(L)}{4 b_q(L)^2 \lambda_q(L) R(\Psi(f,\cdot)) n}\right]^{\frac{1}{4+q}}, \tag{4.4}$$

where $\lambda_q(L) = 2^{\frac{q}{2}-1}\omega_{q-1} \int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr$.

Unfortunately, expression (4.4) can not be used in practise since it depends on the curvature term $R(\Psi(f,\cdot))$ of the unknown density f.

4.2.1 Available bandwidth selectors

The first proposals for data-driven bandwidth selection with directional data are from Hall et al. (1987), who provide cross-validatory selectors. Specifically, Least Squares Cross-Validation (LSCV) and Likelihood Cross-Validation (LCV) selectors are introduced, arising as the minimizers of the cross-validated estimates of the squared error loss and the Kullback-Leibler loss, respectively. The selectors have the following expressions:

$$h_{\mathrm{LSCV}} = \arg\max_{h>0} \mathrm{CV}_2(h), \quad \mathrm{CV}_2(h) = 2n^{-1}\sum_{i=1}^{n} \hat{f}_h^{-i}(\mathbf{X}_i) - \int_{\Omega_q} \hat{f}_h(\mathbf{x})^2\, \omega_q(d\mathbf{x}),$$

$$h_{\mathrm{LCV}} = \arg\max_{h>0} \mathrm{CV}_{\mathrm{KL}}(h), \quad \mathrm{CV}_{\mathrm{KL}}(h) = \sum_{i=1}^{n} \log \hat{f}_h^{-i}(\mathbf{X}_i),$$

where $\hat{f}_h^{-i}$ represents the kernel estimator computed without the i-th observation. See Remark 4.3 for an efficient computation of h_{LSCV}.

Recently, Taylor (2008) proposed a plug-in selector for the case of circular data ($q = 1$) for the estimator with the von Mises kernel. The selector of Taylor (2008) uses from the beginning the assumption that the reference density is a von Mises to construct the AMISE. This contrasts with the classic rule of thumb selector of Silverman (1986), which supposes at the end (*i.e.*, after deriving the AMISE expression) that the reference density is a normal. The bandwidth parameter is chosen by first obtaining an estimation $\hat{\kappa}$ of the concentration parameter κ in the reference density (for example, by maximum likelihood) and using the formula

$$h_{\mathrm{TAY}} = \left[\frac{4\pi^{\frac{1}{2}}\mathcal{I}_0(\hat{\kappa})^2}{3\hat{\kappa}^2\mathcal{I}_2(2\hat{\kappa})n}\right]^{\frac{1}{5}}.$$

Note that the parametrization of Taylor (2008) has been adapted to the context of the estimator (4.2) by denoting by h the inverse of the squared concentration parameter employed in his paper.

More recently, Oliveira et al. (2012) proposed a selector that improves the performance of Taylor (2008) allowing for more flexibility in the reference density, considering a mixture of von Mises. This selector is also devoted to the circular case and is mainly based on two elements. First, the AMISE expansion that Di Marzio et al. (2009) derived for the circular kernel density estimator by the use of Fourier expansions of the circular kernels. This expression has the following form when the kernel is a circular von Mises (the estimator is equivalent to consider $L(r) = e^{-r}$, $q = 1$ and h as the inverse of the squared concentration parameter in (4.2)):

$$\mathrm{AMISE}(h) = \frac{1}{16}\left[1 - \frac{\mathcal{I}_2(h^{-1/2})}{\mathcal{I}_0\left(h^{-1/2}\right)}\right]^2 \int_0^{2\pi} f''(\theta)^2\, d\theta + \frac{\mathcal{I}_0(2h^{-1/2})}{2n\pi\mathcal{I}_0\left(h^{-1/2}\right)^2}. \tag{4.5}$$

The second element is the Expectation-Maximization (EM) algorithm of Banerjee et al. (2005) for fitting mixtures of directional von Mises. The selector, that will be denoted by h_{OLI}, proceeds as follows

i. Use the EM algorithm to fit mixtures from a determined range of components.

ii. Choose the fitted mixture with minimum AIC.

iii. Compute the curvature term in (4.5) using the fitted mixture and seek for the h that minimizes this expression, that will be h_{OLI}.

4.3 A new rule of thumb selector

Using the properties of the von Mises density it is possible to derive a directional analogue to the rule of thumb of Silverman (1986), which is the optimal AMISE bandwidth for normal reference density and normal kernel. The rule is resumed in the following result.

Proposition 4.2 (Rule of thumb)**.** *The curvature term for a von Mises density* $\mathrm{vM}(\boldsymbol{\mu}, \kappa)$ *is*

$$R(\Psi(f_{\mathrm{vM}}(\cdot;\boldsymbol{\mu},\kappa),\cdot)) = \frac{\kappa^{\frac{q+1}{2}}}{2^{q+2}\pi^{\frac{q+1}{2}}\mathcal{I}_{\frac{q-1}{2}}(\kappa)^2 q}\left[2q\mathcal{I}_{\frac{q+1}{2}}(2\kappa) + (2+q)\mathcal{I}_{\frac{q+3}{2}}(2\kappa)\right].$$

If $\hat{\kappa}$ *is a suitable estimator for* κ*, then the rule of thumb selector for the kernel estimator (4.2) with a directional kernel* L *is*

$$h_{\mathrm{ROT}} = \left[\frac{q^2 d_q(L) 2^{q+2}\pi^{\frac{q+1}{2}}\mathcal{I}_{\frac{q-1}{2}}(\hat{\kappa})^2}{\hat{\kappa}^{\frac{q+1}{2}} 4 b_q(L)^2 \lambda_q(L)\left(2q\mathcal{I}_{\frac{q+1}{2}}(2\hat{\kappa}) + (2+q)\mathcal{I}_{\frac{q+3}{2}}(2\hat{\kappa})\right) n}\right]^{\frac{1}{4+q}}.$$

If L *is the von Mises kernel, then:*

$$h_{\mathrm{ROT}} = \begin{cases} \left[\dfrac{4\pi^{\frac{1}{2}}\mathcal{I}_0(\hat{\kappa})^2}{\hat{\kappa}\left[2\mathcal{I}_1(2\hat{\kappa}) + 3\hat{\kappa}\mathcal{I}_2(2\hat{\kappa})\right] n}\right]^{\frac{1}{5}}, & q = 1, \\ \left[\dfrac{8\sinh^2(\hat{\kappa})}{\hat{\kappa}\left[(1+4\hat{\kappa}^2)\sinh(2\hat{\kappa}) - 2\hat{\kappa}\cosh(2\hat{\kappa})\right] n}\right]^{\frac{1}{6}}, & q = 2, \\ \left[\dfrac{4\pi^{\frac{1}{2}}\mathcal{I}_{\frac{q-1}{2}}(\hat{\kappa})^2}{\hat{\kappa}^{\frac{q+1}{2}}\left[2q\mathcal{I}_{\frac{q+1}{2}}(2\hat{\kappa}) + (2+q)\hat{\kappa}\mathcal{I}_{\frac{q+3}{2}}(2\hat{\kappa})\right] n}\right]^{\frac{1}{4+q}}, & q \geq 3. \end{cases} \tag{4.6}$$

The parameter κ *can be estimated by maximum likelihood.*

In view of the expression for h_{ROT} in (4.6), it is interesting to compare it with h_{TAY} when $q = 1$. As it can be seen, both selectors coincide except for one difference: the term $2\mathcal{I}_1(2\hat{\kappa})$ in the sum in the denominator of h_{ROT}. This "extra term" can be explained by examining the way that both selectors are derived. Whereas the selector h_{ROT} derives the bandwidth supposing that the reference density is a von Mises when the AMISE is already derived in a general way, the selector h_{TAY} uses the von Mises assumption to compute it. Therefore, it is expected that the selector h_{ROT} will be more robust against deviations from the von Mises density.

Figure 4.1 collects two graphs exposing these comments, that are also corroborated in Section 4.5. The left plot shows the MISE for h_{TAY} and h_{ROT} for the density $\frac{1}{2}\mathrm{vM}\left((0,1),2\right) + \frac{1}{2}\mathrm{vM}\left((\cos(\theta),\sin(\theta)),2\right)$, where $\theta \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$. This model represents two equally concentrated von Mises densities that spread off from being the same to being antipodal. As it can be seen, the h_{ROT} selector is slightly more accurate when the von Mises model holds ($\theta = \frac{\pi}{2}$) and when the deviation is large ($\theta \in \left[\pi, \frac{3\pi}{2}\right]$). When $\theta \in \left[\frac{\pi}{2}, \pi\right]$, both selectors perform similar. This graph also illustrates the main problem of these selectors: the von Mises density is not flexible enough to capture densities with multimodality and it approximates them by the flat uniform density.

When the density is a vM($\boldsymbol{\mu}, \kappa$), the right plot of Figure 4.1 shows the output of h_{TAY}, h_{ROT}, h_{MISE} and their corresponding errors with respect to κ. The effect of the "extra term" is visible for low values of κ, where MISE(h_{TAY}) presents a local maxima. This corresponds with higher values of h_{TAY} with respect to h_{ROT} and h_{MISE}, which means that the former produces oversmoothed estimations of the density (*i.e.* tend to the uniform case faster). Despite the worse behaviour of h_{TAY}, when the concentration parameter increases the effect of the "extra term" is mitigated and both selectors are almost the same.

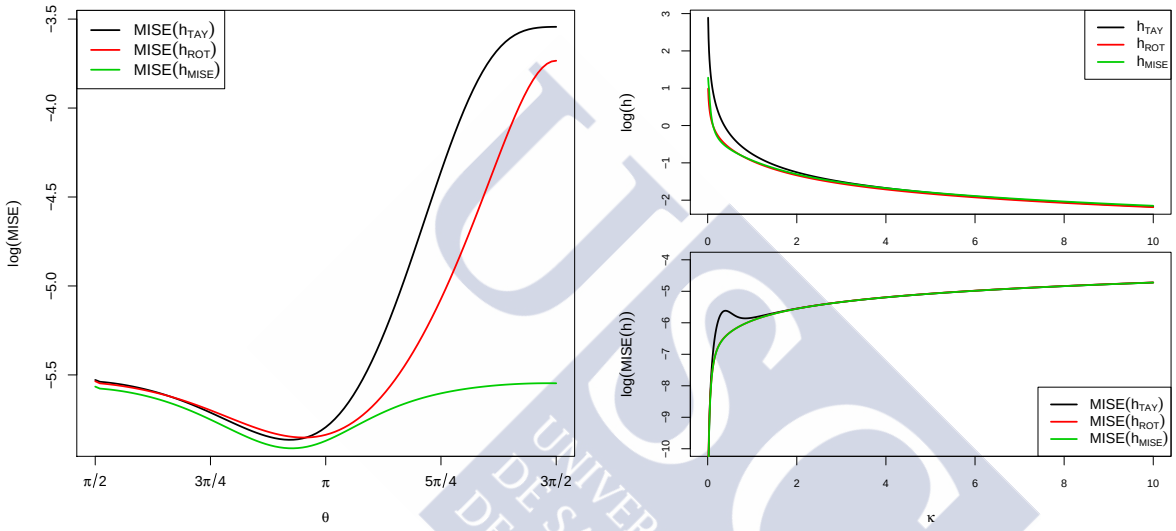


Figure 4.1: The effect of the "extra term" in h_{ROT}. Left plot: logarithm of the curves of MISE(h_{TAY}), MISE(h_{ROT}) and MISE(h_{MISE}) for sample size $n = 250$. The curves are computed by 1000 Monte Carlo samples and h_{MISE} is obtained exactly. The abscissae axis represents the variation of the parameter $\theta \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$, which indexes the reference density $\frac{1}{2}\mathrm{vM}\left((0,1),2\right) + \frac{1}{2}\mathrm{vM}\left((\cos(\theta), \sin(\theta)), 2\right)$. Right plot: logarithm of h_{TAY}, h_{ROT}, h_{MISE} and their corresponding MISE for different values of κ, with $n = 250$.

4.4 Selectors based on mixtures

The results of the previous section show that, although the rule of thumb presents a significant improvement with respect to the Taylor (2008) selector in terms of generality and robustness, it also shares the same drawbacks when the underlying density is not the von Mises model (see Figure 4.1). To overcome these problems, two alternatives for improving h_{ROT} will be considered.

The first one is related with improving the reference density to plug-in into the curvature term. The von Mises density has been proved to be not flexible enough to estimate properly the curvature term in (4.4). This is specially visible when the underlying model is a mixture of antipodal von Mises, but the estimated curvature term is close to zero (the curvature of a uniform density). A modification in this direction is to consider a suitable mixture of von Mises for the reference density, that will be able to capture the curvature of rather complex underlying densities. This idea was employed first by Ćwik and Koronacki (1997) considering mixtures of multivariate normals and by Oliveira et al. (2012) in the circular setting.

The second improvement is concerned with the error criterion for the choice of the bandwidth. Until now, the error criterion considered was the AMISE, which is the usual in the literature of kernel smoothing. However, as Marron and Wand (1992) showed for the linear case and García-Portugués et al. (2013b) did for the directional situation, the AMISE and MISE may differ significantly for moderate and even large sample sizes, with a potential significative misfit between h_{AMISE} and h_{MISE}. Then, a substantial decreasing of the error of the estimator (4.2) is likely to happen if the bandwidth is obtained from the exact MISE, instead of the asymptotic version. Obviously, the problem of this new approach is how to compute exactly the MISE, but this can be done if the reference density is a mixture of von Mises.

The previous two considerations, improve the reference density and the error criterion, will lead to the bandwidth selectors of Asymptotic MIxtures (AMI), denoted by h_{AMI}, and Exact MIxtures (EMI), denoted by h_{EMI}. Before explaining in detail the two proposed selectors, it is required to introduce some notation on mixtures of von Mises.

An M-mixture of von Mises densities with means $\boldsymbol{\mu}_j$, concentration parameters κ_j and weights p_j, with $j = 1, \ldots, M$, is denoted by

$$f_M(\mathbf{x}) = \sum_{j=1}^{M} p_j f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}_j, \kappa_j), \quad \sum_{j=1}^{M} p_j = 1, \quad p_j \geq 0. \tag{4.7}$$

When dealing with mixtures, the tuning parameter is the number of components, M, which can be estimated from the sample. The notation $f_{\widehat{M}}$ will be employed to represent the mixture of $\widehat{M}$ components where the parameters are estimated and $\widehat{M}$ is obtained from the sample. The details of this fitting are explained later in Algorithm 4.3.

Then, the AMI selector follows from modifying the rule of thumb selector to allow fitted mixtures of von Mises. It is stated in the next procedure.

Algorithm 4.1 (AMI selector)**.** *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample of a directional variable $\mathbf{X}$.*

i. Compute a suitable estimation $f_{\widehat{M}}$ using Algorithm 4.3.

ii. For a directional kernel L, set

$$h_{\mathrm{AMI}} = \left[\frac{q d_q(L)}{4 b_q(L)^2 \lambda_q(L) R\Big(\Psi(f_{\widehat{M}}, \cdot)\Big) n}\right]^{\frac{1}{4+q}}$$

and for the von Mises kernel,

$$h_{\mathrm{AMI}} = \left[q 2^q \pi^{\frac{q}{2}} R\Big(\Psi(f_{\widehat{M}}, \cdot)\Big) n\right]^{-\frac{1}{4+q}}.$$

Remark 4.1. *Unfortunately, the curvature term $R(\Psi(f_{\widehat{M}}, \cdot))$ does not admit a simple closed expression, unless for the case where $\widehat{M} = 1$, i.e., when h_{AMI} is equivalent to h_{ROT}. This is due to the cross-product terms between the derivatives of the mixtures that appear in the integrand. However, this issue can be bypassed by using either numerical integration in q-spherical coordinates or Monte Carlo integration to compute $R(\Psi(f_{\widehat{M}}, \cdot))$ for any $\widehat{M}$.*

The EMI selector relies on the exact expression of the MISE for densities of the type (4.7), that will be denoted by

$$\mathrm{MISE}_M(h) = \mathbb{E}\left[\int_{\Omega_q}\left(\hat{f}_h(\mathbf{x}) - f_M(\mathbf{x})\right)^2 \omega_q(d\mathbf{x})\right].$$

Similarly to what Marron and Wand (1992) did for the linear case, García-Portugués et al. (2013b) derived the closed expression of $\mathrm{MISE}_M(h)$ when the directional kernel is the von Mises one. The calculations are based on the convolution properties of the von Mises, which unfortunately are not so straightforward as the ones for the normal, resulting in more complex expressions.

Proposition 4.3 (García-Portugués et al. (2013b))**.** *Let f_M be the density of an M-mixture of directional von Mises (4.7). The exact MISE of the directional kernel estimator (4.2) with von Mises kernel and obtained from a random sample of size n is*

$$\mathrm{MISE}_M(h) = (D_q(h)n)^{-1} + \mathbf{p}^T\left[(1-n^{-1})\boldsymbol{\Psi}_{\mathbf{2}}(h) - 2\boldsymbol{\Psi}_{\mathbf{1}}(h) + \boldsymbol{\Psi}_{\mathbf{0}}(h)\right]\mathbf{p}, \tag{4.8}$$

where $\mathbf{p} = (p_1,\ldots,p_M)^T$ and $D_q(h) = C_q\left(1/h^2\right)^2 C_q\left(2/h^2\right)^{-1}$. The matrices $\boldsymbol{\Psi}_{\mathbf{a}}(h)$, $a = 0,1,2$ have entries:

$$\boldsymbol{\Psi}_{\mathbf{0}}(h) = \left(\frac{C_q(\kappa_i)C_q(\kappa_j)}{C_q(||\kappa_i\boldsymbol{\mu}_i + \kappa_j\boldsymbol{\mu}_j||)}\right)_{ij},\ \boldsymbol{\Psi}_{\mathbf{1}}(h) = \left(\int_{\Omega_q}\frac{C_q\left(1/h^2\right)C_q(\kappa_i)C_q(\kappa_j)}{C_q\left(||\mathbf{x}/h^2 + \kappa_i\boldsymbol{\mu}_i||\right)}e^{\kappa_j\mathbf{x}^T\boldsymbol{\mu}_j}\,\omega_q(d\mathbf{x})\right)_{ij},$$

$$\boldsymbol{\Psi}_{\mathbf{2}}(h) = \left(\int_{\Omega_q}\frac{C_q\left(1/h^2\right)^2 C_q(\kappa_i)C_q(\kappa_j)}{C_q\left(||\mathbf{x}/h^2 + \kappa_i\boldsymbol{\mu}_i||\right)C_q(||\mathbf{x}/h^2 + \kappa_j\boldsymbol{\mu}_j||)}\,\omega_q(d\mathbf{x})\right)_{ij},$$

where C_q is defined in equation (4.1).

Remark 4.2. *A more efficient way to implement (4.8), specially for large sample sizes and higher dimensions, is the following expression:*

$$\mathrm{MISE}_M(h) = (D_q(h)n)^{-1} + \int_{\Omega_q}\left\{\left(\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right] - f_M(\mathbf{x})\right)^2 - \mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]^2\right\}\omega_q(d\mathbf{x}),$$

where the integral is either evaluated numerically using q-spherical coordinates or Monte Carlo integration and $\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]$ is computed using

$$\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right] = \sum_{j=1}^{M} p_j\frac{C_q(\kappa_j)C_q\left(1/h^2\right)}{C_q(||\mathbf{x}/h^2 + \kappa_j\boldsymbol{\mu}_j||)}.$$

Remark 4.3. *By the use of similar techniques, when the kernel is von Mises, the LSCV selector admits an easier expression for the CV_2 loss that avoids the calculation of the integral of $\hat{f}_h^{-i}$:*

$$\mathrm{CV}_2(h) = \frac{2C_q\left(1/h^2\right)}{n}\sum_{i=1}^{n}\sum_{j>i}^{n}\left[\frac{2}{n-1}e^{\mathbf{X}_i^T\mathbf{X}_j/h^2} - \frac{C_q\left(1/h^2\right)}{nC_q\left(||\mathbf{X}_i + \mathbf{X}_j||/h^2\right)}\right] - \frac{C_q(1/h^2)^2}{nC_q(2/h^2)}.$$

Based on the previous result, the philosophy of the EMI selector is the following: using a suitable pilot parametric estimation of the unknown density (given by Algorithm 4.3), build the exact MISE and obtain the bandwidth that minimizes it. This is summarized in the following procedure.

Algorithm 4.2 (EMI selector)**.** *Consider the von Mises kernel and let* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be a random sample of a directional variable* $\mathbf{X}$*.*

i. Compute a suitable estimation $f_{\widehat{M}}$ *using Algorithm 4.3.*

ii. Obtain $h_{\text{EMI}} = \arg\min_{h>0} \text{MISE}_{\widehat{M}}(h)$*.*

4.4.1 Mixtures fitting and selection of the number of components

The EM algorithm of Banerjee et al. (2005), implemented in the `R` package `movMF` (see Hornik and Grün (2014)), provides a complete solution to the problem of estimation of the parameters in a mixture of directional von Mises of dimension q. However, the issue of selecting the number of components of the mixture in an automatic and optimal way is still an open problem.

The propose considered in this work is an heuristic approach based on the Bayesian Information Criterion (BIC), defined as $\text{BIC} = -2l + k \log n$, where l is the log-likelihood of the model and k is the number of parameters. The procedure looks for the fitted mixture with a number of components M that minimizes the BIC. This problem can be summarized as the global minimization of a function (BIC) defined on the naturals (number of components).

The heuristic procedure starts by fitting mixtures from $M = 1$ to $M = M_B$, computing their BIC and providing $\widehat{M}$, the number of components with minimum BIC. Then, in order to ensure that $\widehat{M}$ is a global minimum and not a local one, M_N neighbours next to $\widehat{M}$ are explored (*i.e.* fit mixture, compute BIC and update $\widehat{M}$), if they were not previously explored. This procedure continues until $\widehat{M}$ has at least M_N neighbours at each side with larger BICs. A reasonable compromise for M_B and M_N, checked by simulations, is to set $M_B = \lfloor \log n \rfloor$ and $M_N = 3$. In order to avoid spurious solutions, fitted mixtures with any $\kappa_j > 250$ are removed. The procedure is detailed as follows.

Algorithm 4.3 (Mixture estimation with data-driven selection of the number of components)**.** *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be a random sample of a directional variable* $\mathbf{X}$ *with density* f*.*

i. Set $M_B = \lfloor \log n \rfloor$ *and* M_N *as the user supplies, usually* $M_N = 3$*.*

ii. For M *varying from* 1 *to* M_B*,*

  i. estimate the M*-mixture with the EM algorithm of Banerjee et al. (2005) and*

  ii. compute the BIC of the fitted mixture.

iii. Set $\widehat{M}$ *as the number of components of the mixture with lower BIC.*

iv. If $M_B - M_N < \widehat{M}$*, set* $M_B = M_B + 1$ *and turn back to step ii. Otherwise, end with the final estimation* $f_{\widehat{M}}$*.*

Other informative criteria, such as the Akaike Information Criterion (AIC) and its corrected version, AICc, were checked in the simulation study together with BIC. The BIC turned out to be the best choice to use with the AMI and EMI selectors, as it yielded the minimum errors.

4.5 Comparative study

Along this section, the three new bandwidth selectors will be compared with the already proposed selectors described in Subsection 4.2.1. A collection of directional models, with their corresponding simulation schemes, are considered. Subsection 4.5.1 is devoted to comment the directional models used in the simulation study (all of them are defined for any arbitrary dimension q, not just for the circular or spherical case). These models are also described in the appendix.

For each of the different combinations of dimension, sample size and model, the MISE of each selector was estimated empirically by 1000 Monte Carlo samples, with the same seed for the different selectors. This is used in the computation of $\mathrm{MISE}(h_{\mathrm{MISE}})$, where h_{MISE} is obtained as a numerical minimization of the estimated MISE. The calculus of the ISE was done by: Simpson quadrature rule with 2000 discretization points for $q = 1$; Lebedev and Laikov (1995) rule with 5810 nodes for $q = 2$ and Monte Carlo integration with 10000 sampling points for $q > 2$ (same seed for all the integrations). Finally, the kernel considered in the study is the von Mises.

4.5.1 Directional models

The first models considered are the uniform density in Ω_q and the von Mises density given in (4.1). The analogous of the von Mises for axial data (*i.e.*, directional data where $f(\mathbf{x}) = f(-\mathbf{x})$) is the Watson distribution $\mathrm{W}(\boldsymbol{\mu}, \kappa)$ (Mardia and Jupp, 2000):

$$f_{\mathrm{W}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = M_q(\kappa) \exp\left\{\kappa(\mathbf{x}^T\boldsymbol{\mu})^2\right\},$$

where $M_q(\kappa) = \left(\omega_{q-1}\int_{-1}^{1} e^{\kappa t^2}(1-t^2)^{\frac{q}{2}-1}\,dt\right)^{-1}$. This density has two antipodal modes: $\boldsymbol{\mu}$ and $-\boldsymbol{\mu}$, both of them with concentration parameter $\kappa \geq 0$. A further extension of this density is the called Small Circle distribution $\mathrm{SC}(\boldsymbol{\mu}, \tau, \nu)$ (Bingham and Mardia, 1978):

$$f_{\mathrm{SC}}(\mathbf{x}; \boldsymbol{\mu}, \tau, \nu) = A_q(\tau, \nu) \exp\left\{-\tau(\mathbf{x}^T\boldsymbol{\mu} - \nu)^2\right\},$$

where $A_q(\tau, \nu) = \left(\omega_{q-1}\int_{-1}^{1} e^{-\tau(t-\nu)^2}(1-t^2)^{\frac{q}{2}-1}\,dt\right)^{-1}$, $\nu \in (-1, 1)$ and $\tau \in \mathbb{R}$. For the case $\tau \geq 0$, this density has a kind of modal strip along the $(q-1)$-sphere $\{\mathbf{x} \in \Omega_q : \mathbf{x}^T\boldsymbol{\mu} = \nu\}$.

A common feature of all these densities is that they are rotationally symmetric, that is, their contourlines are $(q-1)$-spheres orthogonal to a particular direction. This characteristic can be exploited by means of the so called tangent-normal decomposition (see Mardia and Jupp (2000)), that leads to the change of variables

$$\begin{cases} \mathbf{x} = t\boldsymbol{\mu} + (1-t^2)^{\frac{1}{2}}\mathbf{B}_{\boldsymbol{\mu}}\boldsymbol{\xi}, \\ \omega_q(d\mathbf{x}) = (1-t^2)^{\frac{q}{2}-1}\,dt\,\omega_{q-1}(d\boldsymbol{\xi}), \end{cases} \tag{4.9}$$

where $\boldsymbol{\mu} \in \Omega_q$ is a fixed vector, $t = \boldsymbol{\mu}^T\mathbf{x}$ (measures the distance of $\mathbf{x}$ from $\boldsymbol{\mu}$), $\boldsymbol{\xi} \in \Omega_{q-1}$ and $\mathbf{B}_{\boldsymbol{\mu}} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$ is the semi-orthonormal matrix ($\mathbf{B}_{\boldsymbol{\mu}}^T\mathbf{B}_{\boldsymbol{\mu}} = \mathbf{I}_q$ and $\mathbf{B}_{\boldsymbol{\mu}}\mathbf{B}_{\boldsymbol{\mu}}^T = \mathbf{I}_{q+1} - \boldsymbol{\mu}\boldsymbol{\mu}^T$, with $\mathbf{I}_q$ the q-identity matrix) resulting from the completion of $\boldsymbol{\mu}$ to the orthonormal basis $\{\boldsymbol{\mu}, \mathbf{b}_1, \ldots, \mathbf{b}_q\}$. The family of rotationally symmetric densities can be parametrized as

$$f_{g_{\boldsymbol{\theta}},\boldsymbol{\mu}}(\mathbf{x}) = g_{\boldsymbol{\theta}}(\boldsymbol{\mu}^T\mathbf{x}), \tag{4.10}$$

where $g_{\boldsymbol{\theta}}$ is a function depending on a vector parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ and such that $\omega_{q-1} g_{\boldsymbol{\theta}}(t)(1-t^2)^{\frac{q}{2}-1}$ is a density in $(-1,1)$, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Using this property, it is easy to simulate from (4.10).

Algorithm 4.4 (Sampling from a rotationally symmetric density)**.** *Let be the rotationally symmetric density (4.10) and consider the notation of (4.9).*

i. Sample T from the density $\omega_{q-1} g_{\boldsymbol{\theta}}(t)(1-t^2)^{\frac{q}{2}-1}$.

ii. Sample $\boldsymbol{\xi}$ from a uniform in Ω_{q-1} ($\Omega_0 = \{-1, 1\}$).

iii. $T\boldsymbol{\mu} + (1-T^2)^{\frac{1}{2}} \mathbf{B}_{\boldsymbol{\mu}} \boldsymbol{\xi}$ is a sample from $f_{g_{\boldsymbol{\theta}}, \boldsymbol{\mu}}$.

Remark 4.4. *Step i can always be performed using the inversion method (Johnson, 1987). This approach can be computationally expensive: it involves solving the root of the distribution function, which is computed from an integral evaluated numerically if no closed expression is available. A reasonable solution to this (for a fixed choice of $g_{\boldsymbol{\theta}}$ and $\boldsymbol{\mu}$) is to evaluate once the quantile function in a dense grid (for example,* 2000 *points equispaced in* $(0,1)$*), save the grid and use it to interpolate using cubic splines the new evaluations, which is computationally fast.*

Extending these ideas for rotationally symmetric models, two new directional densities are proposed. The first one is the Directional Cauchy density $\mathrm{DC}(\boldsymbol{\mu}, \kappa)$, defined as an analogy with the usual Cauchy distribution as

$$f_{\mathrm{DC}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = \frac{1}{D_q(\kappa)(1 + 2\kappa(1 - \mathbf{x}^T\boldsymbol{\mu}))}, \quad D_q(\kappa) = \begin{cases} 2\pi\,(1+4\kappa)^{-1/2}, & q = 1, \\ \pi \log(1+4\kappa)\kappa^{-1}, & q = 2, \\ \omega_{q-1} \int_{-1}^{1} \frac{(1-t^2)^{\frac{q}{2}-1}}{1+2\kappa(1-t)}\, dt, & q > 2, \end{cases}$$

where $\boldsymbol{\mu}$ is the mode direction and $\kappa \geq 0$ the concentration parameter around it ($\kappa = 0$ gives the uniform density). This density shares also some of the characteristics of the usual Cauchy distribution: high concentration around a peaked mode and a power decay of the density. The other proposed density is the Skew Normal Directional density $\mathrm{SND}(\boldsymbol{\mu}, m, \sigma, \lambda)$,

$$f_{\mathrm{SND}}(\mathbf{x}; \boldsymbol{\mu}, m, \sigma, \lambda) = S_q(m, \sigma, \lambda) g_{m,\sigma,\lambda}(\boldsymbol{\mu}^T\mathbf{x}),\ S_q(m, \sigma, \lambda) = \left(\omega_{q-1} \int_{-1}^{1} g_{m,\sigma,\lambda}(t)(1-t^2)^{\frac{q}{2}-1}\, dt \right)^{-1},$$

where $g_{m,\sigma,\lambda}$ is the skew normal density of Azzalini (1985) with location m, scale σ and shape λ that is truncated to the interval $(-1,1)$. The density is inspired by the wrapped skew normal distribution of Pewsey (2006), although it is based on the rotationally symmetry rather than in wrapping techniques. A particular form of this density is an homogeneous "cap" in a neighbourhood of $\boldsymbol{\mu}$ that decreases very fast outside of it.

Non rotationally symmetric densities can be created by mixtures of rotationally symmetric. However, it is interesting to introduce a purely non rotationally symmetric density: the Projected Normal distribution of Pukkila and Rao (1988). Denoted by $\mathrm{PN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the corresponding density is

$$f_{\mathrm{PN}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \mathbf{Q}_3^{-\frac{p}{2}} I_p\left(\mathbf{Q}_2 \mathbf{Q}_3^{-\frac{1}{2}}\right) \exp\left\{-2^{-1}\left(\mathbf{Q}_1 - \mathbf{Q}_2^2 \mathbf{Q}_3^{-1}\right)\right\},$$

where $\mathbf{Q}_1 = \mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}$, $\mathbf{Q}_2 = \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}$, $\mathbf{Q}_3 = \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ and $I_p(\alpha) = \int_0^\infty t^{p-1} \exp\{-2^{-1}(t-\alpha)^2\}\, dt$.

Sampling from this distribution is extremely easy: just sample $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and then project $\mathbf{X}$ to Ω_q by $\mathbf{X}/||\mathbf{X}||$.

The whole collection of models, with 20 densities in total, are detailed in Table 4.5 in Appendix 4.B. Figures 4.2 and 4.3 show the plots of these densities for the circular and spherical cases.

4.5.2 Circular case

For the circular case, the comparative study has been done for the 20 models described in Figure 4.2 (see Table 4.5 to see their densities), for the circular selectors h_{LCV}, h_{LSCV}, h_{TAY}, h_{OLI}, h_{ROT}, h_{AMI} and h_{EMI} and for the sample sizes 100, 250, 500 and 1000. Due to space limitations, only the results for sample size 500 are shown in Table 4.1, and the rest of them are relegated to Appendix 4.C.

In addition, to help summarizing the results a ranking similar to Ranking B of Cao et al. (1994) will be constructed. The ranking will be computed according to the following criteria: for each model, the m bandwidth selectors $h_1, \ldots, h_m$ considered are sorted from the best performance (lowest error) to the worst performance (largest error). The best bandwidth receives m points, the second $m-1$ and so on. These points, denoted by r, are standardized by m and multiplied by the relative performance of each selector compared with the best one. In other words, the points of the selector h_k, if h_{opt} is the best one, are $\frac{r_k}{m}\frac{\mathrm{MISE}(h_{\mathrm{opt}})}{\mathrm{MISE}(h_k)}$. The final score for each selector is the sum of the ranks obtained in all the twenty models (thus, a selector which is the best in all models will have 20 points). With this ranking, it is easy to group the results in a single and easy to read table.

In view of the results, the following conclusions can be extracted. Firstly, h_{ROT} performs well in certain unimodal models such as M3 (von Mises) and M6 (skew normal directional), but its performance is very poor with multimodal models like M15 (Watson). In its particular comparison with h_{TAY}, it can be observed that both selectors share the same order of error, but being h_{ROT} better in all the situations except for one: the uniform model (M1). This is due to the "extra term" commented in Section 4.3: its absence in the denominator makes that $h_{\mathrm{TAY}} \to \infty$ faster than h_{ROT} when the concentration parameter $\kappa \to 0$ and, what is a disadvantage for $\kappa > 0$, turns out in an advantage for the uniform case. With respect to h_{AMI} and h_{EMI}, although their performance becomes more similar when the sample size increases, something expected, h_{EMI} seems to be on average a step ahead from h_{AMI}, specially for low sample sizes. Among the cross-validated selectors, h_{LCV} performs better than h_{LSCV}, a fact that was previously noted by simulation studies carried out by Taylor (2008) and Oliveira et al. (2012). Finally, h_{OLI} presents the most competitive behaviour among the previous proposals in the literature when the sample size is reasonably large (see Table 4.2).

The comparison between the circular selectors is summarized in the scores of Table 4.2. For all the sample sizes considered, h_{EMI} is the most competitive selector, followed by h_{AMI} for all the sample sizes except $n = 100$, where h_{LCV} is the second. The effect of the sample effect is also interesting to comment. For $n = 100$, h_{LCV} and h_{ROT} perform surprisingly well, in contrast with h_{OLI}, which is the second worst selector for this case. When the sample size increases, h_{ROT} and h_{TAY} have a decreasing performance and h_{OLI} stretches differences with h_{AMI}, showing a similar behaviour. This was something expected as both selectors are based on error criteria

that are asymptotically equivalent. The cross-validated selectors show a stable performance for sample sizes larger than $n = 100$.

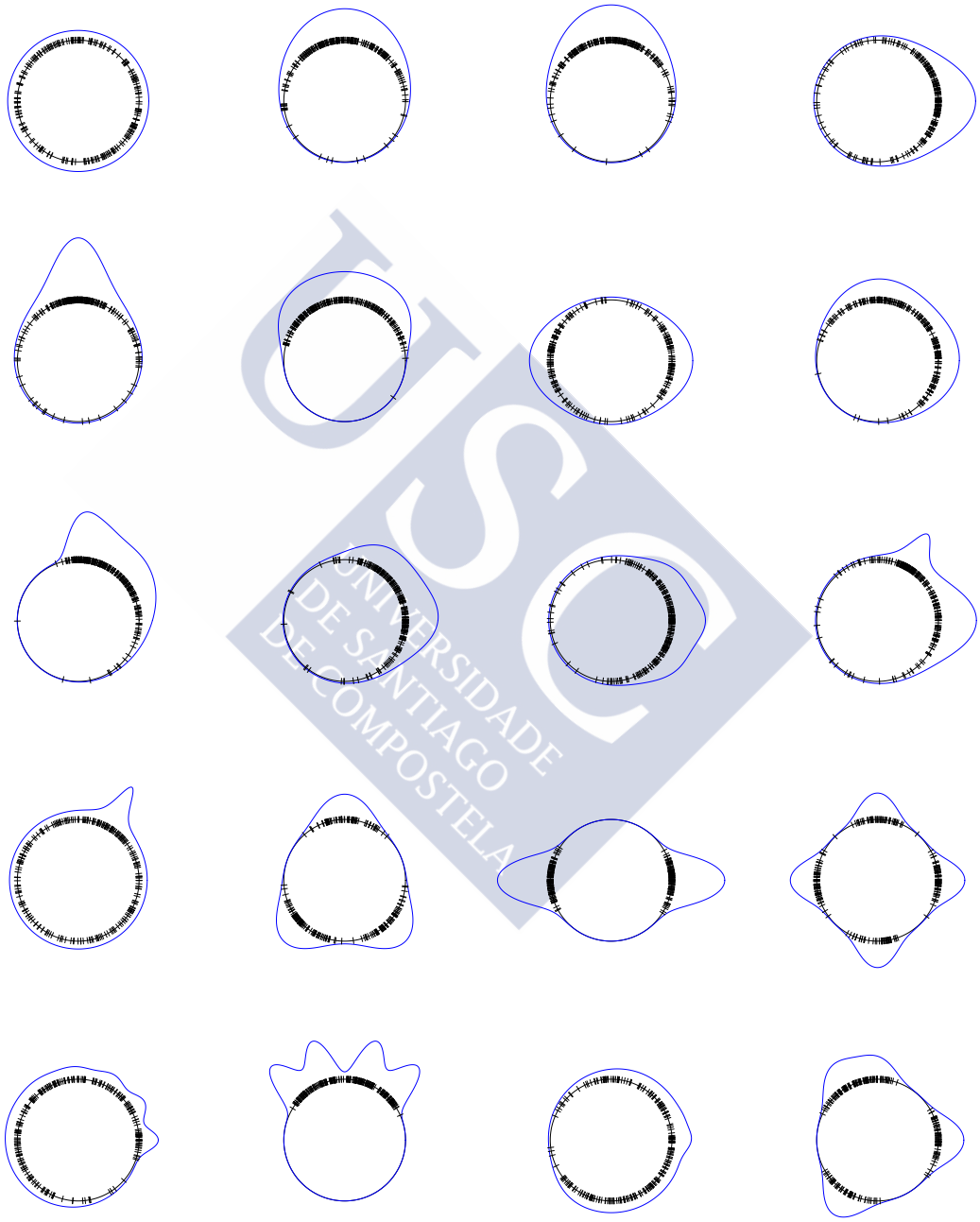


Figure 4.2: Simulation scenarios for the circular case. From left to right and up to down, models M1 to M20. For each model, a sample of size 250 is drawn.

Model	h_{MISE}	h_{LCV}	h_{LSCV}	h_{TAY}	h_{OLI}	h_{ROT}	h_{AMI}	h_{EMI}
M1	0.0000	0.057 (0.12)	0.059 (0.12)	**0.001** (0.00)	0.049 (0.10)	0.020 (0.03)	0.020 (0.03)	0.022 (0.03)
M2	0.2298	0.265 (0.16)	0.297 (0.21)	0.238 (0.15)	0.249 (0.16)	0.235 (0.14)	0.235 (0.14)	**0.234** (0.15)
M3	0.2815	0.338 (0.21)	0.356 (0.24)	0.313 (0.20)	0.301 (0.20)	**0.294** (0.19)	0.298 (0.19)	0.301 (0.19)
M4	0.3429	0.424 (0.23)	0.413 (0.23)	0.683 (0.29)	0.366 (0.19)	0.534 (0.25)	**0.363** (0.19)	0.363 (0.19)
M5	0.5938	0.883 (0.49)	0.708 (0.38)	2.032 (0.61)	0.640 (0.35)	1.723 (0.55)	**0.638** (0.35)	0.645 (0.36)
M6	0.2789	0.374 (0.23)	0.354 (0.23)	0.281 (0.14)	0.341 (0.20)	**0.281** (0.15)	0.323 (0.18)	0.304 (0.17)
M7	0.3013	0.333 (0.16)	0.353 (0.18)	6.677 (0.07)	0.319 (0.15)	5.533 (0.77)	**0.309** (0.15)	0.310 (0.15)
M8	0.2408	0.268 (0.14)	0.298 (0.17)	0.286 (0.12)	0.261 (0.14)	0.254 (0.12)	0.251 (0.12)	**0.248** (0.12)
M9	0.6208	0.913 (0.38)	0.720 (0.34)	1.270 (0.31)	0.685 (0.31)	1.152 (0.30)	0.676 (0.31)	**0.658** (0.30)
M10	0.3180	0.361 (0.17)	0.390 (0.22)	0.366 (0.16)	0.356 (0.18)	**0.341** (0.16)	0.347 (0.17)	0.351 (0.17)
M11	0.3056	0.346 (0.17)	0.372 (0.22)	0.352 (0.17)	0.327 (0.17)	**0.326** (0.15)	0.328 (0.15)	0.332 (0.15)
M12	0.7322	0.910 (0.33)	0.822 (0.30)	1.974 (0.30)	**0.788** (0.27)	1.750 (0.29)	0.854 (0.38)	0.831 (0.39)
M13	0.9349	1.178 (0.37)	1.011 (0.28)	4.625 (0.33)	**1.005** (0.31)	3.733 (0.28)	1.095 (0.55)	1.077 (0.55)
M14	0.5106	0.534 (0.20)	0.570 (0.23)	12.445 (0.53)	0.528 (0.20)	9.085 (1.11)	0.520 (0.19)	**0.517** (0.19)
M15	0.6101	0.663 (0.32)	0.709 (0.35)	44.295 (0.39)	0.648 (0.31)	39.961 (3.66)	0.642 (0.31)	**0.630** (0.30)
M16	0.6006	0.627 (0.23)	0.664 (0.25)	14.293 (0.00)	0.634 (0.22)	14.231 (0.08)	0.627 (0.22)	**0.613** (0.22)
M17	0.5891	**0.631** (0.17)	0.664 (0.18)	1.280 (0.20)	0.692 (0.15)	0.927 (0.10)	0.743 (0.18)	0.715 (0.17)
M18	1.0646	1.130 (0.40)	1.116 (0.39)	4.921 (0.24)	**1.067** (0.38)	4.630 (0.25)	1.104 (0.38)	1.087 (0.38)
M19	0.2718	0.306 (0.12)	0.323 (0.13)	0.524 (0.15)	0.302 (0.12)	0.341 (0.09)	0.301 (0.12)	**0.297** (0.10)
M20	0.5550	0.581 (0.22)	0.618 (0.25)	13.962 (0.91)	0.581 (0.23)	9.299 (1.20)	0.567 (0.22)	**0.562** (0.21)

Table 4.1: Comparative study for the circular case, with sample size $n = 500$. Columns of the selector $\bullet$ represent the MISE($\bullet$) $\times$ 100, with bold type for the minimum of the errors. The standard deviation of the ISE $\times$ 100 is given between parentheses.

q	n	h_{LCV}	h_{LSCV}	h_{TAY}	h_{OLI}	h_{ROT}	h_{AMI}	h_{EMI}
1	100	11.1494	9.4896	6.5143	6.8864	10.2829	11.1327	**14.5329**
	250	9.6357	7.6350	5.2053	10.7883	7.9129	13.0558	**16.0261**
	500	8.6549	7.7280	4.0933	13.2003	6.8351	14.5268	**15.6039**
	1000	9.0128	7.9820	3.7168	14.0234	5.6784	14.7077	**15.4358**
2	100	11.8161	13.4387	$*$	$*$	6.4424	8.1711	**15.3028**
	250	10.2201	12.2789	$*$	$*$	4.3195	11.2453	**17.1272**
	500	8.9001	12.1317	$*$	$*$	3.3156	13.0011	**18.0860**
	1000	8.2036	12.1566	$*$	$*$	2.9175	13.3693	**18.7548**

Table 4.2: Ranking for the selectors for the circular and spherical cases, for sample sizes $n = 100, 250, 500, 1000$. The higher the score in the ranking, the better the performance of the selector. Bold type indicates the best selector.

4.5.3 Spherical case

The comparative study for the spherical case has been done for the directional selectors h_{LSCV}, h_{ROT}, h_{AMI} and h_{EMI}, in the models given in Figure 4.3. As in the previous case, Table 4.2 contains the scores of the selectors for the different sample sizes, Table 4.3 includes the detailed results for $n = 500$ and the rest of the sample sizes are shown in Appendix 4.C.

In this case the results are even more clear. The h_{EMI} selector is by far the best, with an important gap between its competitors for all the sample sizes considered. Further, the effect of computing the exact error instead of the asymptotic one can be appreciated: h_{AMI} only is competitive against the cross-validated selectors for sample sizes larger than $n = 250$, while h_{EMI} remains always the most competitive. In addition, the performance of h_{AMI} seems to decrease

due to the effect of the dimension in the asymptotic error and does not converge so quick as in the circular case to the performance of h_{EMI}.

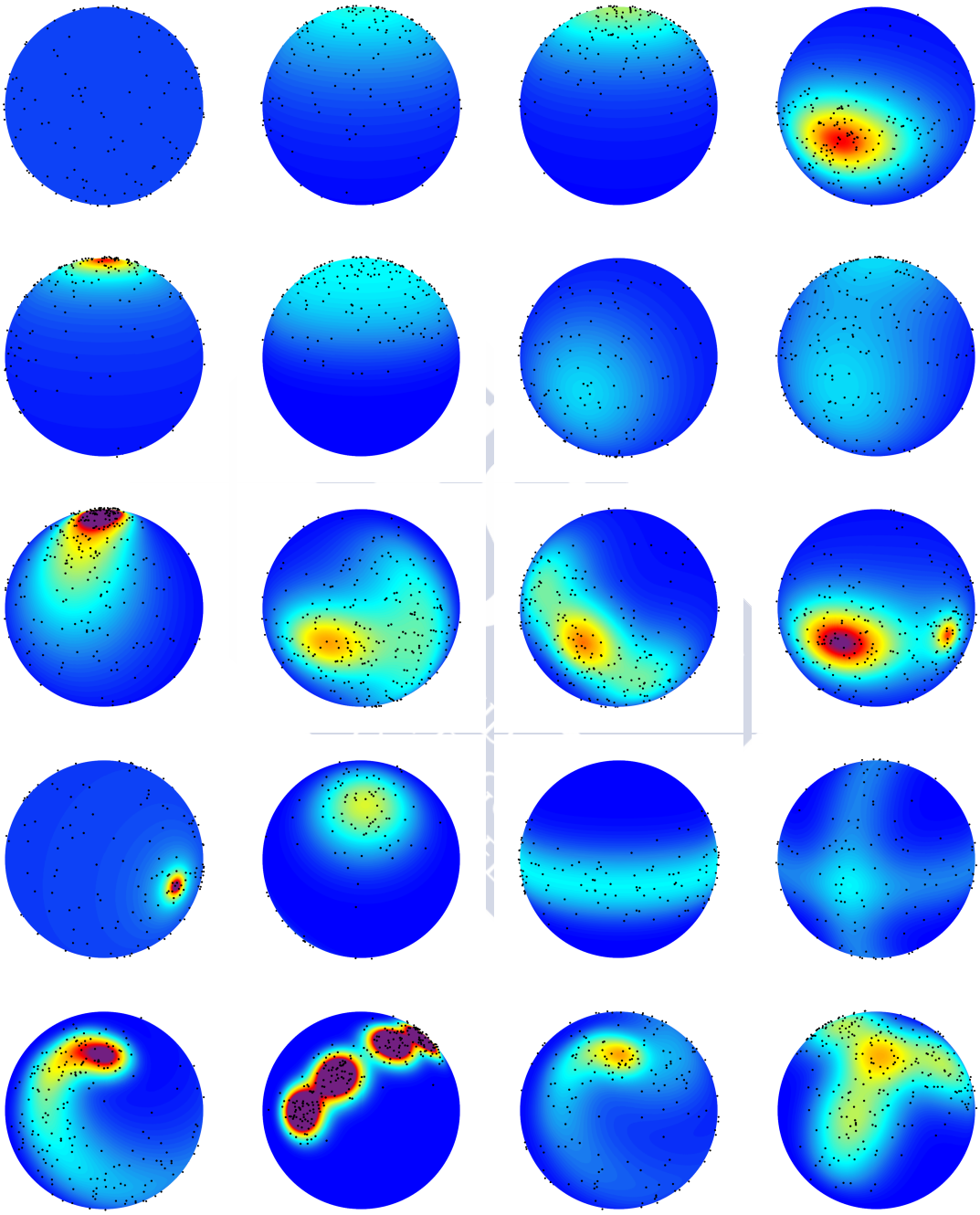


Figure 4.3: Simulation scenarios for the spherical case. From left to right and up to down, models M1 to M20. For each model, a sample of size 250 is drawn.

An interesting fact is that h_{LSCV} performs better than h_{LCV}, contrarily to what happens in the circular case. This phenomena is strengthen with higher dimensions, as it can be seen in the next subsection. A possible explanation is the following. For the standard linear case, LCV

has been proved to be a bad selector in densities with heavy tails (see Cao et al. (1994)) that are likely to produce outliers. In the circular case, the compact support jointly with periodicity may mitigate this situation, something that does not hold when the dimension increases and the sparsity of the observations is more likely. This makes that among the cross-validated selectors h_{LCV} works better for $q = 1$ and h_{LSCV} for $q > 1$.

Model	h_{MISE}	h_{LCV}	h_{LSCV}	h_{ROT}	h_{AMI}	h_{EMI}
M1	0.0000	0.023 (0.06)	0.023 (0.06)	0.041 (0.03)	0.017 (0.02)	**0.014** (0.02)
M2	0.3058	0.331 (0.13)	0.340 (0.15)	0.316 (0.13)	0.312 (0.12)	**0.310** (0.12)
M3	0.4729	0.555 (0.22)	0.525 (0.22)	0.542 (0.21)	**0.481** (0.19)	0.487 (0.19)
M4	1.0441	1.501 (0.46)	1.117 (0.34)	2.588 (0.48)	**1.088** (0.33)	1.093 (0.35)
M5	0.9621	1.417 (0.45)	1.024 (0.31)	2.130 (0.40)	1.016 (0.30)	**1.003** (0.32)
M6	0.4172	0.493 (0.15)	0.450 (0.14)	**0.427** (0.11)	0.496 (0.17)	0.443 (0.13)
M7	0.3927	0.417 (0.13)	0.417 (0.12)	1.656 (0.28)	**0.399** (0.12)	0.400 (0.12)
M8	0.3380	0.352 (0.11)	0.370 (0.12)	0.374 (0.11)	0.352 (0.10)	**0.343** (0.11)
M9	2.6708	5.343 (1.26)	2.834 (0.78)	9.710 (0.92)	2.871 (0.65)	**2.733** (0.72)
M10	0.9698	1.230 (0.31)	**1.036** (0.30)	1.521 (0.29)	1.110 (0.30)	1.101 (0.31)
M11	1.0349	1.312 (0.34)	1.097 (0.29)	2.213 (0.35)	1.158 (0.26)	**1.067** (0.27)
M12	1.5800	2.365 (0.59)	1.668 (0.43)	4.123 (0.54)	**1.642** (0.42)	1.643 (0.44)
M13	1.4085	1.674 (0.26)	1.472 (0.23)	2.211 (0.13)	1.729 (0.40)	**1.464** (0.25)
M14	1.1299	1.176 (0.30)	1.182 (0.30)	8.885 (0.77)	1.160 (0.28)	**1.137** (0.28)
M15	1.1262	**1.155** (0.21)	1.162 (0.21)	7.528 (0.76)	1.302 (0.25)	1.160 (0.21)
M16	0.8637	0.890 (0.14)	**0.887** (0.14)	3.480 (0.22)	0.957 (0.21)	0.889 (0.15)
M17	1.8989	2.514 (0.52)	1.971 (0.42)	6.693 (0.45)	2.060 (0.39)	**1.950** (0.42)
M18	5.0555	5.170 (1.08)	5.279 (1.14)	28.468 (0.79)	5.272 (1.06)	**5.097** (1.08)
M19	1.1259	1.262 (0.26)	**1.177** (0.24)	2.750 (0.24)	1.244 (0.30)	1.199 (0.31)
M20	1.1810	1.214 (0.28)	1.250 (0.30)	2.219 (0.28)	1.246 (0.29)	**1.195** (0.27)

Table 4.3: Comparative study for the spherical case, with sample size $n = 500$. Columns of the selector $\bullet$ represent the MISE($\bullet$) $\times$ 100, with bold type for the minimum of the errors. The standard deviation of the ISE $\times$ 100 is given between parentheses.

4.5.4 The effect of dimension

Finally, the previous selectors are tested in higher dimensions. Table 4.4 summarizes the information for dimensions $q = 3, 4, 5$ and sample size $n = 1000$ (see Table 4.8 in Appendix 4.C for whole results). As it can be seen, h_{EMI} continues performing better than its competitors. Also, as previously commented in the spherical case, h_{AMI} has a lower performance due to the misfit between AMISE and MISE, which gets worse when the sample size is fixed and the dimension increases. h_{LSCV} arises as the second best selector for higher dimensions, outperforming h_{LCV}, as happened in the spherical case.

q	h_{LCV}	h_{LSCV}	h_{ROT}	h_{AMI}	h_{EMI}
3	7.4838	15.3405	4.7658	10.6920	**17.6956**
4	8.8565	15.4862	5.1370	9.1871	**17.4579**
5	10.3262	15.1815	5.4088	7.6616	**15.5301**

Table 4.4: Ranking for the selectors for dimensions $q = 3, 4, 5$ and sample size $n = 1000$. The larger the score in the ranking, the better the performance of the selector. Bold type indicates the best selector.

4.6 Data application

According with the comparative study of the previous section, the h_{EMI} selector poses in average the best performance of all the considered selectors. In this section it will be applied to estimate the density of two real datasets.

4.6.1 Wind direction

Wind direction is a typical example of circular data. The data of this illustration was recorded in the meteorological station of A Mourela (7° 51' 21.91" W, 43° 25' 52.35" N), located near the coal power plant of As Pontes, in the northwest of Spain. The wind direction has a big impact on the dispersion of the pollutants from the coal power plant and a reliable estimation of its unknown density is useful for a further study of the pollutants transportation. The wind direction was measured minutely at the top of a pole of 80 metres during the month of June, 2012. In order to mitigate serial dependence, the data has been hourly averaged by computing the circular mean, resulting in a sample of size 673. The resulting bandwidth is $h_{\mathrm{EMI}} = 0.1896$, obtained from the data-driven mixture of 3 von Mises. Left plot of Figure 4.4 represents the estimated density, which shows a clear predominance of the winds from the west and three main modes. Running time, measured in a 3.5 GHz core, is 1.21 seconds (0.89 for the mixtures fitting and 0.32 for bandwidth optimization).

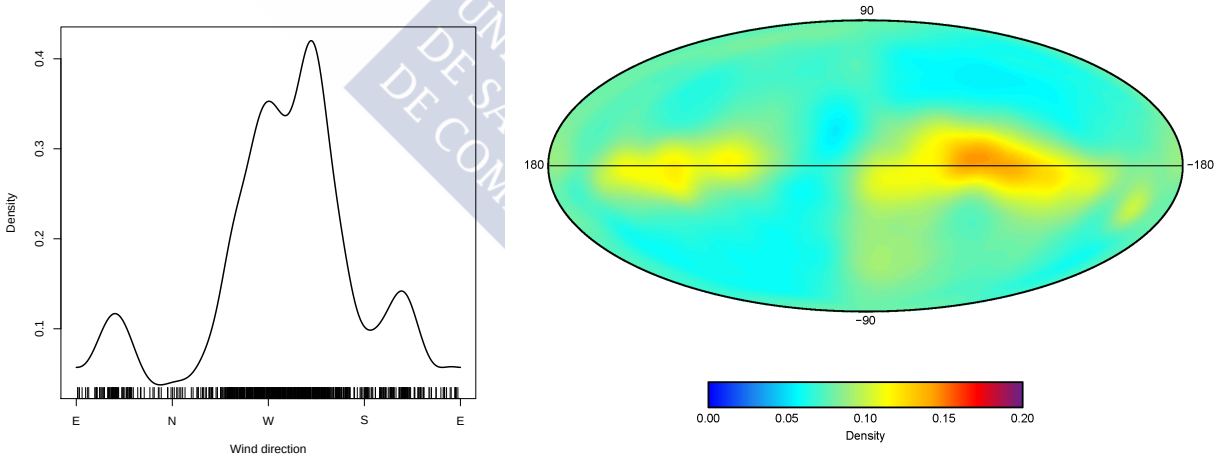


Figure 4.4: Left: density of the wind direction in the meteorological station of A Mourela. Right: density of the stars collected in the Hipparcos catalogue, represented in galactic coordinates and Aitoff projection.

4.6.2 Position of stars

A challenging field where spherical data is present is astronomy. Usually, the position of stars is referred to the position that occupy in the celestial sphere, *i.e.*, the location in the earth surface that arises as the intersection with the imaginary line that joins the centre of the earth with the star. A massive enumeration of near stars is given in the Hipparcos catalogue (Perryman, 1997), that collects the findings of the Hipparcos mission carried out by the European Space Agency in 1989–1993. An improved version of the original dataset, available from van Leeuwen

(2007), contains a corrected collection of the position of the stars on the celestial sphere as well as other star variables.

For many years, most of the statistical tools used to describe this kind of data were histograms adapted to the spherical case, where the choice of the bin width was done manually (see page 328 of Perryman (1997)). In this illustration, a smooth estimation of the spherical density is given using the optimal smoothing of the h_{EMI} selector. Using the 117955 star positions from the dataset of van Leeuwen (2007), the underlying density is approximated with 12 components automatically obtained, resulting the bandwidth $h_{\mathrm{EMI}} = 0.1064$. Note that the analysis of such a large dataset by cross-validatory techniques would demand an enormous amount of computing time and memory resources, whereas the running time for h_{EMI} is reasonable, with 256.01 seconds (247.34 for the mixtures fitting and 8.67 for bandwidth optimization). The right plot of Figure 4.4 shows the density of the position of the measured stars. This plot is given in the Aitoff projection (see Perryman (1997)) and in galactic coordinates, which means that the equator represents the position of the galactic rotation plane. The higher concentrations of stars are located around two spots, that represent the Orion's arm (left) and the Gould's Belt (right) of our galaxy.

4.7 Conclusions

Three new bandwidth selectors for directional data are proposed. The rule of thumb extends and improves significantly the previous proposal of Taylor (2008), but also fails estimating densities with multimodality. On the other hand, the selectors based on mixtures are competitive with the previous proposals in the literature, being the EMI selector the most competitive on average among all, for different sample sizes and dimensions, but specially for low or moderate sample sizes. The performance of AMI selector is one step behind EMI, a difference that is reduced when sample size increases.

In the comparison study, new rotationally symmetric models have been introduced and other interesting conclusions have been obtained. First, LCV is also a competitive selector for the circular case and outperforms LSCV, something that was known in the literature of circular data. However, this situation is reversed for the spherical case and higher dimensions, where LSCV is competitive and performs better than LCV.

The final conclusion of this paper is simple: the EMI bandwidth selector presents a reliable choice for kernel density estimation with directional data and its performance is at least as competitive as the existing proposals until the moment.

Acknowledgements

The author gratefully acknowledges the comments and guidance of professors Rosa M. Crujeiras and Wenceslao González-Manteiga. The work of the author has been supported by FPU grant AP2010–0957 from the Spanish Ministry of Education. Support of Project MTM2008–03010, from the Spanish Ministry of Science and Innovation, Project 10MDS207015PR from Dirección Xeral de I+D, Xunta de Galicia and IAP network StUDyS, from Belgian Science Policy, are acknowledged. The author acknowledges the helpful comments by the Editor and an Associate Editor.

4.A Proofs

Proof of Proposition 4.2. By simple differentiation, the operator (4.3) in a von Mises density $\mathrm{vM}(\boldsymbol{\mu}, \kappa)$ is

$$\Psi\left(f_{\mathrm{vM}}(\cdot;\boldsymbol{\mu},\kappa),\mathbf{x}\right) = \kappa C_q(\kappa) e^{\kappa \mathbf{x}^T\boldsymbol{\mu}} \left(-\mathbf{x}^T\boldsymbol{\mu} + \kappa q^{-1}\left(1-(\mathbf{x}^T\boldsymbol{\mu})^2\right)\right).$$

Then, by the change of variables of (4.9),

$$\begin{aligned} R\left(\Psi(f_{\mathrm{vM}}(\cdot;\boldsymbol{\mu},\kappa),\cdot)\right) &= \int_{\Omega_q} \Psi(f_{\mathrm{vM}}(\cdot;\boldsymbol{\mu},\kappa),\mathbf{x})^2\,\omega_q(d\mathbf{x}) \\ &= \kappa^2 C_q(\kappa)^2 \int_{\Omega_{q-1}} \int_{-1}^{1} e^{2\kappa t}\left(-t+\frac{\kappa}{q}(1-t^2)\right)^2 (1-t^2)^{\frac{q}{2}-1}\,dt\,\omega_{q-1}(d\boldsymbol{\xi}) \\ &= \frac{\kappa^{q+1}}{2^q \pi^{\frac{q}{2}+1}\mathcal{I}_{\frac{q-1}{2}}(\kappa)^2\Gamma\left(\frac{q}{2}\right)} \int_{-1}^{1} e^{2\kappa t}\left(-t+\frac{\kappa}{q}(1-t^2)\right)^2 (1-t^2)^{\frac{q}{2}-1}\,dt. \end{aligned}$$

The integral can divided into three terms expanding the square. After two integrations by parts, the sum of the first two is

$$\int_{-1}^{1} e^{2\kappa t}(1-t^2)^{\frac{q}{2}-1}t^2\,dt - \frac{2\kappa}{q}\int_{-1}^{1} e^{2\kappa t}(1-t^2)^{\frac{q}{2}}t\,dt = \frac{1}{q}\int_{-1}^{1} e^{2\kappa t}(1-t^2)^{\frac{q}{2}}\,dt.$$

This integral and the last term follows immediately by the integral form of the modified Bessel function, yielding

$$R\left(\Psi(f_{\mathrm{vM}}(\cdot;\boldsymbol{\mu},\kappa),\cdot)\right) = \frac{\kappa^{\frac{q+1}{2}}}{2^{q+2}\pi^{\frac{q+1}{2}}\mathcal{I}_{\frac{q-1}{2}}(\kappa)^2 q}\left[2q\mathcal{I}_{\frac{q+1}{2}}(2\kappa) + (2+q)\mathcal{I}_{\frac{q+3}{2}}(2\kappa)\right].$$

The particular case $q = 2$ follows by using $\mathcal{I}_{-\frac{1}{2}}(z) = \sqrt{\frac{2}{\pi z}}\sinh(z)$, $\mathcal{I}_{\frac{1}{2}}(z) = \sqrt{\frac{2}{\pi z}}\cosh(z)$ and relations $\mathcal{I}_{\nu-1}(z) = \mathcal{I}_{\nu+1}(z) + \frac{2\nu}{z}\mathcal{I}_{\nu}(z)$ and $\mathcal{I}_{\nu+1}(z) = \mathcal{I}_{\nu-1}(z) - \frac{2\nu}{z}\mathcal{I}_{\nu}(z)$. Also, for the von Mises kernel $L(r) = e^{-r}$, it is easy to see that

$$\lambda_q(L) = (2\pi)^{\frac{q}{2}}, \quad b_q(L) = \frac{q}{2} \quad \text{and} \quad d_q(L) = 2^{-\frac{q}{2}}.$$

□

4.B Models for the simulation study

Table 4.5 collects the densities of the different models used in the simulation study. Apart from the notations introduced in Section 4.5 for the families of directional densities, the following terminology is used. First, the vector $\mathbf{0}_q$ represent a vector with q zeros. Second, functions ρ_1 and ρ_2 give the polar and spherical parametrization of a vector from a single and a pair of angles, respectively:

$$\rho_1(\theta) = (\cos(\theta), \sin(\theta)),\ \rho_2(\theta,\phi) = (\cos(\theta)\sin(\phi), \sin(\theta)\sin(\phi), \cos(\phi)), \quad \theta \in [0, 2\pi),\ \phi \in [0, \pi).$$

Thirdly, the notation $\#i$ for an index i varying in the ordered set S aims to represent the position of i in S. Finally, the matrix $\boldsymbol{\Sigma}_1$ is such that the first three elements of $\mathrm{diag}(\boldsymbol{\Sigma}_1)$ are $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ and the rest of them are 1. Matrix $\boldsymbol{\Sigma}_2$ are just like matrix $\boldsymbol{\Sigma}_1$ but with the diagonal reversed.

4.C Extended tables for the simulation study

Tables 4.6 and 4.7 show the results for sample sizes 100, 250 and 1000 for the circular and spherical cases, respectively. The case $n = 500$ is collected in Tables 4.1 and 4.3. Finally, Table 4.8 contains the simulation results for sample size $n = 1000$ and dimensions $q = 3, 4, 5$.

Model	Description	Density
M1	Uniform (Unif.)	ω_q^{-1}
M2	Von Mises (vM)	$\mathrm{vM}((\mathbf{0}_q, 1), 2)$
M3	Projected normal (PN), rotationally symmetric	$\mathrm{PN}\left((\mathbf{0}_q, 1), \frac{1}{2} I_{q+1}\right)$
M4	Projected normal, non rotationally symmetric	$\mathrm{PN}((1, \mathbf{0}_q), 2\boldsymbol{\Sigma}_1)$
M5	Directional Cauchy (DC)	$\mathrm{DC}((\mathbf{0}_q, 1), 10)$
M6	Skew normal directional	$\mathrm{SND}\left((\mathbf{0}_q, 1), \frac{1}{2}, \frac{1}{2}, 5\right)$
M7	Watson	$\mathrm{W}((1, \mathbf{0}_q), 2)$
M8	Mixture of two 90° vM	$\frac{1}{2}\mathrm{vM}((\mathbf{0}_q, 1), 3) + \frac{1}{2}\mathrm{vM}((1, \mathbf{0}_q), 3)$
M9	Skewed mixture of vM (8 vM)	
	$q = 1$	$\frac{1}{8}\mathrm{vM}\left((0, 1), \left(\frac{5}{3}\right)^8\right) + \frac{1}{8}\sum_{i\in\{1,2,3,4,6,8,9\}} \mathrm{vM}\left(\rho_1\left(\frac{i\pi}{20}\right), \left(\frac{5}{3}\right)^{\#i}\right)$
	$q > 1$	$\frac{1}{8}\mathrm{vM}\left((\mathbf{0}_{q-1}, 1), \left(\frac{5}{3}\right)^8\right) + \frac{1}{8}\sum_{i\in\{1,2,3,4,6,8,9\}} \mathrm{vM}\left(\left(\mathbf{0}_{q-1}, \rho_2\left(0, \frac{(10-i)\pi}{20}\right)\right), \left(\frac{5}{3}\right)^{\#i}\right)$
M10	Mixture of two PN	$\frac{1}{2}\mathrm{PN}\left((1, \mathbf{0}_q), \boldsymbol{\Sigma}_1\right) + \frac{1}{2}\mathrm{PN}\left(\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, \mathbf{0}_{q-1}\right), \boldsymbol{\Sigma}_2\right)$
M11	Bandage (5 vM)	
	$q = 1$	$\frac{2}{10}\mathrm{vM}\left(\rho_1\left(0\right), 20\right) + \frac{2}{10}\sum_{i\in\{-1,1\}} \mathrm{vM}\left(\rho_1\left(\frac{i\pi}{6}\right), 10\right)$ $+\frac{1}{10}\sum_{i\in\{-1,1\}}\left[\mathrm{vM}\left(\rho_1\left(\frac{i\pi}{4}\right), 5\right) + \mathrm{vM}\left(\rho_1\left(\frac{i\pi}{2}\right), 1\right)\right]$
	$q > 1$	$\frac{2}{10}\mathrm{vM}\left(\left(\rho_2\left(0, \frac{\pi}{2}\right), \mathbf{0}_{q-1}\right), 20\right) + \frac{2}{10}\sum_{i\in\{-1,1\}} \mathrm{vM}\left(\left(\rho_2\left(\frac{i\pi}{6}, \frac{(4+i)\pi}{8}\right), \mathbf{0}_{q-1}\right), 10\right)$ $+\frac{2}{10}\sum_{i\in\{-1,1\}}\left[\mathrm{vM}\left(\left(\rho_2\left(\frac{i\pi}{4}, \frac{i\pi}{3}\right), \mathbf{0}_{q-1}\right), 5\right) + \mathrm{vM}\left(\left(\rho_2\left(\frac{i\pi}{2}, \frac{i\pi}{3}\right), \mathbf{0}_{q-1}\right), 1\right)\right]$
M12	Mixture of PN and DC	$\frac{3}{4}\mathrm{PN}((1, \mathbf{0}_q), \boldsymbol{\Sigma}_1) + \frac{1}{4}\mathrm{DC}\left(\left(\frac{1}{2}, \frac{\sqrt{3}}{2}, \mathbf{0}_{q-1}\right), 50\right)$
M13	Mixture of Unif. and DC	$\frac{4}{5}\omega_q^{-1} + \frac{1}{5}\mathrm{DC}\left(\left(\frac{1}{2}, \frac{\sqrt{3}}{2}, \mathbf{0}_{q-1}\right), 100\right)$
M14	Trimodal (3 vM)	$\frac{1}{3}\mathrm{vM}((\mathbf{0}_q, 1), 10) + \frac{1}{3}\mathrm{vM}\left(\left(\mathbf{0}_{q-1}, \rho_1\left(\frac{5\pi}{4}\right)\right), 10\right) + \frac{1}{3}\mathrm{vM}\left(\left(\mathbf{0}_{q-1}, \rho_1\left(\frac{7\pi}{4}\right)\right), 10\right)$
M15	Small circle	$\mathrm{SC}((\mathbf{0}_q, 1), 10)$
M16	Double small circle	$\frac{1}{2}\mathrm{SC}((\mathbf{0}_q, 1), 10) + \frac{1}{2}\mathrm{SC}((1, \mathbf{0}_q), 10)$
M17	Spiral (10 vM)	
	$q = 1$	$\frac{1}{10}\sum_{i=0}^{9} \mathrm{vM}\left(\rho_1\left(\frac{3\pi i}{18}\right), \left(\frac{3}{2}\right)^{10-i}\right)$
	$q > 1$	$\frac{1}{10}\sum_{i=0}^{9} \mathrm{vM}\left(\left(\rho_2\left(\frac{3\pi i}{18}, \frac{3\pi i}{36}\right), \mathbf{0}_{q-1}\right), \left(\frac{3}{2}\right)^{10-i}\right)$
M18	Claw (4 vM)	$\frac{1}{4}\sum_{i=0}^{1}\left[\mathrm{vM}\left(\left(\mathbf{0}_{q-1}, \rho_1\left(\frac{(2i+1)\pi}{4}\right), 50\right)\right) + \mathrm{vM}\left(\left(\mathbf{0}_{q-1}, \rho_1\left(\frac{(i+2)\pi}{5}\right)\right), 50\right)\right]$
M19	Double spiral (20 vM)	
	$q = 1$	$\frac{1}{20}\sum_{i=0}^{9}\left[\mathrm{vM}\left(\rho_1\left(\frac{3\pi i}{18}\right), \left(\frac{3}{2}\right)^{10-i}\right) + \mathrm{vM}\left(\rho_1\left(-\frac{3\pi i}{18}\right), 10\right)\right]$
	$q > 1$	$\frac{1}{20}\sum_{i=0}^{9}\left[\mathrm{vM}\left(\left(\rho_2\left(\frac{3\pi i}{18}, \frac{3\pi i}{36}\right), \mathbf{0}_{q-1}\right), \left(\frac{3}{2}\right)^{10-i}\right)\right.$ $\left.+\mathrm{vM}\left(\left(\rho_2\left(\frac{3\pi i}{18}, -\frac{3\pi i}{36}\right), \mathbf{0}_{q-1}\right), 10\right)\right]$
M20	Windmill (4 vM)	
	$q = 1$	$\frac{2}{11}\mathrm{vM}\left((0, 1), 20\right) + \frac{1}{11}\sum_{i=1}^{3} \mathrm{vM}\left(\rho_1\left(\frac{2i\pi}{3}\right), 15\right)$
	$q > 1$	$\frac{2}{11}\mathrm{vM}\left((\mathbf{0}_q, 1), 20\right) + \frac{1}{11}\sum_{i=1}^{3}\sum_{j\in\{3,5,6\}} \mathrm{vM}\left(\left(\rho_2\left(\frac{2i\pi}{3}, \frac{\pi}{j}\right), \mathbf{0}_{q-1}\right), 15\right)$

Table 4.5: Directional densities considered in the simulation study.

Model	h_{MISE}	h_{LCV}	h_{LSCV}	h_{TAY}	h_{OLI}	h_{ROT}	h_{AMI}	h_{EMI}
M1	0.0000	0.286 (0.56)	0.293 (0.60)	**0.019** (0.06)	0.680 (1.03)	0.109 (0.16)	0.115 (0.20)	0.120 (0.20)
M2	0.7525	0.938 (0.62)	1.140 (1.02)	0.803 (0.57)	1.229 (1.15)	**0.790** (0.55)	0.812 (0.67)	0.802 (0.62)
M3	0.8828	1.152 (0.72)	1.282 (1.09)	0.985 (0.66)	1.345 (1.24)	**0.936** (0.62)	0.953 (0.64)	0.962 (0.65)
M4	1.1173	**1.492** (0.90)	1.526 (1.05)	1.890 (1.02)	1.558 (1.18)	1.542 (0.86)	1.585 (1.00)	1.551 (0.94)
M5	1.9219	3.375 (1.71)	2.547 (1.61)	4.536 (1.79)	**2.357** (1.38)	3.982 (1.61)	2.415 (1.54)	2.384 (1.57)
M6	0.8810	1.264 (0.95)	1.228 (1.03)	**0.907** (0.53)	1.654 (1.37)	0.918 (0.55)	1.203 (0.90)	1.073 (0.74)
M7	0.9914	1.145 (0.64)	1.248 (0.79)	6.584 (0.20)	1.317 (0.96)	5.571 (0.74)	1.091 (0.67)	**1.064** (0.61)
M8	0.7534	0.885 (0.55)	1.046 (0.84)	0.823 (0.40)	1.255 (1.12)	**0.775** (0.43)	0.935 (0.59)	0.859 (0.50)
M9	1.8991	2.727 (1.19)	2.386 (1.35)	2.743 (0.89)	2.516 (1.40)	2.574 (0.88)	2.401 (1.22)	**2.245** (1.13)
M10	0.9551	1.154 (0.62)	1.355 (1.16)	1.029 (0.54)	1.510 (1.22)	**0.995** (0.54)	1.060 (0.80)	1.048 (0.68)
M11	0.8626	1.078 (0.68)	1.200 (0.95)	0.976 (0.63)	1.363 (1.24)	**0.913** (0.58)	0.932 (0.61)	0.933 (0.60)
M12	2.1940	2.995 (1.00)	**2.731** (1.33)	3.461 (0.79)	2.967 (1.42)	3.217 (0.77)	3.278 (1.02)	3.265 (0.93)
M13	2.6714	3.305 (1.01)	**3.142** (1.03)	5.012 (0.46)	3.405 (1.06)	4.486 (0.51)	3.879 (1.10)	3.611 (1.18)
M14	1.7224	1.872 (0.84)	2.028 (1.04)	12.790 (0.52)	2.076 (1.14)	11.392 (1.12)	1.853 (0.86)	**1.789** (0.82)
M15	2.3079	2.608 (1.50)	2.892 (1.74)	43.759 (1.41)	2.701 (1.64)	39.639 (3.94)	2.483 (1.36)	**2.408** (1.33)
M16	2.2045	2.354 (0.98)	2.585 (1.23)	14.312 (0.05)	2.620 (1.21)	14.303 (0.03)	2.490 (1.04)	**2.325** (0.97)
M17	1.2089	1.448 (0.56)	1.488 (0.65)	1.740 (0.41)	1.953 (1.11)	**1.372** (0.32)	1.523 (0.61)	1.413 (0.41)
M18	3.5569	**3.929** (1.56)	4.108 (1.58)	7.254 (0.72)	4.082 (1.61)	7.028 (0.74)	4.360 (1.47)	4.380 (1.44)
M19	0.6717	0.875 (0.53)	0.914 (0.56)	1.114 (0.50)	1.320 (1.09)	**0.766** (0.32)	0.887 (0.55)	0.813 (0.41)
M20	1.8640	2.034 (0.93)	2.203 (1.16)	14.834 (0.89)	2.360 (1.28)	12.778 (1.56)	2.046 (0.93)	**1.943** (0.86)
M1	0.0000	0.114 (0.22)	0.117 (0.23)	**0.004** (0.01)	0.137 (0.27)	0.040 (0.07)	0.040 (0.07)	0.043 (0.07)
M2	0.3760	0.456 (0.29)	0.513 (0.41)	0.393 (0.27)	0.444 (0.35)	0.387 (0.26)	0.387 (0.26)	**0.387** (0.26)
M3	0.4492	0.557 (0.36)	0.617 (0.48)	0.495 (0.32)	0.531 (0.40)	**0.468** (0.30)	0.478 (0.32)	0.483 (0.32)
M4	0.5716	0.738 (0.43)	0.718 (0.46)	1.066 (0.52)	**0.648** (0.38)	0.850 (0.44)	0.671 (0.42)	0.667 (0.42)
M5	1.0223	1.594 (0.88)	1.294 (0.73)	2.961 (1.01)	1.139 (0.63)	2.550 (0.92)	**1.125** (0.63)	1.133 (0.65)
M6	0.4702	0.658 (0.43)	0.626 (0.44)	**0.476** (0.25)	0.686 (0.48)	0.478 (0.26)	0.586 (0.36)	0.536 (0.31)
M7	0.5067	0.573 (0.29)	0.632 (0.37)	6.641 (0.12)	0.573 (0.32)	5.496 (0.79)	0.529 (0.28)	**0.527** (0.26)
M8	0.4070	0.459 (0.25)	0.533 (0.38)	0.462 (0.21)	0.492 (0.34)	**0.421** (0.22)	0.449 (0.27)	0.432 (0.25)
M9	1.0059	1.521 (0.61)	1.228 (0.60)	1.766 (0.47)	1.185 (0.56)	1.624 (0.46)	1.146 (0.54)	**1.096** (0.50)
M10	0.5141	0.600 (0.31)	0.659 (0.43)	0.568 (0.28)	0.644 (0.39)	**0.539** (0.28)	0.551 (0.31)	0.555 (0.29)
M11	0.4758	0.562 (0.31)	0.607 (0.36)	0.532 (0.29)	0.550 (0.34)	**0.498** (0.27)	0.506 (0.27)	0.511 (0.27)
M12	1.1958	1.549 (0.55)	1.388 (0.56)	2.501 (0.44)	**1.366** (0.53)	2.261 (0.43)	1.781 (0.72)	1.735 (0.78)
M13	1.5100	1.912 (0.61)	**1.692** (0.52)	4.813 (0.40)	1.740 (0.55)	4.055 (0.38)	1.981 (0.93)	1.896 (0.94)
M14	0.8561	0.908 (0.37)	0.981 (0.45)	12.619 (0.55)	0.925 (0.43)	10.234 (1.22)	0.886 (0.37)	**0.874** (0.36)
M15	1.0708	1.190 (0.63)	1.297 (0.69)	44.200 (0.60)	1.184 (0.64)	40.171 (3.55)	1.139 (0.59)	**1.113** (0.58)
M16	1.0426	1.088 (0.42)	1.156 (0.47)	14.296 (0.01)	1.132 (0.43)	14.246 (0.06)	1.107 (0.41)	**1.071** (0.40)
M17	0.9007	**0.973** (0.25)	1.017 (0.29)	1.474 (0.29)	1.046 (0.34)	1.078 (0.16)	1.067 (0.27)	1.005 (0.20)
M18	1.8266	1.974 (0.78)	1.989 (0.74)	5.904 (0.36)	**1.878** (0.72)	5.641 (0.36)	2.015 (0.80)	1.960 (0.80)
M19	0.4184	0.494 (0.24)	0.529 (0.30)	0.729 (0.27)	0.522 (0.28)	0.482 (0.16)	0.503 (0.23)	**0.476** (0.19)
M20	0.9527	1.009 (0.41)	1.075 (0.46)	14.453 (0.88)	1.042 (0.44)	11.014 (1.41)	0.988 (0.40)	**0.969** (0.39)
M1	0.0000	0.032 (0.06)	0.031 (0.06)	**0.000** (0.00)	0.023 (0.05)	0.010 (0.01)	0.010 (0.01)	0.011 (0.02)
M2	0.1386	0.157 (0.09)	0.176 (0.13)	0.142 (0.09)	0.144 (0.09)	0.140 (0.08)	0.140 (0.08)	**0.140** (0.08)
M3	0.1664	0.190 (0.12)	0.201 (0.13)	0.183 (0.11)	**0.172** (0.10)	0.173 (0.11)	0.176 (0.11)	0.177 (0.11)
M4	0.2051	0.247 (0.13)	0.240 (0.13)	0.432 (0.18)	0.215 (0.11)	0.331 (0.15)	**0.214** (0.11)	0.214 (0.11)
M5	0.3545	0.502 (0.27)	0.405 (0.20)	1.389 (0.37)	**0.373** (0.19)	1.160 (0.34)	0.374 (0.19)	0.378 (0.20)
M6	0.1718	0.219 (0.13)	0.212 (0.13)	0.173 (0.09)	0.194 (0.10)	**0.172** (0.09)	0.193 (0.10)	0.184 (0.10)
M7	0.1793	0.195 (0.09)	0.205 (0.10)	6.694 (0.04)	0.186 (0.08)	5.545 (0.78)	**0.183** (0.08)	0.184 (0.08)
M8	0.1472	0.160 (0.08)	0.175 (0.09)	0.180 (0.08)	0.153 (0.08)	0.157 (0.07)	0.151 (0.07)	**0.150** (0.07)
M9	0.3762	0.500 (0.21)	0.417 (0.18)	0.897 (0.20)	0.400 (0.17)	0.802 (0.20)	0.402 (0.17)	**0.392** (0.17)
M10	0.1916	0.213 (0.10)	0.232 (0.13)	0.227 (0.10)	**0.205** (0.10)	0.209 (0.10)	0.213 (0.10)	0.215 (0.10)
M11	0.1926	0.215 (0.10)	0.229 (0.12)	0.232 (0.10)	**0.202** (0.09)	0.212 (0.09)	0.208 (0.09)	0.210 (0.09)
M12	0.4395	0.534 (0.18)	0.481 (0.16)	1.507 (0.20)	**0.459** (0.15)	1.313 (0.19)	0.465 (0.16)	0.463 (0.16)
M13	0.5716	0.718 (0.23)	0.602 (0.17)	4.362 (0.27)	**0.591** (0.19)	3.406 (0.21)	0.609 (0.21)	0.616 (0.22)
M14	0.2951	0.306 (0.11)	0.320 (0.13)	12.094 (0.55)	0.299 (0.11)	7.599 (0.87)	0.298 (0.11)	**0.297** (0.11)
M15	0.3526	0.384 (0.19)	0.395 (0.19)	44.368 (0.25)	0.376 (0.18)	40.070 (3.76)	0.373 (0.18)	**0.367** (0.17)
M16	0.3408	0.355 (0.12)	0.372 (0.13)	14.293 (0.00)	0.358 (0.12)	14.217 (0.10)	0.356 (0.12)	**0.349** (0.12)
M17	0.3638	**0.385** (0.10)	0.393 (0.11)	1.111 (0.12)	0.486 (0.12)	0.815 (0.05)	0.468 (0.13)	0.466 (0.15)
M18	0.6353	0.670 (0.23)	0.651 (0.22)	3.963 (0.17)	0.688 (0.23)	3.665 (0.18)	0.649 (0.22)	**0.644** (0.22)
M19	0.1769	0.194 (0.07)	0.201 (0.08)	0.383 (0.09)	0.189 (0.06)	0.248 (0.06)	**0.188** (0.06)	0.193 (0.06)
M20	0.3286	0.339 (0.12)	0.359 (0.14)	13.143 (0.92)	0.335 (0.12)	7.403 (0.87)	0.333 (0.12)	**0.331** (0.12)

Table 4.6: Comparative study for the circular case, with up to down blocks corresponding to sample sizes 100, 250 and 1000, respectively. Columns of the selector $\bullet$ represent the MISE($\bullet$) $\times$ 100, with bold type for the minimum of the errors. The standard deviation of the ISE $\times$ 100 is given between parentheses.

Model	h_{MISE}	h_{LCV}	h_{LSCV}	h_{ROT}	h_{AMI}	h_{EMI}
M1	0.0000	**0.074** (0.23)	0.121 (0.32)	0.149 (0.12)	0.145 (0.72)	0.083 (0.20)
M2	0.8133	0.920 (0.45)	1.047 (0.73)	**0.838** (0.41)	0.889 (0.73)	0.851 (0.47)
M3	1.2481	1.465 (0.63)	1.527 (0.81)	1.361 (0.61)	1.334 (0.89)	**1.309** (0.62)
M4	2.6742	3.777 (1.31)	3.087 (1.35)	4.867 (1.24)	3.161 (1.40)	**3.022** (1.30)
M5	2.3676	3.260 (1.10)	**2.760** (1.09)	3.770 (0.96)	3.353 (1.57)	2.992 (1.22)
M6	1.0781	1.362 (0.61)	1.295 (0.67)	**1.092** (0.38)	1.304 (0.90)	1.158 (0.47)
M7	1.0165	1.179 (0.58)	1.176 (0.47)	2.199 (0.33)	1.186 (0.92)	**1.078** (0.43)
M8	0.8731	0.958 (0.40)	1.048 (0.53)	**0.921** (0.34)	1.056 (0.64)	0.947 (0.41)
M9	6.7507	11.184 (2.90)	7.647 (2.67)	14.962 (1.99)	8.056 (2.53)	**7.186** (2.14)
M10	2.3766	2.851 (0.79)	2.751 (1.13)	2.946 (0.73)	**2.653** (1.20)	2.660 (0.81)
M11	2.5706	3.266 (0.87)	**2.926** (1.13)	4.029 (0.83)	3.588 (1.88)	3.263 (1.08)
M12	3.8591	5.629 (1.56)	**4.312** (1.49)	6.991 (1.30)	4.996 (2.14)	4.559 (1.76)
M13	2.4304	2.689 (0.37)	2.681 (0.53)	**2.629** (0.23)	3.691 (2.32)	2.769 (0.49)
M14	3.0506	3.200 (0.90)	3.364 (1.17)	12.687 (1.31)	3.376 (0.95)	**3.128** (0.88)
M15	2.8638	**3.031** (0.70)	3.083 (0.81)	8.674 (0.64)	4.156 (1.91)	3.125 (0.80)
M16	2.1417	**2.263** (0.49)	2.300 (0.56)	3.928 (0.20)	3.033 (1.35)	2.911 (0.94)
M17	4.6150	5.730 (1.24)	**5.030** (1.35)	9.316 (0.93)	5.560 (2.16)	5.047 (1.32)
M18	13.2289	**13.572** (3.49)	14.586 (4.42)	40.641 (1.60)	15.735 (4.23)	13.717 (3.54)
M19	2.5921	2.907 (0.65)	**2.861** (0.76)	3.883 (0.48)	4.067 (1.67)	3.570 (0.96)
M20	3.0018	**3.174** (0.89)	3.373 (1.09)	4.261 (0.73)	3.850 (1.56)	3.316 (0.88)
M1	0.0000	0.046 (0.11)	0.047 (0.11)	0.070 (0.05)	0.038 (0.12)	**0.027** (0.04)
M2	0.4615	0.508 (0.23)	0.536 (0.28)	0.477 (0.22)	0.475 (0.20)	**0.471** (0.21)
M3	0.7147	0.844 (0.36)	0.817 (0.40)	0.808 (0.34)	**0.729** (0.31)	0.739 (0.32)
M4	1.6108	2.271 (0.71)	1.753 (0.59)	3.462 (0.73)	**1.704** (0.55)	1.710 (0.58)
M5	1.4470	2.070 (0.70)	1.587 (0.55)	2.776 (0.62)	1.642 (0.60)	**1.555** (0.57)
M6	0.6433	0.775 (0.27)	0.721 (0.27)	**0.655** (0.20)	0.733 (0.31)	0.677 (0.22)
M7	0.5942	0.634 (0.20)	0.643 (0.20)	1.871 (0.31)	0.613 (0.19)	**0.607** (0.19)
M8	0.5106	0.538 (0.18)	0.573 (0.22)	0.552 (0.18)	0.556 (0.22)	**0.524** (0.17)
M9	3.9353	7.525 (1.76)	4.242 (1.20)	11.752 (1.27)	4.392 (1.06)	**4.087** (1.09)
M10	1.4223	1.783 (0.47)	**1.567** (0.49)	2.016 (0.42)	1.647 (0.58)	1.667 (0.44)
M11	1.5512	1.951 (0.52)	1.687 (0.49)	2.888 (0.53)	1.858 (0.65)	**1.661** (0.47)
M12	2.3194	3.469 (0.92)	2.516 (0.76)	5.174 (0.82)	2.497 (0.70)	**2.421** (0.71)
M13	1.8303	2.070 (0.32)	**1.953** (0.33)	2.383 (0.17)	2.673 (0.91)	2.033 (0.41)
M14	1.7407	1.811 (0.47)	1.856 (0.51)	10.627 (1.13)	1.824 (0.46)	**1.761** (0.46)
M15	1.6979	**1.759** (0.34)	1.773 (0.36)	8.061 (0.73)	2.111 (0.55)	1.785 (0.35)
M16	1.2929	**1.338** (0.24)	1.344 (0.24)	3.675 (0.21)	1.677 (0.43)	1.704 (0.41)
M17	2.7758	3.619 (0.77)	2.927 (0.66)	7.752 (0.60)	3.122 (0.71)	**2.901** (0.65)
M18	7.7070	7.889 (1.78)	8.149 (1.92)	33.585 (1.08)	8.272 (1.77)	**7.803** (1.74)
M19	1.6228	1.801 (0.37)	**1.720** (0.35)	3.211 (0.33)	2.136 (0.76)	1.951 (0.75)
M20	1.7820	1.845 (0.48)	1.924 (0.54)	2.961 (0.43)	1.996 (0.57)	**1.828** (0.48)
M1	0.0000	0.013 (0.03)	0.013 (0.03	0.023 (0.02)	0.008 (0.01)	**0.007** (0.01)
M2	0.2002	0.214 (0.08)	0.217 (0.08	0.207 (0.08)	0.203 (0.07)	**0.202** (0.07)
M3	0.3069	0.356 (0.13)	0.328 (0.12	0.357 (0.12)	**0.311** (0.11)	0.315 (0.11)
M4	0.6790	0.983 (0.30)	0.714 (0.21	1.912 (0.32)	0.704 (0.20)	**0.699** (0.21)
M5	0.6403	0.960 (0.28)	0.671 (0.19	1.622 (0.26)	**0.658** (0.18)	0.661 (0.20)
M6	0.2789	0.324 (0.09)	0.297 (0.09	**0.288** (0.07)	0.327 (0.09)	0.295 (0.08)
M7	0.2559	0.270 (0.08)	0.268 (0.07	1.452 (0.28)	**0.258** (0.07)	0.261 (0.07)
M8	0.2196	0.227 (0.07)	0.234 (0.07	0.249 (0.07)	0.225 (0.06)	**0.221** (0.07)
M9	1.7392	3.617 (0.77)	1.811 (0.48)	7.790 (0.65)	1.827 (0.42)	**1.768** (0.46)
M10	0.6306	0.794 (0.20)	0.666 (0.18)	1.104 (0.19)	0.661 (0.17)	**0.655** (0.18)
M11	0.6922	0.869 (0.22)	0.722 (0.18)	1.687 (0.23)	0.733 (0.15)	**0.704** (0.17)
M12	1.0539	1.576 (0.36)	1.093 (0.24)	3.211 (0.37)	1.094 (0.25)	**1.093** (0.26)
M13	1.0348	1.302 (0.19)	1.069 (0.15)	2.045 (0.10)	1.136 (0.16)	**1.061** (0.16)
M14	0.7395	0.773 (0.18)	0.764 (0.18)	7.232 (0.51)	0.751 (0.16)	**0.742** (0.17)
M15	0.7442	**0.756** (0.12)	0.762 (0.13)	7.040 (0.77)	0.817 (0.13)	0.757 (0.12)
M16	0.5674	0.581 (0.08)	0.578 (0.08)	3.290 (0.24)	0.618 (0.10)	**0.573** (0.08)
M17	1.2668	1.696 (0.34)	1.304 (0.26)	5.657 (0.32)	1.350 (0.23)	**1.289** (0.26)
M18	3.2966	3.396 (0.67)	3.381 (0.67)	23.727 (0.59)	3.373 (0.64)	**3.311** (0.65)
M19	0.7614	0.851 (0.16)	**0.784** (0.14)	2.308 (0.17)	0.809 (0.14)	0.788 (0.16)
M20	0.7675	0.782 (0.17)	0.800 (0.18)	1.614 (0.19)	0.787 (0.17)	**0.772** (0.17)

Table 4.7: Comparative study for the spherical case, with up to down blocks corresponding to sample sizes 100, 250 and 1000, respectively. Columns of the selector $\bullet$ represent the MISE($\bullet$) $\times$ 100, with bold type for the minimum of the errors. The standard deviation of the ISE$\times$100 is given between parentheses.

Model	h_{MISE}	h_{LCV}	h_{LSCV}	h_{ROT}	h_{AMI}	h_{EMI}
M1	0.0000	0.006 (0.02)	0.006 (0.02)	0.008 (0.01)	0.008 (0.01)	**0.005** (0.01)
M2	0.2201	0.230 (0.06)	0.229 (0.06)	0.224 (0.05)	0.224 (0.05)	**0.221** (0.06)
M3	0.4536	0.528 (0.13)	0.470 (0.12)	**0.456** (0.11)	0.456 (0.11)	0.462 (0.12)
M4	1.2583	1.568 (0.23)	**1.281** (0.19)	2.350 (0.24)	1.291 (0.18)	1.288 (0.20)
M5	0.6392	0.846 (0.18)	0.657 (0.14)	0.909 (0.16)	0.667 (0.12)	**0.651** (0.14)
M6	0.3575	0.401 (0.07)	0.370 (0.06)	0.387 (0.06)	0.397 (0.08)	**0.370** (0.06)
M7	0.2808	0.294 (0.06)	0.288 (0.06)	1.425 (0.18)	**0.283** (0.06)	0.284 (0.06)
M8	0.2623	0.269 (0.06)	0.271 (0.06)	0.266 (0.06)	0.271 (0.05)	**0.264** (0.06)
M9	6.9786	17.556 (2.08)	7.139 (1.49)	19.477 (1.75)	7.303 (1.29)	**7.064** (1.44)
M10	1.2743	1.567 (0.25)	**1.302** (0.20)	1.903 (0.26)	1.308 (0.21)	1.303 (0.22)
M11	1.4724	2.117 (0.34)	1.506 (0.26)	2.465 (0.34)	1.612 (0.22)	**1.485** (0.26)
M12	1.4808	1.922 (0.26)	**1.506** (0.20)	2.500 (0.26)	1.522 (0.20)	1.512 (0.22)
M13	0.4761	0.509 (0.06)	**0.492** (0.05)	0.594 (0.04)	0.666 (0.22)	0.505 (0.07)
M14	1.4327	1.543 (0.25)	1.459 (0.24)	10.893 (0.94)	1.466 (0.22)	**1.436** (0.23)
M15	0.8662	0.883 (0.08)	0.879 (0.08)	4.351 (0.35)	0.993 (0.10)	**0.877** (0.08)
M16	0.5830	0.592 (0.06)	**0.591** (0.06)	1.895 (0.12)	0.636 (0.07)	0.706 (0.08)
M17	5.2376	10.032 (1.16)	5.373 (1.01)	19.474 (0.94)	5.506 (0.85)	**5.320** (0.92)
M18	13.3175	13.954 (2.18)	13.529 (2.13)	43.085 (2.21)	13.713 (2.09)	**13.357** (2.11)
M19	2.6298	3.743 (0.47)	**2.699** (0.48)	7.433 (0.43)	2.758 (0.45)	2.701 (0.48)
M20	2.1853	2.276 (0.33)	2.222 (0.32)	4.823 (0.32)	2.268 (0.29)	**2.193** (0.31)
M1	0.0000	**0.004** (0.01)	0.004 (0.01)	0.008 (0.01)	0.008 (0.01)	0.004 (0.00)
M2	0.2356	0.243 (0.05)	0.243 (0.05)	0.243 (0.05)	0.243 (0.05)	**0.237** (0.05)
M3	0.6224	0.705 (0.14)	0.637 (0.12)	0.629 (0.11)	0.629 (0.11)	**0.629** (0.12)
M4	1.8521	2.023 (0.17)	**1.870** (0.16)	2.899 (0.18)	1.927 (0.17)	1.886 (0.16)
M5	0.4269	0.488 (0.09)	0.437 (0.07)	0.479 (0.08)	0.495 (0.08)	**0.436** (0.08)
M6	0.4448	0.486 (0.06)	**0.455** (0.06)	0.507 (0.05)	0.507 (0.05)	0.463 (0.05)
M7	0.2765	0.287 (0.05)	0.282 (0.05)	0.819 (0.09)	0.284 (0.04)	**0.278** (0.05)
M8	0.2953	0.301 (0.05)	0.303 (0.05)	**0.296** (0.05)	0.311 (0.05)	0.297 (0.05)
M9	9.9498	19.713 (1.88)	11.588 (1.53)	19.502 (1.66)	14.847 (4.08)	**10.004** (1.43)
M10	2.3370	2.629 (0.23)	**2.364** (0.21)	3.338 (0.24)	2.521 (0.31)	2.369 (0.22)
M11	2.9424	4.518 (0.57)	2.993 (0.43)	4.557 (0.54)	3.358 (0.38)	**2.963** (0.43)
M12	1.9610	2.290 (0.21)	**1.980** (0.17)	2.769 (0.21)	2.071 (0.21)	1.996 (0.19)
M13	0.1732	0.180 (0.02)	**0.180** (0.02)	0.180 (0.02)	0.207 (0.12)	0.187 (0.02)
M14	2.2517	2.393 (0.28)	2.278 (0.26)	11.171 (0.81)	2.391 (0.25)	**2.255** (0.26)
M15	0.8263	0.842 (0.05)	0.836 (0.05)	2.384 (0.16)	0.883 (0.07)	**0.834** (0.05)
M16	0.5189	**0.525** (0.04)	0.526 (0.04)	1.069 (0.06)	0.527 (0.06)	0.568 (0.06)
M17	18.2152	38.755 (2.66)	18.815 (3.26)	55.977 (1.91)	19.331 (3.03)	**18.403** (2.94)
M18	74.2135	87.710 (16.21)	77.381 (16.68)	215.090 (11.43)	75.489 (16.94)	**74.833** (16.58)
M19	7.8653	12.311 (0.97)	8.093 (1.26)	18.594 (0.72)	8.325 (1.23)	**8.061** (1.22)
M20	4.1058	4.235 (0.44)	4.152 (0.44)	7.416 (0.45)	4.461 (0.42)	**4.115** (0.43)
M1	0.0000	**0.001** (0.00)	0.004 (0.01)	0.010 (0.01)	0.010 (0.01)	0.004 (0.00)
M2	0.2539	0.260 (0.05)	0.260 (0.05)	0.266 (0.04)	0.266 (0.04)	**0.255** (0.05)
M3	0.8986	1.004 (0.17)	0.912 (0.15)	0.915 (0.14)	0.915 (0.14)	**0.905** (0.15)
M4	2.4674	2.522 (0.14)	**2.484** (0.14)	3.340 (0.15)	2.620 (0.18)	2.521 (0.15)
M5	0.3296	0.356 (0.06)	**0.336** (0.05)	0.339 (0.05)	0.379 (0.07)	0.352 (0.06)
M6	0.5562	0.592 (0.07)	**0.565** (0.06)	0.666 (0.05)	0.666 (0.05)	0.586 (0.06)
M7	0.2715	0.280 (0.04)	0.276 (0.04)	0.551 (0.05)	0.287 (0.03)	**0.273** (0.04)
M8	0.3342	0.341 (0.05)	0.341 (0.05)	0.336 (0.05)	0.358 (0.04)	**0.335** (0.05)
M9	12.4415	23.477 (2.05)	20.881 (3.64)	21.104 (1.89)	31.948 (19.84)	**12.539** (1.57)
M10	2.6473	2.775 (0.17)	**2.669** (0.16)	3.108 (0.18)	2.800 (0.28)	2.964 (0.28)
M11	6.3810	10.020 (0.94)	6.454 (0.87)	9.398 (0.92)	7.191 (0.88)	**6.430** (0.82)
M12	2.4791	2.700 (0.18)	**2.496** (0.16)	3.072 (0.18)	2.580 (0.18)	2.569 (0.18)
M13	0.0704	0.085 (0.01)	0.076 (0.01)	**0.071** (0.01)	0.071 (0.01)	0.072 (0.01)
M14	3.1635	3.300 (0.33)	3.205 (0.31)	11.378 (0.73)	3.576 (0.30)	**3.167** (0.31)
M15	0.8038	0.816 (0.04)	**0.812** (0.04)	1.625 (0.09)	0.879 (0.22)	0.926 (0.28)
M16	0.4712	**0.476** (0.03)	0.477 (0.03)	0.738 (0.04)	0.683 (0.11)	0.775 (0.15)
M17	20.2938	37.587 (2.73)	22.014 (3.55)	52.149 (2.25)	31.485 (15.35)	**20.508** (3.34)
M18	74.4378	74.868 (13.64)	83.389 (16.68)	133.737 (9.48)	108.546 (36.67)	**74.639** (13.26)
M19	8.3636	11.984 (0.98)	8.832 (1.14)	17.672 (0.81)	12.430 (5.23)	**8.494** (1.08)
M20	7.4037	7.701 (0.73)	7.467 (0.70)	11.502 (0.68)	8.450 (0.82)	**7.415** (0.71)

Table 4.8: Comparative study for higher dimensions with sample size $n = 1000$: up to down blocks correspond to dimensions $q = 3, 4, 5$. Columns of the selector $\bullet$ represent the MISE($\bullet$) $\times 100$, with bold type for the minimum of the errors. The standard deviation of the ISE$\times 100$ is given between parentheses.

References

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, 12(2):171–178.

Bai, Z. D., Rao, C. R., and Zhao, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.*, 27(1):24–39.

Banerjee, A., Dhillon, I. S., Ghosh, J., and Sra, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions. *J. Mach. Learn. Res.*, 6:1345–1382.

Bingham, C. and Mardia, K. V. (1978). A small circle distribution on the sphere. *Biometrika*, 65(2):379–389.

Cabella, P. and Marinucci, D. (2009). Statistical challenges in the analysis of cosmic microwave background radiation. *Ann. Appl. Stat.*, 3(1):61–95.

Cao, R., Cuevas, A., and Gonzalez Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Comput. Statist. Data Anal.*, 17(2):153–176.

Chacón, J. E. and Duong, T. (2013). Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electron. J. Stat.*, 7:499–532.

Chiu, S.-T. (1996). A comparative review of bandwidth selection for kernel density estimation. *Statist. Sinica*, 6(1):129–145.

Ćwik, J. and Koronacki, J. (1997). A combined adaptive-mixtures/plug-in estimator of multivariate probability densities. *Comput. Statist. Data Anal.*, 26(2):199–218.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.*, 79(19):2066–2075.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2011). Kernel density estimation on the torus. *J. Statist. Plann. Inference*, 141(6):2156–2173.

Durastanti, C., Lan, X., and Marinucci, D. (2013). Needlet-Whittle estimates on the unit sphere. *Electron. J. Stat.*, 7:597–646.

Fernández-Durán, J. J. (2004). Circular distributions based on nonnegative trigonometric sums. *Biometrics*, 60(2):499–503.

Fernández-Durán, J. J. (2007). Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics*, 63(2):579–585.

Fernández-Durán, J. J. and Gregorio-Domínguez, M. M. (2010). Maximum likelihood estimation of nonnegative trigonometric sum models using a Newton-like algorithm on manifolds. *Electron. J. Stat.*, 4:1402–1410.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013a). Exploring wind direction and SO_2 concentration by circular-linear density estimation. *Stoch. Environ. Res. Risk Assess.*, 27(5):1055–1067.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013b). Kernel density estimation for directional-linear data. *J. Multivariate Anal.*, 121:152–175.

Hall, P., Watson, G. S., and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762.

Hornik, K. and Grün, B. (2014). movMF: an R package for fitting mixtures of von Mises-Fisher distributions. *J. Stat. Softw.*, 58(10):1–31.

Horová, I., Koláček, J., and Vopatová, K. (2013). Full bandwidth matrix selectors for gradient kernel density estimate. *Comput. Statist. Data Anal.*, 57(1):364–376.

Jammalamadaka, S. R. and Lund, U. J. (2006). The effect of wind direction on ozone levels: a case study. *Environ. Ecol. Stat.*, 13(3):287–298.

Johnson, M. E. (1987). *Multivariate statistical simulation*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. John Wiley & Sons, New York.

Jones, C., Marron, J. S., and Sheather, S. J. (1996). Progress in data-based bandwidth selection for kernel density estimation. *Computation. Stat.*, (11):337–381.

Jupp, P. E. and Mardia, K. V. (1989). A unified view of the theory of directional statistics, 1975-1988. *Int. Stat. Rev.*, 57(3):261–294.

Klemelä, J. (2000). Estimation of densities and derivatives of densities with directional data. *J. Multivariate Anal.*, 73(1):18–40.

Lebedev, V. I. and Laikov, D. N. (1995). A quadrature formula for the sphere of the 131st algebraic order of accuracy. *Dokl. Math.*, 59(3):477–481.

Mardia, K. V. and Jupp, P. E. (2000). *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, second edition.

Marron, J. S. and Wand, M. P. (1992). Exact mean integrated squared error. *Ann. Statist.*, 20(2):712–736.

Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2012). A plug-in rule for bandwidth selection in circular density estimation. *Comput. Statist. Data Anal.*, 56(12):3898–3908.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33(3):1065–1076.

Perryman, M. A. C. (1997). *The Hipparcos and Tycho catalogues*, volume 1200 of *ESA SP*. ESA Publication Division, Noordwijk.

Pewsey, A. (2006). Modelling asymmetrically distributed circular data using the wrapped skew-normal distribution. *Environ. Ecol. Stat.*, 13(3):257–269.

Pukkila, T. M. and Rao, C. R. (1988). Pattern recognition based on scale invariant discriminant functions. *Inform. Sci.*, 45(3):379–389.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27(3):832–837.

Scott, D. W. (1992). *Multivariate density estimation*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. John Wiley & Sons, New York.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Taylor, C. C. (2008). Automatic bandwidth selection for circular density estimation. *Comput. Statist. Data Anal.*, 52(7):3493–3500.

van Leeuwen, F. (2007). *Hipparcos, the new reduction of the raw data*, volume 350 of *Astrophysics and Space Science Library*. Springer, Dordrecht.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Watson, G. S. (1983). *Statistics on spheres*, volume 6 of *University of Arkansas Lecture Notes in the Mathematical Sciences*. John Wiley & Sons, New York.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Chapter 5

A nonparametric test for directional-linear independence

Abstract

A nonparametric test for assessing the independence between a directional random variable (circular or spherical, as particular cases) and a linear one is proposed in this paper. The statistic is based on the squared distance between nonparametric kernel density estimates and its calibration is done by a permutation approach. The size and power characteristics of various variants of the test are investigated and compared with those for classical correlation-based tests of independence in an extensive simulation study. Finally, the best-performing variant of the new test is applied in the analysis of the relation between the orientation and size of Portuguese wildfires.

Reference

García-Portugués, E., Barros, A. M. G., Crujeiras, R. M., González-Manteiga, W. and Pereira, J. M. C. (2014). A test for directional-linear independence, with applications to wildfire orientation and size. *Stoch. Environ. Res. Risk Assess.*, 28(5):1261–1275.

Contents

5.1 Introduction 108
5.2 Background to kernel density estimation 110
5.2.1 Linear kernel density estimation 110
5.2.2 Directional kernel density estimation 110
5.2.3 Directional-linear kernel density estimation. 111
5.3 A test for directional-linear independence 112
5.3.1 The test statistic 112
5.3.2 Calibration of the test 113
5.3.3 Simulation study 115
5.4 Real data analysis 120
5.4.1 Data description 120
5.4.2 Results 121
5.5 Discussion 122
5.A Proof of Lemma 5.1 124
References 125

5.1 Introduction

Characterization of wildfire orientation patterns at landscape scale has important management implications (Moreira et al., 2001; Lloret et al., 2002; Moreira et al., 2011). It has been shown that landscape fuel reduction treatments will only be successful if strategically placed in order to intersect fire spread in the heading direction (Finney, 2001; Schmidt et al., 2008).

Barros et al. (2012) assessed the existence of preferential fire perimeter orientation at watershed level, to support the spatial layout of fuelbreak networks. Their analysis identified clusters of watersheds where fire perimeters were preferentially aligned along the NE/SW and the SE/NW axes. Those watersheds included fire perimeters that together account for roughly 65% of the overall burnt area in Portugal, over the period from 1975 to 2005, while in the remaining watersheds fire perimeters were randomly aligned. In Figure 5.1, some descriptive maps of the data of interest are displayed. The left plot shows the total area burnt in each watershed, whereas the middle plot represents the mean slope of the fires in each region. Finally, the right plot indicates which watersheds exhibit a preferred fire orientation, versus a random orientation, according to Barros et al. (2012). The authors argued that spatial patterns of fire perimeter orientation found in the 31-year dataset could be explained by dominant weather during the Portuguese fire season (Pereira et al., 2005). However, given that fire perimeter orientation analysis is event-based (*i.e.*, it is based on the orientation of each fire event) all perimeters are treated equally independently of their size. In this paper, a test for assessing independence between wildfire size and orientation is presented, complementing the work of Barros et al. (2012). Furthermore, orientation of the wildfire will be considered in two-dimensional and three-dimensional spaces.

Spatial characterization of a wildfire, by means of its main orientation, and the associated burnt area, must be handled by non-standard statistical approaches, given the special nature of fire orientation. Specifically, it can be measured as an angle in the plane (two-dimensional orientation) or as a pair of angles identifying a direction in the three-dimensional sphere, if the main slope of the wildfire is taken into account. Hence, appropriate methods for handling circular and, more generally, directional data must be considered, jointly with suitable combinations of directional and linear techniques.

The analysis of the relation between directional and linear variables has been classically approached through the construction of circular-linear correlation coefficients. The adaptation of the classical linear correlation coefficient to the circular-linear setting was introduced by Mardia (1976) and Johnson and Wehrly (1977) and further studied by Liddell and Ord (1978), who obtained its exact distribution under certain parametric assumptions. For the circular-linear case, a rank-based test of association was also proposed by Mardia (1976), who derived its asymptotic distribution. Later, Fisher and Lee (1981) adapted Kendall's τ as a measure of circular-linear association based on the notion of concordance in the cylinder. To the best of the authors' knowledge, these three tests are the only available for testing the independence in directional-linear variables. As they are based on correlation coefficients, these tests are only powerful against deviations in the conditional expectation that can be measured by the corresponding coefficient. As a consequence, none of these tests are able to capture all possible types of dependence, neither for the conditional expectation nor for more complex types of dependence.

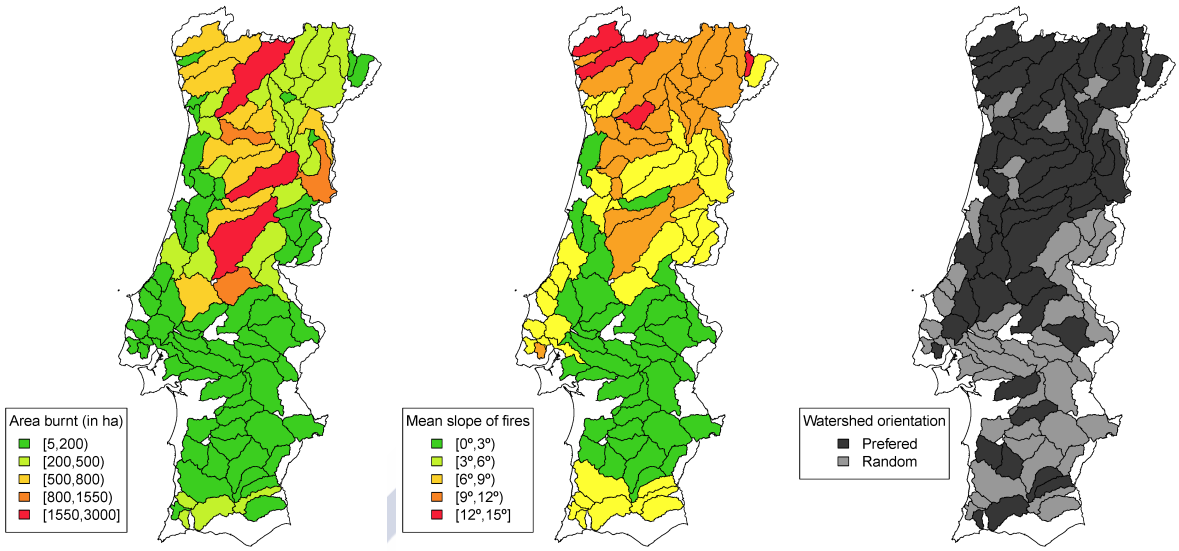


Figure 5.1: Descriptive maps of wildfires in Portugal with the 102 watersheds delineated by Barros et al. (2012). The left map shows the number of hectares burnt from fire perimeters associated with each watershed. Each fire perimeter is associated with the watershed that contains its centroid. The center map represents the mean slope of the fires of each watershed, where the slope is measured in degrees (0° stands for plain slope and 90° for a vertical one). Finally, the right map shows watersheds where fires display preferential alignment according to Barros et al. (2012).

From a different perspective, circular and linear variables can also be jointly modeled by the construction of circular-linear distributions. Johnson and Wehrly (1978) introduced a method for deriving circular-linear densities with specified marginals. A new family of circular-linear distributions based on nonnegative trigonometric sums, which proved to be more flexible in capturing the data structure, was proposed by Fernández-Durán (2007), adapting the method by Johnson and Wehrly (1978). More recently, García-Portugués et al. (2013a) exploited the copula representation of the Johnson and Wehrly (1978) family, allowing for a completely nonparametric estimator, which was applied to analyze SO_2 concentrations and wind direction. Nevertheless, the aforementioned methods are designed for the circular-linear case, whereas in our context, a more general tool for handling directional-linear relations is needed, provided that wildfire orientation may be reported in two or three dimensions.

In this paper, the assessment of the relation between a directional (circular or spherical, as particular cases) and a linear variable is approached through the construction of a formal test to check directional-linear independence. Inspired by the ideas of Rosenblatt (1975) and Rosenblatt and Wahlen (1992) for the linear setting (see also Ahmad and Li (1997)), the proposed test statistic is based on a nonparametric directional-linear kernel density estimator and an $\mathcal{L}_2$ distance is taken as a discrepancy measure between the joint estimator and the one constructed under the independence hypothesis. The new test presents some interesting advantages: it is designed in a general fashion for directional variables of all dimensions and it is able to capture all kinds of deviations from independence by virtue the nonparametric density estimation. Besides, one gets a kernel density estimate as a spin-off, which provides further information about the form of dependence when independence is rejected.

The remainder of the paper is organized as follows. In Section 5.2, some background to kernel density estimation, for linear, directional and directional-linear data is presented. Section 5.3 is devoted to the introduction of the test statistic, introducing a simplified version of the test and describing in detail its practical application. The finite sample performance of the test, in terms of size and power, is assessed through a simulation study for circular-linear and spherical-linear variables. Application to real data is provided in Section 5.4, including data description and results, focusing on the assessment of independence between wildfire orientation and burnt area size in Portugal. Some discussion and final comments are given in Section 5.5.

5.2 Background to kernel density estimation

In the linear setting, the basic building block for the independence test introduced by Rosenblatt (1975) is a kernel density estimator. Independence between two linear random variables is assessed through an $\mathcal{L}_2$ distance between a bidimensional kernel density estimator and the product of the marginal kernel density estimators. In order to extend such a procedure to the directional-linear case, kernel density estimation for linear, directional and directional-linear variables is required. A brief background on kernel density estimators will be provided in this section.

5.2.1 Linear kernel density estimation

The well-known kernel density estimator for linear data was introduced by Rosenblatt (1956) and Parzen (1962). Given a random sample $Z_1, \ldots, Z_n$ from a linear random variable Z (*i.e.* with support $\text{supp}(Z) \subseteq \mathbb{R}$) with density f, the kernel density estimator at a point $z \in \mathbb{R}$ is defined as

$$\hat{f}_g(z) = \frac{1}{ng} \sum_{i=1}^{n} K\left(\frac{z - Z_i}{g}\right),$$

where K is a kernel function, usually a symmetric density about the origin, and $g > 0$ is the smoothing or bandwidth parameter, which controls the roughness of the estimator. Properties of this estimator have been deeply studied (see Silverman (1986) or Wand and Jones (1995) for comprehensive reviews). It is also well known that the choice of kernel (normal, Epanechnikov, etc.) has little effect on the overall shape of the kernel density estimate. However, the bandwidth is a key tuning parameter: large values produce oversmoothed estimates of f, whereas small values provide undersmoothed curves. Comprehensive reviews on bandwidth selection are given in Cao et al. (1994), Chiu (1996) and Jones et al. (1996), among others.

5.2.2 Directional kernel density estimation

Denote by $\mathbf{X}$ a directional random variable with density f. The support of such a variable is the q-dimensional sphere, namely $\Omega_q = \{\mathbf{x} \in \mathbb{R}^{q+1} : x_1^2 + \cdots + x_{q+1}^2 = 1\}$, endowed with the Lebesgue measure in Ω_q, that will be denoted by ω_q. Therefore, a directional density is a nonnegative function that satisfies $\int_{\Omega_q} f(\mathbf{x})\, \omega_q(d\mathbf{x}) = 1$.

The directional kernel density estimator was introduced by Hall et al. (1987) and Bai et al. (1988). Given a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$, of a directional variable $\mathbf{X}$ with $\text{supp}(\mathbf{X}) \subseteq \Omega_q$ and

density f, at a point $\mathbf{x} \in \Omega_q$ the estimator is given by

$$\hat{f}_h(\mathbf{x}) = \frac{c_{h,q}(L)}{n} \sum_{i=1}^{n} L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right), \tag{5.1}$$

where L is the directional kernel, $h > 0$ is the bandwidth parameter and $c_{h,q}(L)$ is a normalizing constant depending on the kernel L, the bandwidth h and the sphere dimension q. The scalar product of two vectors, $\mathbf{x}$ and $\mathbf{y}$, is denoted by $\mathbf{x}^T\mathbf{y}$, where ^{T} denotes the transpose operator.

A common choice for the directional kernel is $L(r) = e^{-r}$, $r \geq 0$, also known as the von Mises kernel due to its relation with the von Mises-Fisher distribution (see Watson (1983)). In a q-dimensional sphere, the von Mises density $\mathrm{vM}(\boldsymbol{\mu}, \kappa)$ is given by

$$f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = C_q(\kappa) \exp\left\{\kappa \mathbf{x}^T \boldsymbol{\mu}\right\}, \quad C_q(\kappa) = \kappa^{\frac{q-1}{2}} \left[(2\pi)^{\frac{q+1}{2}} \mathcal{I}_{\frac{q-1}{2}}(\kappa)\right]^{-1}, \tag{5.2}$$

where $\boldsymbol{\mu} \in \Omega_q$ is the mean direction, $\kappa \geq 0$ is the concentration parameter around the mean and $\mathcal{I}_\nu$ is the modified Bessel function of order ν,

$$\mathcal{I}_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\pi^{1/2}\Gamma\left(\nu + \frac{1}{2}\right)} \int_{-1}^{1} (1-t^2)^{\nu - \frac{1}{2}} e^{zt}\, dt.$$

For the von Mises kernel, the value of $c_{h,q}(L)$ is $C_q\left(1/h^2\right) e^{1/h^2}$ and the directional estimator (5.1) can be interpreted as a mixture of von Mises-Fisher densities:

$$\hat{f}_h(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} f_{\mathrm{vM}}\left(\mathbf{x}; \mathbf{X}_i, 1/h^2\right).$$

Note that large values of h provide a small concentration parameter, which results in a uniform model in the sphere, whereas small values of h give high concentrations around the sample observations, providing an undersmoothed curve. Cross-validation rules based on Likelihood Cross Validation (LCV) and Least Squares Cross Validation (LSCV) for bandwidth selection were discussed by Hall et al. (1987).

5.2.3 Directional-linear kernel density estimation.

Consider a directional-linear random variable, $(\mathbf{X}, Z)$ with support $\mathrm{supp}(\mathbf{X}, Z) \subseteq \Omega_q \times \mathbb{R}$ and joint density f. For the simple case of circular data ($q = 1$), the support of the variable is the cylinder and, in general, the support is a multidimensional cylinder. Following the ideas in the previous sections for the linear and directional cases, given a random sample $(\mathbf{X}_1, Z_1), \ldots, (\mathbf{X}_n, Z_n)$, the directional-linear kernel density estimator at a point $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$ can be defined as

$$\hat{f}_{h,g}(\mathbf{x}, z) = \frac{c_{h,q}(L)}{ng} \sum_{i=1}^{n} LK\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}, \frac{z - Z_i}{g}\right), \tag{5.3}$$

where LK is a directional-linear kernel, $g > 0$ is the linear bandwidth parameter, $h > 0$ is the directional bandwidth and $c_{h,q}(L)$ is the directional normalizing constant. The estimator (5.3) was introduced by García-Portugués et al. (2013b), who also studied its asymptotic properties in terms of bias and variance, and established its asymptotic normality.

A product kernel $LK(\cdot,\cdot) = L(\cdot) \times K(\cdot)$, specifically, the von Mises-normal kernel

$$LK(r,t) = e^{-r} \times \phi_1(t), \quad r \in [0,\infty),\, t \in \mathbb{R},$$

will be considered throughout this paper in order to simplify computations, where ϕ_σ denotes the density of a normal with zero mean and standard deviation σ. Similarly to the linear and directional kernel density estimators, a smoothing parameter (bidimensional, in this case) is involved in the estimator construction. The cross-validation procedures introduced by Hall et al. (1987) can be adapted to the directional-linear setting, yielding the following bandwidth selectors:

$$(h,g)_{\mathrm{LCV}} = \arg\max_{h,g>0} \sum_{i=1}^{n} \log \hat{f}_{h,g}^{-i}(\mathbf{X}_i, Z_i),$$

$$(h,g)_{\mathrm{LSCV}} = \arg\max_{h,g>0} \left[2n^{-1} \sum_{i=1}^{n} \hat{f}_{h,g}^{-i}(\mathbf{X}_i, Z_i) - \int_{\Omega_q \times \mathbb{R}} \hat{f}_{h,g}(\mathbf{x},z)^2 \, dz \, \omega_q(d\mathbf{x}) \right],$$

where $f_{h,g}^{-i}$ represents the kernel density estimator computed without the i-th datum.

5.3 A test for directional-linear independence

The new test statistic for assessing independence between a directional and a linear variable is described in this section.

5.3.1 The test statistic

Consider the joint directional-linear density $f_{(\mathbf{X},Z)}$ for the variable $(\mathbf{X}, Z)$. $f_{\mathbf{X}}$ and f_Z denote the directional and linear marginal densities, respectively. The null hypothesis of independence between both components can be stated as

$$H_0 : f_{(\mathbf{X},Z)}(\mathbf{x},z) = f_{\mathbf{X}}(\mathbf{x}) f_Z(z), \quad \forall (\mathbf{x},z) \in \Omega_q \times \mathbb{R}$$

and the alternative hypothesis as

$$H_a : f_{(\mathbf{X},Z)}(\mathbf{x},z) \neq f_{\mathbf{X}}(\mathbf{x}) f_Z(z), \quad \text{for any } (\mathbf{x},z) \in \Omega_q \times \mathbb{R}.$$

Following the idea of Rosenblatt (1975), a natural statistic to test H_0 arises from considering the $\mathcal{L}_2$ distance between the nonparametric estimation of the joint density $f_{(\mathbf{X},Z)}$ by the directional-linear kernel estimator (5.3), denoted by $\hat{f}_{(\mathbf{X},Z);h,g}$, and the nonparametric estimation of $f_{(\mathbf{X},Z)}$ under H_0, given by the product of the marginal directional and linear kernel estimators, denoted by $\hat{f}_{\mathbf{X};h}$ and $\hat{f}_{Z;g}$, respectively. We therefore propose the following test statistic:

$$T_n = \Delta_2\Big(\hat{f}_{(\mathbf{X},Z);h,g}, \hat{f}_{\mathbf{X};h}\hat{f}_{Z;g}\Big), \tag{5.4}$$

where Δ_2 stands for the squared $\mathcal{L}_2$ distance in $\Omega_q \times \mathbb{R}$ between two functions f_1 and f_2:

$$\Delta_2(f_1,f_2) = \int_{\Omega_q \times \mathbb{R}} \left(f_1(\mathbf{x},z) - f_2(\mathbf{x},z)\right)^2 \omega_q(d\mathbf{x})\, dz.$$

The test statistic depends on a pair of bandwidths (h,g), which is used for the directional-linear estimator, and whose components are also considered for the marginal directional and

linear kernel density estimators. Under the null hypothesis of independence, H_0, it holds that $\mathbb{E}[\hat{f}_{(\mathbf{X},Z);h,g}(\mathbf{x},z)] = \mathbb{E}[\hat{f}_{\mathbf{X};h}(\mathbf{x})]\mathbb{E}[\hat{f}_{Z;g}(z)]$.

Asymptotic properties of (5.4) have been studied by García-Portugués et al. (2014), who proved its asymptotic normality under independence, but with a slow rate of convergence that does not encourage its use in practice. For that reason, a calibration mechanism will be needed for the practical application of the test.

In addition, the construction of T_n requires the calculation of an integral over $\Omega_q \times \mathbb{R}$, which may pose computational problems since it involves the calculation of several nested integrals. However, if the kernel estimators are obtained using von Mises and normal kernels, then an easy to compute expression for T_n can be obtained, as stated in the following lemma.

Lemma 5.1. *If the kernel estimators involved in (5.4), obtained from a random sample $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$ of $(\mathbf{X}, Z)$, are constructed with von Mises and normal kernels, the following expression for T_n holds:*

$$T_n = \mathbf{1}_n\left(\frac{1}{n^2}\boldsymbol{\Psi}(h)\circ\boldsymbol{\Omega}(g) - \frac{2}{n^3}\boldsymbol{\Psi}(h)\boldsymbol{\Omega}(g) + \frac{1}{n^4}\boldsymbol{\Psi}(h)\mathbf{1}_n\mathbf{1}_n^T\boldsymbol{\Omega}(g)\right)\mathbf{1}_n^T, \tag{5.5}$$

where $\circ$ denotes the Hadamard product and $\boldsymbol{\Psi}(h)$ and $\boldsymbol{\Omega}(g)$ are $n \times n$ matrices given by

$$\boldsymbol{\Psi}(h) = \left(\frac{C_q\left(1/h^2\right)^2}{C_q\left(||\mathbf{X}_i + \mathbf{X}_j||/h^2\right)}\right)_{ij}, \quad \boldsymbol{\Omega}(g) = \left(\phi_{\sqrt{2}g}\left(Z_i - Z_j\right)\right)_{ij},$$

where $\mathbf{1}_n$ is a vector of n ones and C_q is the normalizing function (5.2).

The proof of this result can be seen in Appendix 5.A. Note that expression (5.5) for T_n only requires matrix operations. This will be the expression used for computing the test statistic. It should also be noted that the effect of the dimension q appears only in the definition of C_q and in $||\mathbf{X}_i + \mathbf{X}_j||$, and both are easily scalable for large q. Thus, an important advantage of (5.5) is that computing requirements are similar for different dimensions q, something which is not the case if (5.4) is employed with numerical integration.

5.3.2 Calibration of the test

The null hypothesis of independence is stated in a nonparametric way, which determines the resampling methods used for calibration. However, as the null hypothesis is of a non-interaction kind, a permutation approach (which is not at all foreign to hypothesis testing) seems a reliable option. If $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$ is a random sample from the directional-linear variable $(\mathbf{X}, Z)$ and σ is a random permutation of n elements, then $\{(\mathbf{X}_i, Z_{\sigma(i)})\}_{i=1}^n$, represents the resulting σ-permuted sample. T_n^σ denotes the test statistic computed from the σ-permuted random sample. Under the assumption of independence between the directional and linear components, it is reasonable to expect that the distribution of T_n is similar to the distribution of T_n^σ, which can be easily approximated by Monte Carlo methods.

In addition to its simplicity, the main advantage of the use of permutations is its easy implementation using Lemma 5.1, as it is possible to reuse the computation of the matrices $\boldsymbol{\Psi}(h)$ and

$\boldsymbol{\Omega}(g)$ needed for T_n to compute a σ-permuted statistic T_n^σ. In virtue of expression (5.5) and the definition of T_n^σ, the σ-permuted test statistic is given by

$$T_n^\sigma = \mathbf{1}_n\left(\frac{1}{n^2}\boldsymbol{\Psi}(h)\circ\boldsymbol{\Omega}^\sigma(g) - \frac{2}{n^3}\boldsymbol{\Psi}(h)\boldsymbol{\Omega}^\sigma(g) + \frac{1}{n^4}\boldsymbol{\Psi}(h)\mathbf{1}_n\mathbf{1}_n^T\boldsymbol{\Omega}^\sigma(g)\right)\mathbf{1}_n^T,$$

where the ij-th entry of the matrix $\boldsymbol{\Omega}^\sigma(g)$ is the $\sigma(i)\sigma(j)$-entry of $\boldsymbol{\Omega}(g)$. For the computation of $\boldsymbol{\Psi}(h)$ and $\boldsymbol{\Omega}(g)$, symmetry properties reduce the number of computations and can also be used to optimize the products $\boldsymbol{\Psi}(h)\circ\boldsymbol{\Omega}^\sigma(g)$ and $\boldsymbol{\Psi}(h)\boldsymbol{\Omega}^\sigma(g)$. The last addend of T_n^σ is the same as that of T_n and there is no need to recompute it. The testing procedure can be summarized in the following algorithm.

Algorithm 5.1. *Let $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$ be a random sample from a directional-linear variable $(\mathbf{X}, Z)$.*

i. Obtain a suitable pair of bandwidths (h, g).

ii. Compute the observed value of T_n from (5.5), with kernel density estimators taking bandwidths (h, g).

iii. Permutation calibration. For $b = 1, \ldots, B \leq n!$, compute $T_n^{\sigma_b}$ with bandwidths (h, g) for a random permutation σ_b.

iv. Approximate the p-value by $\#\{T_n \leq T_n^{\sigma_b}\}/B$, where $\#$ denotes the cardinal of the set.

In steps *ii* and *iii*, a pair of bandwidths must be chosen. For the directional-linear case, as commented in Section 5.2, cross-validation bandwidths, namely $(h, g)_{\text{LCV}}$ and $(h, g)_{\text{LSCV}}$, can be considered. However, as usually happens with cross-validatory bandwidths, these selectors tend to provide undersmoothed estimators, something which a priori is not desirable as introduces a substantial variability in the statistic T_n.

To mitigate this problem, a more sophisticated bandwidth selector will be introduced. Considering the von Mises-normal kernel, the bootstrap version for the Mean Integrated Squared Error (MISE) of the directional-linear kernel density estimator (5.3) was derived by García-Portugués et al. (2013b):

$$\begin{aligned}\text{MISE}^*_{h_p,g_p}(h,g) = &\left(C_q(1/h^2)^2 C_q(2/h^2)^{-1} 2\pi^{\frac{1}{2}} gn\right)^{-1} \\ &+ n^{-2}\mathbf{1}_n\left[(1-n^{-1})\boldsymbol{\Psi}^*_{\mathbf{2}}(h)\circ\boldsymbol{\Omega}^*_{\mathbf{2}}(g) - 2\boldsymbol{\Psi}^*_{\mathbf{1}}(h)\circ\boldsymbol{\Omega}^*_{\mathbf{1}}(g) + \boldsymbol{\Psi}^*_{\mathbf{0}}\circ\boldsymbol{\Omega}^*_{\mathbf{0}}\right]\mathbf{1}_n^T,\end{aligned}$$

where matrices $\boldsymbol{\Psi}^*_{\mathbf{a}}(h)$ and $\boldsymbol{\Omega}^*_{\mathbf{a}}(g)$, $a = 0, 1, 2$ are

$$\boldsymbol{\Psi}^*_{\mathbf{0}} = \left(\frac{C_q(1/h_p^2)^2}{C_q(||\mathbf{X}_i + \mathbf{X}_j||/h_p^2)}\right)_{ij}, \quad \boldsymbol{\Psi}^*_{\mathbf{1}}(h) = \left(\int_{\Omega_q}\frac{C_q(1/h^2)C_q(1/h_p^2)^2 e^{\mathbf{x}^T\mathbf{X}_j/h_p^2}}{C_q(||\mathbf{x}/h^2 + \mathbf{X}_i/h_p^2||)}\,\omega_q(d\mathbf{x})\right)_{ij},$$

$$\boldsymbol{\Psi}^*_{\mathbf{2}}(h) = \left(\int_{\Omega_q}\frac{C_q(1/h^2)^2 C_q(1/h_p^2)^2}{C_q(||\mathbf{x}/h^2 + \mathbf{X}_i/h_p^2||)C_q(||\mathbf{x}/h^2 + \mathbf{X}_j/h_p^2||)}\,\omega_q(d\mathbf{x})\right)_{ij},$$

$$\boldsymbol{\Omega}^*_{\mathbf{0}} = \left(\phi_{\sqrt{2}g_p}(Z_i - Z_j)\right)_{ij}, \quad \boldsymbol{\Omega}^*_{\mathbf{a}}(g) = \left(\phi_{\sigma_{a,g}}(Z_i - Z_j)\right)_{ij},$$

with $\sigma_{a,g} = (ag^2 + 2g_p^2)^{\frac{1}{2}}$, $a = 1, 2$, and (h_p, g_p) a given pair of pilot bandwidths. Then, the estimation bandwidths are obtained as

$$(h, g)_{bo} = \arg\min_{h,g>0}\text{MISE}^*_{h_p,g_p}(h, g).$$

The choice of (h_p, g_p) is needed in order to compute $(h, g)_{bo}$. This must be done by a joint criterion for two important reasons. Firstly, to avoid the predominance of smoothing in one component that may dominate the other (this could happen, for example, if the directional variable is uniform, as in that case the optimal bandwidth tends to infinity). Secondly, to obtain a test with more power against deviations from independence. Based on these comments, a new bandwidth selector, named Bootstrap Likelihood Cross Validation (BLCV), is introduced:

$$(h, g)_{\text{BLCV}} = \arg\min_{h,g>0} \text{MISE}^*_{(h,g)_{\text{MLCV}}}(h, g),$$

where the pair of bandwidths $(h, g)_{\text{MLCV}}$ are obtained by enlarging the order of $(h, g)_{\text{LCV}}$ to be of the kind $(\mathcal{O}(n^{-1/(6+q)}), \mathcal{O}(n^{-1/7}))$, the order that one would expect for a pair of directional-linear pilot bandwidths. For the linear component, this can be seen in the paper by Cao (1993), where the pilot bandwidth is proved to be $g_p = \mathcal{O}(n^{-1/7})$, larger than the order of the optimal estimation bandwidth, $n^{-1/5}$. For the directional case there is no pilot bandwidth available, but considering that the order of the optimal estimation bandwidth is $n^{-1/(4+q)}$ (García-Portugués et al., 2013b), then a plausible conjecture is $h_p = \mathcal{O}(n^{-1/(6+q)})$.

5.3.3 Simulation study

Six different directional-linear models were considered in the simulation study. The models are indexed by a δ parameter that measures the degree of deviation from the independence, where $\delta = 0$ represents independence and $\delta > 0$ accounts for different degrees of dependence. The models show three kind of possible deviations from the independence: first order deviations, that is, deviations in the conditional expectation (M1, M2 and M3); second order deviations or conditional variance deviations (M4 and M5) and first and second order deviations (M6). In order to clarify notation, $\phi(\cdot; m, \sigma)$ and $f_{\mathcal{LN}}(\cdot; m, \sigma)$ represent the density of a normal and a log-normal, with mean/log-scale m and standard deviation/shape σ. Notation $\mathbf{0}_q$ represents a vector of q zeros.

M1. $f_1(\mathbf{x}, z) = \phi(z; \delta(2 + \mathbf{x}^T\boldsymbol{\mu}), \sigma) \times f_{\text{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa)$, with $\boldsymbol{\mu} = (\mathbf{0}_q, 1)$, $\kappa = 1$ and $\sigma = 1$.

M2. $f_2(\mathbf{x}, z) = f_{\mathcal{LN}}(z; \delta(1 + (\mathbf{x}^T\boldsymbol{\mu})^2), \sigma) \times f_{\text{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa)$, with $\boldsymbol{\mu} = (-1, \mathbf{0}_q)$, $\kappa = 0$ and $\sigma = \frac{1}{4}$.

M3. $f_3(\mathbf{x}, z) = [r f_{\mathcal{LN}}(z; \delta(1 + (\mathbf{x}^T\boldsymbol{\mu}_1)^3), \sigma_1) + (1 - r)\phi(z; m, \sigma_2)] \times [p f_{\text{vM}}(\mathbf{x}; \boldsymbol{\mu}_1, \kappa_1) + (1 - p) \times f_{\text{vM}}(\mathbf{x}; \boldsymbol{\mu}_2, \kappa_2)]$, with $\boldsymbol{\mu}_1 = (\mathbf{0}_q, 1)$, $\boldsymbol{\mu}_2 = (\mathbf{0}_q, -1)$, $\kappa_1 = 2$, $\kappa_2 = 1$, $p = \frac{3}{4}$, $\sigma_1 = \sigma_2 = \frac{1}{4}$, $m = 1$ and $r = \frac{1}{4}$.

M4. $f_4(\mathbf{x}, z) = \phi(z; m, \frac{1}{4} + \delta(1 - (\mathbf{x}^T\boldsymbol{\mu}_r)^3)) \times f_{\text{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa)$, with $\boldsymbol{\mu} = (\mathbf{0}_q, 1)$, $\kappa = 1$, $\boldsymbol{\mu}_r = (-1, \mathbf{0}_q)$ and $m = 0$.

M5. $f_5(\mathbf{x}, z) = f_{\mathcal{LN}}(z; m, (5 + \delta\mathbf{x}^T(3\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1))^{-1}) \times [p f_{\text{vM}}(\mathbf{x}; \boldsymbol{\mu}_1, \kappa_1) + (1 - p) f_{\text{vM}}(\mathbf{x}; \boldsymbol{\mu}_2, \kappa_2)]$, with $\boldsymbol{\mu}_1 = (\mathbf{0}_q, 1)$, $\boldsymbol{\mu}_2 = (\mathbf{0}_q, -1)$, $\kappa_1 = \kappa_2 = 2$, $p = \frac{1}{2}$ and $m = 0$.

M6. $f_6(\mathbf{x}, z) = [r f_{\mathcal{LN}}(z; m, \sigma) + (1 - r)\phi(z; \delta(2 + \mathbf{x}^T\boldsymbol{\mu}_r), \frac{1}{4} + \delta(\mathbf{x}^T\boldsymbol{\mu}_r)^2)] \times f_{\text{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa)$, with $\boldsymbol{\mu} = (\mathbf{0}_q, 1)$, $\boldsymbol{\mu}_r = (-1, \mathbf{0}_q)$, $\kappa = 1$, $p = \frac{3}{4}$, $\sigma = \frac{1}{2}$, $m = 0$ and $r = \frac{3}{4}$.

The choice of the models was done in order to capture situations with heteroskedasticity, skewness in the linear component and different types of von Mises mixtures in the directional component.

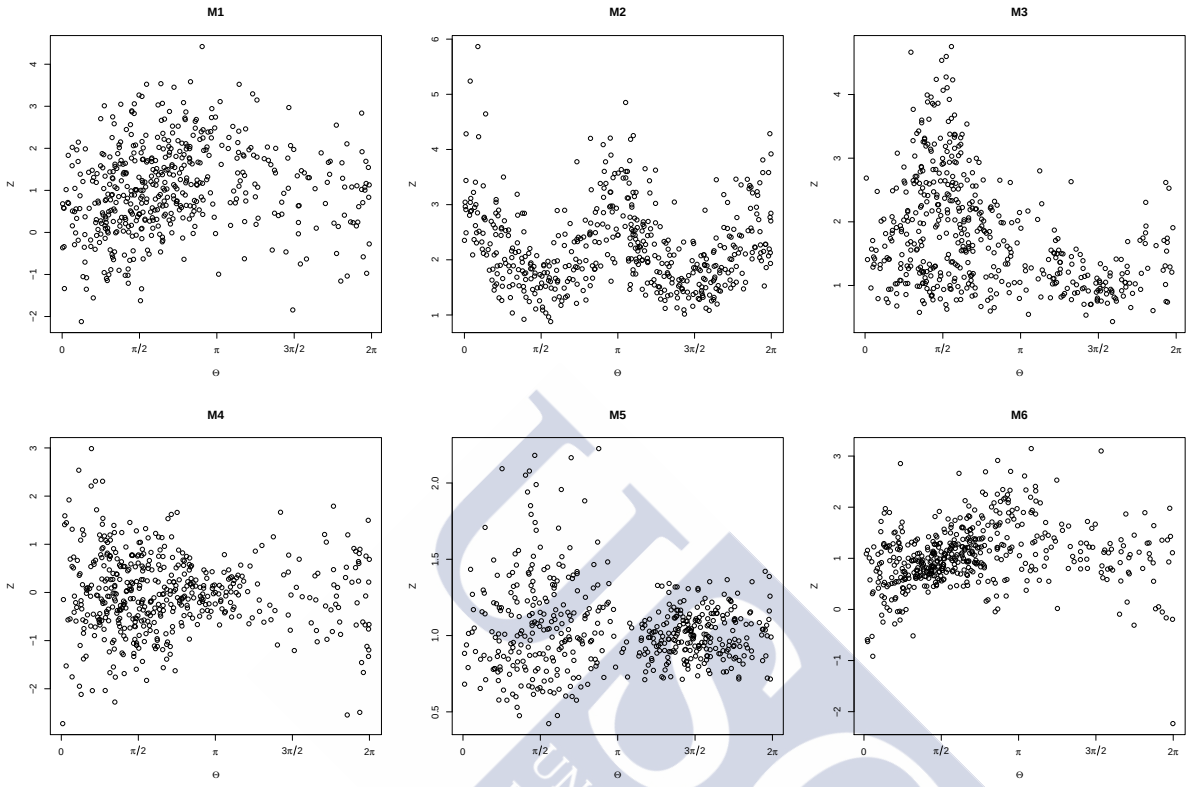


Figure 5.2: Random samples of $n = 500$ points for the simulation models in the circular-linear case, with $\delta = 0.50$ (situation with dependence). From left to right and up to down, M1 to M6. M1, M2 and M3 present a deviation from the independence in terms of the conditional expectation; M4 and M5 account for a deviation in terms of the conditional variance and M6 includes deviations both in conditional expectation and variance.

For the proposed models, different deviations from independence have been considered, by setting $\delta = 0, 0.25, 0.50$. The proposed test statistic has been computed for all the models and sample sizes $n = 50$, 100, 200, 500, 1000. The new test based on the permutation resampling described in Algorithm 5.1, depending on the bandwidth choice, is denoted by T_n^{LCV} and T_n^{BLCV}. The number of permutations considered was $B = 1000$ and the number of Monte Carlo replicates was $M = 1000$. Both the circular-linear and spherical-linear cases were explored. For the circular-linear case, the test was compared with the three tests available for circular-linear association, described as follows:

- Circular-linear correlation coefficient from Mardia (1976) and Johnson and Wehrly (1977), denoted by R_n^2.
- Rank circular-linear correlation coefficient from Mardia (1976), denoted by U_n.
- λ_{4n} measure of cylindrical association of Fisher and Lee (1981), implemented with its incomplete version λ_{4n}^* considering $m = 5000$ random 4-tuples.

Although there exists an exact distribution for R_n^2 under certain normality assumptions on the linear response and asymptotic distributions for U_n and λ_{4n}^*, for a fair comparison, the calibration of these tests has also been done by permutations ($B = 1000$). The exact and asymptotic

distributions for R_n^2 and U_n were also tried instead of the permutation approach, providing empirical levels and powers quite similar to the ones based on permutations.

The proportion of rejections under $H_{k,\delta}$ (for model number k with δ deviation) is reported in Tables 5.1 and 5.2, for the circular-linear and spherical-linear cases, with different sample sizes. In the circular-linear case, the empirical size is close to the nominal level for all the competing tests. The T_n^{LCV} test for this case shows in general a satisfactory behavior under the null hypothesis, except for some cases in M1, M4 and M5, where the test tends to reject the null hypothesis more than expected. This is mostly corrected by T_n^{BLCV}, with a decrease of power with respect to T_n^{LCV} in M2. For the spherical-linear case, the improvement in size approximation T_n^{BLCV} is notable, specially for small sample sizes. If the tests maintain the nominal significance level of 5%, it is expected that approximately 95% of the observed proportions of rejections under the null hypothesis (*i.e.* when $\delta = 0$) to lie within the interval $(0.036, 0.064)$ to three decimal places.

Regarding power, the test for R_n^2 is the most powerful one for M1 and M3, although the performance of T_n^{LCV} and T_n^{BLCV}, specially for M3, is quite similar. This was to be expected, as the circular-linear association tests should present more power against deviations of the first order. However, for M2, M4 and M5, all these tests are not able to distinguish the alternatives and the rejection ratios are close to the nominal level, resulting in λ_{4n}^* being the test with better behavior among them. In contrast, T_n^{LCV} and T_n^{BLCV} correctly detect the deviations from the null. In M6, R_n^2 is only the most competitive for the situation with $n = 50$, with T_n^{LCV} and T_n^{BLCV} the most competitive for the remainder of situations. U_n shows a similar performance to R_n^2, but with more power in M2 and less in M5. λ_{4n}^* is less affected than R_n^2 and U_n by the change of models, but also has lower power than them for M1, M3 and M6. The results for the spherical-linear case are quite similar to the previous ones for the empirical size, but with lower power in comparison with the circular-linear scenario, something expected as a consequence of the difference in dimensionality.

Some final comments on the simulation results follow. For the different sample sizes and dimensions, the running times for T_n^{LCV} and T_n^{BLCV} are collected in Table 5.3. Computation times for T_n^{LCV} are very similar for different dimensions q, whereas T_n^{BLCV} is affected by q due to the choice of the bandwidths $(h, g)_{\mathrm{BLCV}}$. The choice of the kernels was corroborated to be non-important for testing, as similar results were obtained for the test T_n^{LCV} using the directional-linear kernel $LK(r,t) = (1-r)\mathbf{1}_{[0,1]}(r) \times \frac{3}{4}(1-t^2)\mathbf{1}_{[-1,1]}(t)$. Cross-validatory bandwidths LSCV and BLSCV were also tried in the simulation study, providing worse results (this is also what usually happens with directional data, as it can be seen in García-Portugués (2013)). Finally, it is worth mentioning that bootstrap calibration was also tried as an alternative to the permutation approach, using a pair of bandwidths for estimation and another pair for the smooth resampling. The results in terms of size, power and computing times were substantially worse than the ones obtained for permutations.

In conclusion, both T_n^{LCV} and T_n^{BLCV} tests show a competitive behavior in all the simulation models, sample sizes and dimensions considered, only being outperformed by R_n^2 in M1 and M3. Nevertheless, for those models, the rejection rates of both tests are in general close to the ones of R_n^2. The test T_n^{BLCV} corrects the over rejection of T_n^{LCV} in certain simulation models, without a significant loss in power but at the expense of a high computational cost. Finally, the classical tests R_n^2, U_n and λ_{4n}^* presented critical problems on detecting second order and some

n	Model	Circular-linear					Spherical-linear	
		R_n^2	U_n	λ_{4n}^*	T_n^{LCV}	T_n^{BLCV}	T_n^{LCV}	T_n^{BLCV}
50	$H_{1,0.00}$	0.047	0.040	0.047	0.050	0.047	0.059	0.045
	$H_{2,0.00}$	0.051	0.044	0.042	0.055	0.052	0.057	0.052
	$H_{3,0.00}$	0.047	0.045	0.051	0.059	0.054	0.051	0.048
	$H_{4,0.00}$	0.047	0.040	0.047	0.050	0.046	0.059	0.046
	$H_{5,0.00}$	0.042	0.045	0.053	0.055	0.047	0.070	0.057
	$H_{6,0.00}$	0.058	0.065	0.062	0.057	0.055	0.065	0.054
	$H_{1,0.25}$	0.162	0.120	0.092	0.132	0.139	0.099	0.094
	$H_{2,0.25}$	0.055	0.074	0.088	0.143	0.071	0.072	0.050
	$H_{3,0.25}$	0.535	0.538	0.365	0.511	0.543	0.238	0.246
	$H_{4,0.25}$	0.051	0.044	0.067	0.239	0.234	0.103	0.097
	$H_{5,0.25}$	0.046	0.049	0.059	0.128	0.121	0.110	0.094
	$H_{6,0.25}$	0.354	0.332	0.239	0.436	0.432	0.284	0.275
	$H_{1,0.50}$	0.512	0.412	0.235	0.378	0.421	0.231	0.253
	$H_{2,0.50}$	0.054	0.124	0.261	0.633	0.291	0.219	0.078
	$H_{3,0.50}$	0.925	0.845	0.734	0.929	0.949	0.662	0.666
	$H_{4,0.50}$	0.058	0.050	0.081	0.424	0.420	0.149	0.139
	$H_{5,0.50}$	0.055	0.059	0.094	0.501	0.491	0.320	0.298
	$H_{6,0.50}$	0.782	0.706	0.536	0.754	0.756	0.556	0.540
100	$H_{1,0.00}$	0.052	0.054	0.063	0.068	0.061	0.072	0.068
	$H_{2,0.00}$	0.044	0.046	0.052	0.053	0.048	0.051	0.055
	$H_{3,0.00}$	0.047	0.050	0.046	0.061	0.054	0.064	0.049
	$H_{4,0.00}$	0.052	0.054	0.063	0.067	0.060	0.072	0.071
	$H_{5,0.00}$	0.056	0.050	0.057	0.073	0.063	0.074	0.063
	$H_{6,0.00}$	0.046	0.046	0.050	0.062	0.059	0.077	0.071
	$H_{1,0.25}$	0.291	0.227	0.102	0.211	0.213	0.155	0.163
	$H_{2,0.25}$	0.051	0.073	0.092	0.263	0.094	0.114	0.067
	$H_{3,0.25}$	0.889	0.851	0.407	0.805	0.849	0.487	0.500
	$H_{4,0.25}$	0.060	0.049	0.074	0.478	0.484	0.222	0.219
	$H_{5,0.25}$	0.063	0.050	0.067	0.260	0.251	0.171	0.171
	$H_{6,0.25}$	0.547	0.574	0.283	0.720	0.718	0.492	0.479
	$H_{1,0.50}$	0.847	0.721	0.290	0.669	0.718	0.416	0.460
	$H_{2,0.50}$	0.053	0.122	0.279	0.940	0.660	0.530	0.123
	$H_{3,0.50}$	1.000	0.997	0.872	0.999	0.999	0.942	0.957
	$H_{4,0.50}$	0.058	0.053	0.103	0.784	0.803	0.341	0.355
	$H_{5,0.50}$	0.083	0.056	0.107	0.836	0.860	0.602	0.630
	$H_{6,0.50}$	0.965	0.951	0.642	0.968	0.967	0.864	0.845
200	$H_{1,0.00}$	0.049	0.056	0.064	0.057	0.054	0.065	0.060
	$H_{2,0.00}$	0.055	0.063	0.055	0.053	0.054	0.046	0.041
	$H_{3,0.00}$	0.051	0.054	0.058	0.053	0.050	0.045	0.042
	$H_{4,0.00}$	0.049	0.056	0.064	0.057	0.053	0.065	0.057
	$H_{5,0.00}$	0.049	0.056	0.043	0.066	0.063	0.060	0.061
	$H_{6,0.00}$	0.048	0.059	0.052	0.049	0.046	0.054	0.048
	$H_{1,0.25}$	0.529	0.444	0.099	0.349	0.373	0.192	0.208
	$H_{2,0.25}$	0.058	0.081	0.106	0.551	0.154	0.178	0.052
	$H_{3,0.25}$	0.996	0.995	0.431	0.980	0.987	0.795	0.818
	$H_{4,0.25}$	0.054	0.057	0.085	0.839	0.862	0.337	0.348
	$H_{5,0.25}$	0.056	0.052	0.058	0.459	0.487	0.303	0.343
	$H_{6,0.25}$	0.830	0.896	0.277	0.974	0.971	0.842	0.830
	$H_{1,0.50}$	0.982	0.957	0.299	0.924	0.940	0.721	0.750
	$H_{2,0.50}$	0.061	0.145	0.325	0.999	0.967	0.899	0.249
	$H_{3,0.50}$	1.000	1.000	0.913	1.000	1.000	1.000	1.000
	$H_{4,0.50}$	0.053	0.058	0.122	0.981	0.984	0.595	0.618
	$H_{5,0.50}$	0.124	0.051	0.105	0.991	0.995	0.921	0.950
	$H_{6,0.50}$	1.000	1.000	0.691	1.000	1.000	0.994	0.993

Table 5.1: Proportion of rejections for the R_n^2, U_n, λ_{4n}^*, T_n^{LCV} and T_n^{BLCV} tests of independence for sample sizes $n = 50, 100, 200$ for a nominal significance level of 5%. For the six different models the values of the deviation from independence parameter are $\delta = 0$ (independence), 0.25 and 0.50. Each proportion was calculated using $B = 1000$ permutations for each of $M = 1000$ random samples of size n simulated from the specified model.

n	Model	Circular-linear					Spherical-linear	
		R_n^2	U_n	λ_{4n}^*	T_n^{LCV}	T_n^{BLCV}	T_n^{LCV}	T_n^{BLCV}
500	$H_{1,0.00}$	0.053	0.060	0.054	0.069	0.064	0.055	0.045
	$H_{2,0.00}$	0.060	0.050	0.053	0.062	0.055	0.046	0.048
	$H_{3,0.00}$	0.054	0.059	0.064	0.050	0.044	0.045	0.042
	$H_{4,0.00}$	0.053	0.060	0.054	0.069	0.062	0.056	0.050
	$H_{5,0.00}$	0.042	0.038	0.058	0.050	0.047	0.051	0.059
	$H_{6,0.00}$	0.052	0.050	0.053	0.059	0.062	0.059	0.055
	$H_{1,0.25}$	0.916	0.842	0.088	0.698	0.727	0.422	0.447
	$H_{2,0.25}$	0.050	0.073	0.095	0.973	0.511	0.557	0.073
	$H_{3,0.25}$	1.000	1.000	0.443	1.000	1.000	0.995	0.997
	$H_{4,0.25}$	0.057	0.060	0.077	0.999	0.999	0.764	0.786
	$H_{5,0.25}$	0.080	0.038	0.068	0.850	0.865	0.679	0.750
	$H_{6,0.25}$	0.998	1.000	0.263	1.000	1.000	1.000	0.998
	$H_{1,0.50}$	1.000	1.000	0.264	0.998	0.999	0.982	0.986
	$H_{2,0.50}$	0.053	0.125	0.322	1.000	1.000	1.000	0.910
	$H_{3,0.50}$	1.000	1.000	0.942	1.000	1.000	1.000	1.000
	$H_{4,0.50}$	0.064	0.060	0.090	1.000	1.000	0.982	0.987
	$H_{5,0.50}$	0.258	0.043	0.108	1.000	1.000	1.000	1.000
	$H_{6,0.50}$	1.000	1.000	0.709	1.000	1.000	1.000	1.000
1000	$H_{1,0.00}$	0.059	0.053	0.060	0.056	0.057	0.061	0.060
	$H_{2,0.00}$	0.043	0.042	0.070	0.045	0.046	0.058	0.051
	$H_{3,0.00}$	0.063	0.054	0.062	0.057	0.054	0.038	0.037
	$H_{4,0.00}$	0.059	0.053	0.060	0.056	0.054	0.061	0.054
	$H_{5,0.00}$	0.055	0.060	0.047	0.053	0.051	0.078	0.074
	$H_{6,0.00}$	0.045	0.047	0.054	0.057	0.058	0.052	0.048
	$H_{1,0.25}$	0.997	0.992	0.084	0.938	0.947	0.730	0.747
	$H_{2,0.25}$	0.046	0.067	0.109	1.000	0.910	0.947	0.123
	$H_{3,0.25}$	1.000	1.000	0.459	1.000	1.000	1.000	1.000
	$H_{4,0.25}$	0.061	0.052	0.074	1.000	1.000	0.989	0.991
	$H_{5,0.25}$	0.129	0.059	0.059	0.993	0.995	0.936	0.971
	$H_{6,0.25}$	1.000	1.000	0.257	1.000	1.000	1.000	1.000
	$H_{1,0.50}$	1.000	1.000	0.281	1.000	1.000	1.000	1.000
	$H_{2,0.50}$	0.049	0.125	0.305	1.000	1.000	1.000	1.000
	$H_{3,0.50}$	1.000	1.000	0.954	1.000	1.000	1.000	1.000
	$H_{4,0.50}$	0.058	0.057	0.099	1.000	1.000	1.000	1.000
	$H_{5,0.50}$	0.486	0.058	0.106	1.000	1.000	1.000	1.000
	$H_{6,0.50}$	1.000	1.000	0.751	1.000	1.000	1.000	1.000

Table 5.2: Proportion of rejections for the R_n^2, U_n, λ_{4n}^*, T_n^{LCV} and T_n^{BLCV} tests of independence for sample sizes $n = 500, 1000$ for a nominal significance level of 5%. For the six different models the values of the deviation from independence parameter are $\delta = 0$ (independence), 0.25 and 0.50. Each proportion was calculated using $B = 1000$ permutations for each of $M = 1000$ random samples of size n simulated from the specified model.

Test	q	Sample size				
		50	100	200	500	1000
T_n^{LCV}	1	0.17	0.25	0.93	6.15	28.41
	2	0.09	0.27	0.93	6.14	28.83
T_n^{BLCV}	1	0.66	1.11	2.05	9.12	33.79
	2	4.27	5.18	9.98	26.27	71.89

Table 5.3: Computing times (in seconds) for T_n^{LCV} and T_n^{BLCV} as a function of sample size and dimension q, with $q = 1$ for the circular-linear case and $q = 2$ for the spherical-linear case. The tests were run with $B = 1000$ permutations and the times were measured in a 3.5 GHz core.

first order deviations from the independence. For all those reasons, the final recommendation is to preferably use the test T_n^{BLCV} for inference on directional-linear independence and T_n^{LCV} for a less computing intensive exploratory analysis.

5.4 Real data analysis

5.4.1 Data description

The original Portuguese fire atlas, covering the period from 1975 to 2005, is the longest annual and country-wide cartographic fire database in Europe (Pereira and Santos, 2003). Annual wildfire maps were derived from Landsat data, which represents the world's longest and continuously acquired collection of moderate resolution land remote sensing data, providing a unique resource for those who work in forestry, mapping and global change research. For each year in the dataset, Landsat imagery covering Portugal's mainland was acquired after the end of the fire season, thus providing a snapshot of the fires that occurred during the season. Annual fire perimeters were derived through a semi-automatic procedure that starts with supervised image classification, followed by manual editing of classification results. Minimum Mapping Unit (MMU), *i.e.*, the size of the smallest fire mapped, changed according to available data. Between 1975 and 1983 (the MultiSpectral Scanner era), spatial resolution of satellite images is 80 meters and MMU of 35 hectares. From 1984 onwards with data availability at spatial resolution of 30 meters (Thematic Mapper and Enhanced Thematic Mapper era) MMU is 5 hectares, allowing to map a larger number of smaller fires than in the 1975–1983 era. Below an MMU of approximately 5 hectares the burnt area classification errors increase substantially, and given the very skewed nature of fire size distribution, the 5 hectares threshold ensures that over 90% of total area actually burnt is mapped. For consistency, and due to discrepancies in minimum mapping unit between 1975–1983 and 1985–2005, in this study only fire perimeters mapped in the latter period were considered, which results in 26870 fire perimeters.

This application is based on the watershed delineation proposed by Barros et al. (2012). In their work, watersheds were derived from the Shuttle Radar Topography Mission (SRTM) digital terrain model (Farr et al., 2007) using the ArcGIS hydrology toolbox (ESRI, 2009). Minimum watershed size was interactively increased so that each watershed contained a minimum of 25 fire observations (see the cited work for more details). Fire perimeters straddling watershed boundaries were allocated to the watershed that contained its centroid.

The orientation of fire perimeters and watersheds was determined by principal component analysis, following the approach proposed by Luo (1998, pages 131–136). Specifically, principal component analysis was applied to the points that constitute the object's boundary (fire or watershed), with orientation given by the first principal component (PC1). Boundary points can be represented either in bidimensional space defined by each vertex's latitude and longitude coordinates, or in tridimensional space, taking also into account the altitude. Then, the PC1 corresponds to an axis that passes through the center of mass of the object and maximizes the variance of the projected vertices, represented in $\mathbb{R}^2$ or in $\mathbb{R}^3$. The fact of computing the PC1 also in $\mathbb{R}^3$ aims to take into account the variability of fires according to their slope, which, as the center plot of Figure 5.1 shows, presents marked differences between regions. Then, the orientation of the object is taken as the direction given by its PC1.

It is important to notice that an orientation is an axial observation, and that some conversion is needed for applying the directional-linear independence test. In the two-dimensional case, the orientations can be encoded by an angular variable $\Theta \in [0, \pi)$, with period π, so 2Θ is a circular variable. Then, with this codification, the angles 0, $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$ represent the E/W, NE/SW, N/S and NW/SE orientations, respectively. In the three-dimensional space, the orientation is coded by a pair of angles (Θ, Φ) using spherical coordinates, where $\Theta \in [0, \pi)$ plays the same role as the previous setting and $\Phi \in [0, \frac{\pi}{2}]$ measures the inclination ($\Phi = \frac{\pi}{2}$ for flat slope and $\Phi = 0$ for vertical; only positive angles are considered as the slope of a certain angle ω equals the slope of $-\omega$). Therefore, points with spherical coordinates $(2\Theta, \Phi)$, which lie on the upper semisphere, can be regarded as a realization of a spherical variable.

5.4.2 Results

The null hypothesis of independence between wildfire orientation and its burnt area (in log scale) is rejected, either using orientations in $\mathbb{R}^2$ or in $\mathbb{R}^3$, with a common p-value 0.000. The test is carried out using the bandwidth selector BLCV (considered from now on) and all the 26870 observations for years 1985–2005, ignoring stratification by watershed, and with $B = 1000$ permutations. The p-values for the null hypothesis of independence between the orientation of a watershed and the total burnt area of fires within the region are 0.008 and 0.000 for orientations in $\mathbb{R}^2$ and in $\mathbb{R}^3$, respectively. Therefore, the null hypothesis is emphatically rejected.

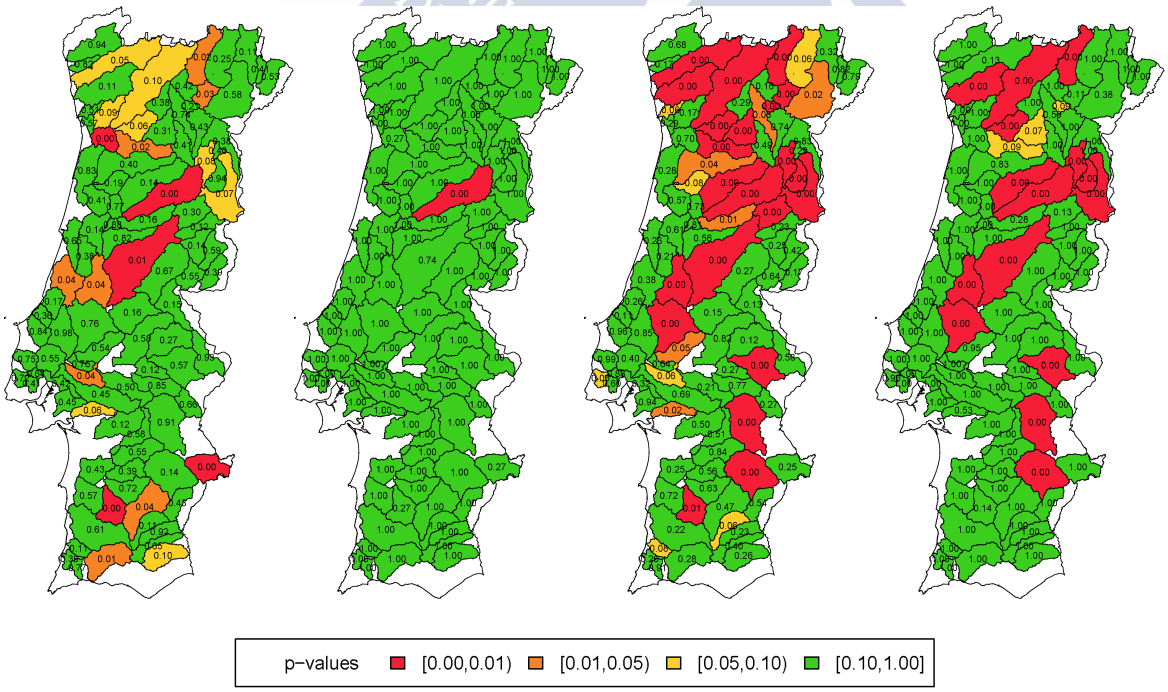


Figure 5.3: p-values from the independence test for the first principal component PC1 of the fire perimeter and the burnt area (on a log scale), by watersheds. From left to right, the first and second maps represent the circular-linear p-values (PC1 in $\mathbb{R}^2$) and their corrected versions using the FDR, respectively. The third and fourth maps represent the spherical-linear situation (PC1 in $\mathbb{R}^3$), with uncorrected and corrected p-values by FDR, respectively.

After identifying the presence of dependence between wildfire orientation and size, it is possible to carry out a watershed-based spatial analysis by applying the test to each watershed, in order to detect if the presence of dependence is homogeneous, or if it is only related to some particular areas. Figure 5.3 represents maps of p-values of the test applied to the observations of each watershed, using PC1 in $\mathbb{R}^2$ and in $\mathbb{R}^3$ (from left to right, first and third plots of Figure 5.3, respectively). The maps reveal the presence of 13 and 27 watersheds where the null hypothesis of independence is rejected with significance level $\alpha = 0.05$, for the circular-linear and the spherical-linear cases, respectively. This shows that the presence of dependence between fire orientation and size is not homogeneous and it is located in specific watersheds (see Figure 5.4). It is also interesting to note that the inclusion of the altitude coordinate in the computation of the PC1 leads to a richer detection of dependence between the wildfire orientation and size at the watershed level. This is due to the negative relation between the fire slope and size (see Figure 5.4), as large fires tend to have a flatter PC1 in $\mathbb{R}^3$ because they occur over highly variable terrain. Finally, the resulting p-values from the watershed analysis can also be adjusted using the False Discovery Rate (FDR) procedure of Benjamini and Yekutieli (2001) (from left to right, second and fourth plots of Figure 5.3). It is also possible to combine the p-values of the unadjusted maps with the FDR to test for independence between the wildfire orientation and the log-burnt area. The resulting p-values are 0.000 for the circular-linear and spherical-linear cases.

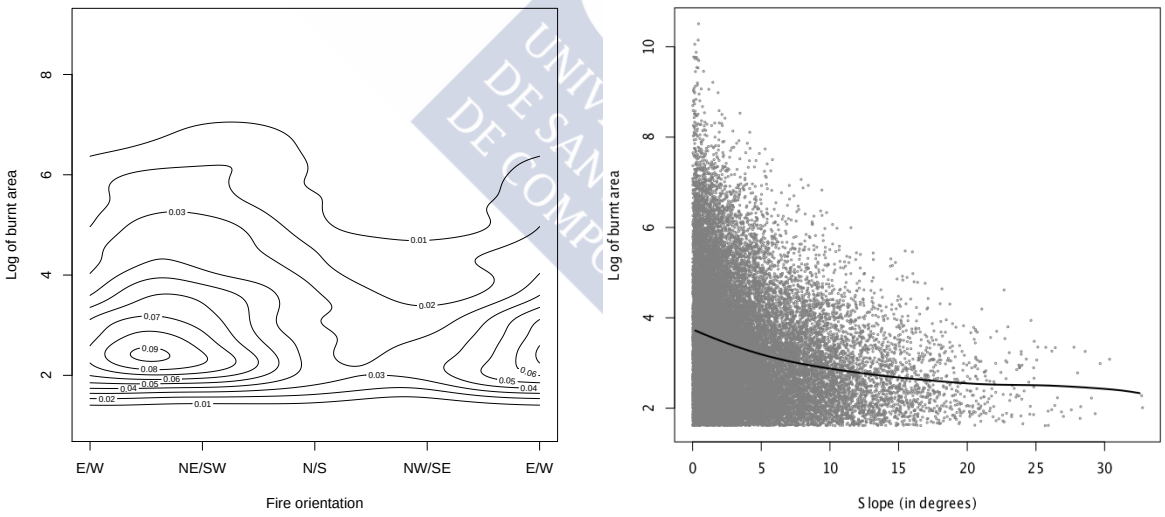


Figure 5.4: Left: density contour plot for fires in watershed number 31, the watershed in the second plot on the left of Figure 5.3 with p-value = 0.000. The number of fires in the watershed is $n = 1543$. The contour plot shows that the size of the area burnt is related with the orientation of the fires in the watershed. Right: scatter plot of the fires slope and the burnt area for the whole dataset, with a nonparametric kernel regression curve showing the negative correlation between fire slope and size.

5.5 Discussion

A nonparametric test for assessing independence between a directional and a linear component has been proposed, and its finite sample performance has been investigated in a simulation study. Simulation results support a satisfactory behavior of the permutation test implemented

with LCV and BLCV bandwidths, in most cases outperforming the available circular-linear testing proposals, and being competitive in other cases. The proposed BLCV bandwidths presents better results in terms of empirical size, although further study is required in bandwidth selection. In addition, when the null hypothesis of independence is rejected, the kernel density estimate can be used to explore the form of dependence, at least for the circular-linear and spherical-linear cases.

The application of the test to the entire wildfire orientation and size dataset makes possible the detection of dependence between these two variables, for both two-dimensional or three-dimensional orientation. The same conclusion holds for watershed orientation and total area burnt. A detailed study of each watershed allows for a more specific insight into the problem. The evidence of independence between fire size and fire orientation in some watersheds suggests that an event-based analysis (such as the work of Barros et al. (2012)) should yield results similar to those that would be expected from an area-based analysis. On the other hand, detection of dependence between fire size and orientation in watersheds with uniform orientation (Barros et al., 2012) highlights cases where there may be a mixture of orientations. In such cases, an analysis taking fire size into account might find evidence of preferential orientation in fire perimeters. In watersheds where fire events show preferential orientation (non-uniform distribution) and there is dependence between size and orientation, fire orientation distributions are structured in relation to fire size, especially considering the typically asymmetric nature of fire size distributions, dominated by a small number of very large events (Strauss et al., 1989). In these cases, an area-weighted analysis of fire perimeter orientation might lead to different results than those found by Barros et al. (2012). When altitude is included in calculation of the PC1 in $\mathbb{R}^3$ it highlights the negative relation between fire slope and size, which is mostly due to the fact that larger fires present flatter PC1. Slope has a skewed distribution, with low mean value and a relatively long right tail. Thus, while small fires usually occur on high slopes, large fires on consistently steep areas are unlikely.

Finally, it can be argued that the data are probably not independent and identically distributed over space and time. Unfortunately, given the data gathering procedure (detailed at the beginning of Section 5.4) dependence patterns cannot be clearly identified. Accounting for temporal or spatial dependence directly in the directional-linear kernel estimator and in the testing procedure is an open problem.

Acknowledgments

The authors acknowledge the support of Project MTM2008–03010, from the Spanish Ministry of Science and Innovation, Project 10MDS207015PR from Dirección Xeral de I+D, Xunta de Galicia and IAP network StUDyS, from Belgian Science Policy. Work of E. García-Portugués has been supported by FPU grant AP2010–0957 from the Spanish Ministry of Education and work of A. M. G. Barros, by Ph.D. Grant SFRH/BD/40398/2007 from the Fundação para a Ciência e Tecnologia. J. M. C. Pereira participated in this research under the framework of research projects "Forest fire under climate, social and economic changes in Europe, the Mediterranean and other fire-affected areas of the world (FUME)", EC FP7 Grant Agreement No. 243888 and "Fire-Land-Atmosphere Inter-Relationships: understanding processes to predict wildfire regimes in Portugal" (FLAIR), PTDC/AAC/AMB/104702/2008. Authors gratefully acknowledge an anonymous referee for the suggestion of employing permutations for the test calibration and

the careful revision of the paper. Authors also acknowledge the suggestions raised by another referee.

5.A Proof of Lemma 5.1

Proof. The closed expression (just involving matrix computations) for T_n is obtained by splitting the calculus into three addends:

$$\begin{aligned}
T_n &= \int_{\Omega_q\times\mathbb{R}} \left(\hat{f}_{(\mathbf{X},Z);h,g}(\mathbf{x},z) - \hat{f}_{\mathbf{X};h}(\mathbf{x})\hat{f}_{Z;g}(z)\right)^2 dz\,\omega_q(d\mathbf{x}) \\
&= \int_{\Omega_q\times\mathbb{R}} \left(\frac{c_{h,q}(L)}{ng}\sum_{i=1}^n L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right) K\left(\frac{z-Z_i}{g}\right) - \hat{f}_{\mathbf{X};h}(\mathbf{x})\hat{f}_{Z;g}(z)\right)^2 dz\,\omega_q(d\mathbf{x}) \\
&= \sum_{i=1}^n\sum_{j=1}^n \int_{\Omega_q\times\mathbb{R}} \frac{c_{h,q}(L)^2}{n^2g^2} L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right) K\left(\frac{z-Z_i}{g}\right) L\left(\frac{1-\mathbf{x}^T\mathbf{X}_j}{h^2}\right) K\left(\frac{z-Z_j}{g}\right) dz\,\omega_q(d\mathbf{x}) \\
&\quad - 2\sum_{i=1}^n \int_{\Omega_q\times\mathbb{R}} \frac{c_{h,q}(L)}{ng} L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right) K\left(\frac{z-Z_i}{g}\right) \hat{f}_{\mathbf{X};h}(\mathbf{x})\hat{f}_{Z;g}(z)\,dz\,\omega_q(d\mathbf{x}) \\
&\quad + \int_{\Omega_q\times\mathbb{R}} \hat{f}_{\mathbf{X};h}(\mathbf{x})^2\hat{f}_{Z;g}(z)^2\,dz\,\omega_q(d\mathbf{x}) \\
&= (6) - (7) + (8).
\end{aligned}$$

The first addend is

$$\begin{aligned}
(6) &= \frac{c_{h,q}(L)^2}{n^2g^2}\sum_{i=1}^n\sum_{j=1}^n \int_{\Omega_q\times\mathbb{R}} L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right) K\left(\frac{z-Z_i}{g}\right) L\left(\frac{1-\mathbf{x}^T\mathbf{X}_j}{h^2}\right) K\left(\frac{z-Z_j}{g}\right) dz\,\omega_q(d\mathbf{x}) \\
&= \frac{c_{h,q}(L)^2}{n^2g^2}\sum_{i=1}^n\sum_{j=1}^n \int_{\Omega_q} e^{-2/h^2} e^{\mathbf{x}^T(\mathbf{X}_i+\mathbf{X}_j)/h^2}\,\omega_q(d\mathbf{x}) \times \int_{\mathbb{R}} K\left(\frac{z-Z_i}{g}\right) K\left(\frac{z-Z_j}{g}\right) dz \\
&= \frac{c_{h,q}(L)^2}{n^2} e^{-2/h^2}\sum_{i=1}^n\sum_{j=1}^n \frac{\phi_{\sqrt{2}g}(Z_i-Z_j)}{C_q\left(\left|\left|\mathbf{X}_i+\mathbf{X}_j\right|\right|/h^2\right)} \\
&= \frac{C_q\left(1/h^2\right)^2}{n^2}\sum_{i=1}^n\sum_{j=1}^n \frac{\phi_{\sqrt{2}g}(Z_i-Z_j)}{C_q\left(\left|\left|\mathbf{X}_i+\mathbf{X}_j\right|\right|/h^2\right)}.
\end{aligned}$$

For the second addend,

$$\begin{aligned}
(7) &= 2\sum_{i=1}^n \int_{\Omega_q\times\mathbb{R}} \frac{c_{h,q}(L)}{ng} L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right) K\left(\frac{z-Z_i}{g}\right) \hat{f}_{\mathbf{X};h}(\mathbf{x})\hat{f}_{Z;g}(z)\,dz\,\omega_q(d\mathbf{x}) \\
&= 2\frac{c_{h,q}(L)}{ng}\sum_{i=1}^n \int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right) \hat{f}_{\mathbf{X};h}(\mathbf{x})\,\omega_q(d\mathbf{x}) \times \int_{\mathbb{R}} K\left(\frac{z-Z_i}{g}\right) \hat{f}_{Z;g}(z)\,dz \\
&= 2\frac{c_{h,q}(L)}{ng}\sum_{i=1}^n \left\{\left[\int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right) \frac{c_{h,q}(L)}{n}\sum_{j=1}^n L\left(\frac{1-\mathbf{x}^T\mathbf{X}_j}{h^2}\right) \omega_q(d\mathbf{x})\right]\right. \\
&\quad \left.\times \left[\int_{\mathbb{R}} K\left(\frac{z-Z_i}{g}\right)\frac{1}{ng}\sum_{k=1}^n K\left(\frac{z-Z_k}{g}\right) dz\right]\right\}
\end{aligned}$$

$$= \frac{2}{n^3}\sum_{i=1}^{n}\left\{\left[\int_{\Omega_q} c_{h,q}(L)^2 e^{-2/h^2} e^{\mathbf{x}^T(\mathbf{X}_i+\mathbf{X}_j)/h^2}\,\omega_q(d\mathbf{x})\right]\times\left[\sum_{k=1}^{n}\phi_{\sqrt{2}g}(Z_i-Z_k)\right]\right\}$$
$$= \frac{2}{n^3}\sum_{i=1}^{n}\left\{\left[\sum_{j=1}^{n}\frac{C_q\left(1/h^2\right)^2}{C_q\left(||\mathbf{X}_i+\mathbf{X}_j||/h^2\right)}\right]\times\left[\sum_{k=1}^{n}\phi_{\sqrt{2}g}(Z_i-Z_k)\right]\right\}.$$

Finally, the third addend is obtained as

$$\begin{aligned}(8) &= \int_{\Omega_q\times\mathbb{R}} \hat{f}_{\mathbf{X};h}(\mathbf{x})^2\hat{f}_{Z;g}(z)^2\,dz\,\omega_q(d\mathbf{x})\\ &= \int_{\Omega_q}\hat{f}_{\mathbf{X};h}(\mathbf{x})^2\,\omega_q(d\mathbf{x})\times\int_{\mathbb{R}}\hat{f}_{Z;g}(z)^2\,dz\\ &= \left[\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{c_{h,q}(L)^2}{n^2}\int_{\Omega_q}L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right)L\left(\frac{1-\mathbf{x}^T\mathbf{X}_j}{h^2}\right)\omega_q(d\mathbf{x})\right]\\ &\quad\times\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{n^2g^2}\int_{\mathbb{R}}K\left(\frac{z-Z_i}{g}\right)K\left(\frac{z-Z_j}{g}\right)dz\right]\\ &= \left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{C_q\left(1/h^2\right)^2}{C_q\left(||\mathbf{X}_i+\mathbf{X}_j||/h^2\right)}\right]\times\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{\sqrt{2}g}(Z_i-Z_j)\right].\end{aligned}$$

From the previous results and after applying some matrix algebra, it turns out that

$$T_n = \mathbf{1}_n\left(\frac{1}{n^2}\mathbf{\Psi}(h)\circ\mathbf{\Omega}(g) - \frac{2}{n^3}\mathbf{\Psi}(h)\mathbf{\Omega}(g) + \frac{1}{n^4}\mathbf{\Psi}(h)\mathbf{1}_n\mathbf{1}_n^T\mathbf{\Omega}(g)\right)\mathbf{1}_n^T,$$

where:

$$\mathbf{\Psi}(h) = \left(\frac{C_q\left(1/h^2\right)^2}{C_q\left(||\mathbf{X}_i+\mathbf{X}_j||/h^2\right)}\right)_{ij},\quad \mathbf{\Omega}(g) = \left(\phi_{\sqrt{2}g}\left(Z_i-Z_j\right)\right)_{ij}.$$

□

References

Ahmad, I. A. and Li, Q. (1997). Testing independence by nonparametric kernel method. *Statist. Probab. Lett.*, 34(2):201–210.

Bai, Z. D., Rao, C. R., and Zhao, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.*, 27(1):24–39.

Barros, A. M. G., Pereira, J. M. C., and Lund, U. J. (2012). Identifying geographical patterns of wildfire orientation: a watershed-based analysis. *Forest. Ecol. Manag.*, 264:98–107.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188.

Cao, R. (1993). Bootstrapping the mean integrated squared error. *J. Multivariate Anal.*, 45(1):137–160.

Cao, R., Cuevas, A., and Gonzalez Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Comput. Statist. Data Anal.*, 17(2):153–176.

Chiu, S.-T. (1996). A comparative review of bandwidth selection for kernel density estimation. *Statist. Sinica*, 6(1):129–145.

ESRI (2009). *ArcMap 9.2l*. Environmental Systems Resource Institute, Redlands.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Douglas, A. (2007). The shuttle radar topography mission. *Rev. Geophys.*, 45(2).

Fernández-Durán, J. J. (2007). Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics*, 63(2):579–585.

Finney, M. A. (2001). Design of regular landscape fuel treatment patterns for modifying fire growth and behavior. *Forest Sci.*, 47(2):219–228.

Fisher, N. I. and Lee, A. J. (1981). Nonparametric measures of angular-linear association. *Biometrika*, 68(3):629–636.

García-Portugués, E. (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron. J. Stat.*, 7:1655–1685.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013a). Exploring wind direction and SO_2 concentration by circular-linear density estimation. *Stoch. Environ. Res. Risk Assess.*, 27(5):1055–1067.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013b). Kernel density estimation for directional-linear data. *J. Multivariate Anal.*, 121:152–175.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2014). Central limit theorems for directional and linear data with applications. *Statist. Sinica*, to appear.

Hall, P., Watson, G. S., and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762.

Johnson, R. A. and Wehrly, T. (1977). Measures and models for angular correlation and angular-linear correlation. *J. Roy. Statist. Soc. Ser. B*, 39(2):222–229.

Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *J. Amer. Statist. Assoc.*, 73(363):602–606.

Jones, C., Marron, J. S., and Sheather, S. J. (1996). Progress in data-based bandwidth selection for kernel density estimation. *Computation. Stat.*, (11):337–381.

Liddell, I. G. and Ord, J. K. (1978). Linear-circular correlation coefficients: some further results. *Biometrika*, 65(2):448–450.

Lloret, F., Calvo, E., Pons, X., and Diaz-Delgado, R. (2002). Wildfires and landscape patterns in the Eastern Iberian Peninsula. *Landscape Ecol.*, 17(8):745–759.

Luo, D. (1998). *Pattern recognition and image processing*. Horwood Series in Engineering Science. Horwood, Chichester.

Mardia, K. V. (1976). Linear-circular correlation coefficients and rhythmometry. *Biometrika*, 63(2):403–405.

Moreira, F., Rego, F. C., and Ferreira, P. G. (2001). Temporal (1985-1995) pattern of change in a cultural landscape of northwestern Portugal: implications for fire occurrence. *Landscape Ecol.*, 16:557–567.

Moreira, F., Viedma, O., Arianoutsou, M., Curt, T., Koutsias, N., Rigolot, E., Barbati, A., Corona, P., Vaz, P., Xanthopoulos, G., Mouillot, F., and Bilgili, E. (2011). Landscape-wildfire interactions in southern Europe: implications for landscape management. *J. Environ. Manag.*, 92(10):2389–2402.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33(3):1065–1076.

Pereira, J. and Santos, T. (2003). Fire risk and burned area mapping in Portugal. Technical report, Direção-Geral das Florestas, Lisboa.

Pereira, M. G., Trigo, R. M., da Camara, C. C., Pereira, J. M. C., and Leite, S. M. (2005). Synoptic patterns associated with large summer forest fires in portugal. *Agr. Forest Meteorol.*, 129(1):11–25.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27(3):832–837.

Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.*, 3(1):1–14.

Rosenblatt, M. and Wahlen, B. E. (1992). A nonparametric measure of independence under a hypothesis of independent components. *Statist. Probab. Lett.*, 15(3):245–252.

Schmidt, D. A., Taylor, A. H., and Skinner, C. N. (2008). The influence of fuels treatment and landscape arrangement on simulated fire behavior, Southern Cascade range, California. *Forest. Ecol. Manag.*, 255(8–9):3170–3184.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Strauss, D., Bednar, L., and Mees, R. (1989). Do one percent of forest fires cause ninety-nine percent of the damage? *Forest Sci.*, 35(2):319–328.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Watson, G. S. (1983). *Statistics on spheres*, volume 6 of *University of Arkansas Lecture Notes in the Mathematical Sciences*. John Wiley & Sons, New York.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 6

Central limit theorems for directional and linear random variables with applications

Abstract

A central limit theorem for the integrated squared error of the directional-linear kernel density estimator is established. The result enables the construction and analysis of two testing procedures based on the squared loss: a nonparametric independence test for directional and linear random variables and a goodness-of-fit test for parametric families of directional-linear densities. Limit distributions for both test statistics, and a consistent bootstrap strategy for the goodness-of-fit test, are developed for the directional-linear case and adapted to the directional-directional setting. Finite sample performance for the goodness-of-fit test is illustrated in a simulation study. This test is also applied to real datasets from biology and environmental sciences.

Reference

García-Portugués, E., Crujeiras, R. M. and González-Manteiga, W. (2014). Central limit theorems for directional and linear random variables with applications. *Statist. Sinica*, to appear.

Contents

6.1 Introduction 130
6.2 Background 131
6.3 Central limit theorem for the integrated squared error 133
6.3.1 Main result 133
6.3.2 Extensions of Theorem 6.1 134
6.4 Testing independence with directional random variables 135
6.5 Goodness-of-fit test with directional random variables 136
6.5.1 Testing a simple null hypothesis 136
6.5.2 Composite null hypothesis 137
6.5.3 Calibration in practise 138
6.5.4 Extensions to directional-directional models 138

6.6 Simulation study 139
6.7 Real data application 142
6.A Sketches of the main proofs 143
6.A.1 CLT for the integrated squared error 143
6.A.2 Testing independence with directional data 145
6.A.3 Goodness-of-fit test for models with directional data 146
References 147

6.1 Introduction

Statistical inference on random variables comprises estimation and testing procedures that allow to characterize the underlying distribution, regardless the variables nature and/or dimension. Specifically, density estimation stands out as one of the fundamental problems in statistical inference, where parametric and nonparametric approaches have been explored. Within the nonparametric alternatives, kernel density estimation (see Silverman (1986), Scott (1992) or Wand and Jones (1995), as comprehensive references for scalar random variables) provides a simple and intuitive way to explore and do inference on random variables. Among other contexts, kernel density estimation has been also adapted to directional data (see Mardia and Jupp (2000)), that is, data in the q-dimensional sphere. This kind of data arises, for example, in meteorology, when measuring wind direction, or in proteomics, when studying the angles in protein structure (circular data, $q = 1$, see Fernández-Durán (2007)); in astronomy, stars positions in the celestial sphere ($q = 2$, see García-Portugués (2013)) are also directional data, and even in text mining (large q, see Chapter 6 in Srivastava and Sahami (2009)) directional data examples can be found. Some early works on kernel density estimation with directional data are the papers by Hall et al. (1987) and Bai et al. (1988), who introduced kernel density estimators and their properties (bias, variance and uniformly strong consistency, among others). The estimation of the density derivatives was studied by Klemelä (2000), and Zhao and Wu (2001) stated a Central Limit Theorem (CLT) for the Integrated Squared Error (ISE) of the directional kernel density estimator. Some more recent works deal with the bandwidth selection problem, such as Taylor (2008) and Oliveira et al. (2012), devoted to circular data and García-Portugués (2013), for a general dimension. In some contexts, joint density models for directional and linear random variables are useful (e.g. for describing wind direction and SO_2 concentration (García-Portugués et al., 2013a)). In this setting, a kernel density estimator for directional-linear data was proposed and analysed by García-Portugués et al. (2013b).

Regardless of estimation purposes, kernel density estimators have been extensively used for the development of goodness-of-fit tests (see González-Manteiga and Crujeiras (2013) for a review) and independence tests. For example, Bickel and Rosenblatt (1973) and Fan (1994), provided goodness-of-fit tests for parametric densities for real random variables. Similarly, in the directional setting, Boente et al. (2014) presented a goodness-of-fit test for parametric directional densities. For assessing independence between two linear random variables, Rosenblatt (1975) proposed a test statistic based on the squared difference between the joint kernel density estimator and the product of the marginal ones (see also Rosenblatt and Wahlen (1992)). This idea was adapted to the directional-linear setting by García-Portugués et al. (2014), who derived a permutation independence test and compared its performance in practise with the testing pro-

posals given by Mardia (1976), Johnson and Wehrly (1978) and Fisher and Lee (1981) in this context.

The main device for the goodness-of-fit and independence tests presented above is the CLT for the ISE of the kernel density estimator. Hence, the main aim of this work is to provide such a result for the directional-linear kernel estimator introduced by García-Portugués et al. (2013b) and use it to derive a goodness-of-fit test for parametric families of directional-linear densities and an independence test for directional and linear variables. The CLT is obtained by proving an extended version of Theorem 1 in Hall (1984). The goodness-of-fit test follows by taking the ISE between the joint kernel estimator and a smoothed parametric estimate of the unknown density as a test statistic. For the independence test, the test statistic introduced in García-Portugués et al. (2014) will be considered and its asymptotic properties will be studied. Jointly with the asymptotic distribution, a bootstrap resampling strategy to calibrate the goodness-of-fit test will be investigated. Finite sample performance of the goodness-of-fit test is checked through an extensive simulation study, and this methodology is applied to analyse real datasets from forestry and proteomics. In addition, the results obtained for the directional-linear case are also adapted to the directional-directional context.

This paper is organized as follows. Section 6.2 presents some background on kernel density estimation for directional and linear random variables. Section 6.3 includes the CLT for the ISE of the directional-linear estimator and its extension to the directional-directional setting. The independence test for directional and linear variables is presented in Section 6.4. The goodness-of-fit test for simple and composite null hypotheses, its bootstrap calibration and extensions are given in Section 6.5. The empirical performance of the goodness-of-fit test is illustrated with a simulation study in Section 6.6 and with applications to real datasets in Section 6.7. Appendix 6.A collects the outline of the main proofs. Technical lemmas and further details on simulations and data analysis are provided as supplementary material, as well as the extensions of the independence test.

6.2 Background

For the sake of simplicity, f will denote the target density along the paper, which may be linear, directional, directional-linear or directional-directional, depending on the context.

Let Z denote a linear random variable with support $\operatorname{supp}(Z) \subseteq \mathbb{R}$ and density f and let $Z_1, \ldots, Z_n$ be a random sample of Z. The linear kernel density estimator is defined as

$$\hat{f}_g(z) = \frac{1}{ng} \sum_{i=1}^{n} K\left(\frac{z - Z_i}{g}\right), \quad z \in \mathbb{R},$$

where K denotes the kernel function and $g > 0$ is the bandwidth parameter, which controls the smoothness of the estimator (see Silverman (1986), among others).

Let $\mathbf{X}$ denote a directional random variable with density f and support the q-dimensional sphere, denoted by $\Omega_q = \{\mathbf{x} \in \mathbb{R}^{q+1} : x_1^2 + \cdots + x_{q+1}^2 = 1\}$. The Lebesgue measure in Ω_q is denoted by ω_q and, therefore, a directional density satisfies $\int_{\Omega_q} f(\mathbf{x})\, \omega_q(d\mathbf{x}) = 1$. When there is no possible confusion, ω_q will also denote the surface area of Ω_q: $\omega_q = \omega_q(\Omega_q) = 2\pi^{\frac{q+1}{2}}/\Gamma(\frac{q+1}{2})$. The

directional kernel density estimator introduced by Hall et al. (1987) and Bai et al. (1988) for a directional density f, based on a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ in the q-sphere, is defined as

$$\hat{f}_h(\mathbf{x}) = \frac{c_{h,q}(L)}{n} \sum_{i=1}^{n} L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right), \quad \mathbf{x} \in \Omega_q,$$

where L is the directional kernel, $h > 0$ is the bandwidth parameter and the scalar product of two vectors, $\mathbf{x}$ and $\mathbf{y}$, is denoted by $\mathbf{x}^T\mathbf{y}$, where $\mathbf{x}^T$ is the transpose of the column vector $\mathbf{x}$. $c_{h,q}(L)$ is a normalizing constant depending on the kernel L, the bandwidth h and the dimension q. Specifically, Bai et al. (1988) indicate that the inverse of the normalizing constant is given by

$$c_{h,q}(L)^{-1} = \lambda_{h,q}(L)h^q \sim \lambda_q(L)h^q, \tag{6.1}$$

where $\lambda_{h,q}(L) = \omega_{q-1}\int_0^{2h^{-2}} L(r)r^{\frac{q}{2}-1}(2-rh^2)^{\frac{q}{2}-1}\,dr$ and $\lambda_q(L) = 2^{\frac{q}{2}-1}\omega_{q-1}\int_0^{\infty} L(r)r^{\frac{q}{2}-1}\,dr$. The notation $a_n \sim b_n$ means that $a_n = b_n(1+o(1))$, as $n \to \infty$.

A usual choice for the directional kernel is $L(r) = e^{-r}$, also known as the von Mises kernel due to its relation with the von Mises-Fisher density (see Watson (1983)), $\mathrm{vM}(\boldsymbol{\mu}, \kappa)$, given by

$$f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = C_q(\kappa)\exp\left\{\kappa\mathbf{x}^T\boldsymbol{\mu}\right\}, \quad C_q(\kappa) = \frac{\kappa^{\frac{q-1}{2}}}{(2\pi)^{\frac{q+1}{2}}\mathcal{I}_{\frac{q-1}{2}}(\kappa)},$$

being $\boldsymbol{\mu} \in \Omega_q$ the directional mean, $\kappa > 0$ the concentration parameter around the mean and $\mathcal{I}_\nu$ the modified Bessel function of order ν.

The kernel estimator for a directional-linear density f based on a random sample $(\mathbf{X}_1, Z_1), \ldots, (\mathbf{X}_n, Z_n)$, with $(\mathbf{X}_i, Z_i) \in \Omega_q \times \mathbb{R}$, $i = 1, \ldots, n$, was proposed by García-Portugués et al. (2013b):

$$\hat{f}_{h,g}(\mathbf{x}, z) = \frac{c_{h,q}(L)}{ng} \sum_{i=1}^{n} LK\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}, \frac{z-Z_i}{g}\right), \quad (\mathbf{x}, z) \in \Omega_q \times \mathbb{R}, \tag{6.2}$$

where LK is a directional-linear kernel, h and g are the bandwidths for the directional and the linear components, respectively, and $c_{h,q}(L)$ is the normalizing constant. For the sake of simplicity, the product kernel $LK(\cdot,\cdot) = L(\cdot) \times K(\cdot)$ is considered. Asymptotic properties of this estimator were studied by García-Portugués et al. (2013b). In order to quantify the error of the density estimator, the ISE, defined as

$$\mathrm{ISE}\left[\hat{f}_{h,g}\right] = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - f(\mathbf{x}, z)\right)^2 dz\,\omega_q(d\mathbf{x}),$$

can be used. In this expression, the integral is taken with respect to the product measure $\omega_q \times m_{\mathbb{R}}$, with $m_{\mathbb{R}}$ denoting the usual Lebesgue measure in $\mathbb{R}$.

Finally, it is possible to define a directional-directional kernel density estimator at $(\mathbf{x}, \mathbf{y}) \in \Omega_{q_1} \times \Omega_{q_2}$ from a random sample $(\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n)$, with $(\mathbf{X}_i, \mathbf{Y}_i) \in \Omega_{q_1} \times \Omega_{q_2}$, $i = 1, \ldots, n$, that comes from a directional-directional density f:

$$\hat{f}_{h_1,h_2}(\mathbf{x}, \mathbf{y}) = \frac{c_{h_1,q_1}(L_1)c_{h_2,q_2}(L_2)}{n} \sum_{i=1}^{n} L_1\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h_1^2}\right) \times L_2\left(\frac{1-\mathbf{y}^T\mathbf{Y}_i}{h_2^2}\right).$$

To fix the notation that will be used along the paper, note that $R(\varphi)$ denotes the integral of the squared function φ^2 along its domain. The following integrals will be also needed:

$$\mu_2(K) = \int_{\mathbb{R}} z^2 K(z)\, dz, \quad b_q(L) = \frac{\int_0^\infty L(r) r^{\frac{q}{2}}\, dr}{\int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr}.$$

Density derivatives of different orders will be denoted as follows:

$$\boldsymbol{\nabla} f(\mathbf{x}, z) = \left(\frac{\partial f(\mathbf{x}, z)}{\partial x_1}, \ldots, \frac{\partial f(\mathbf{x}, z)}{\partial x_{q+1}}, \frac{\partial f(\mathbf{x}, z)}{\partial z} \right)^T = \left(\boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}, z), \nabla_z f(\mathbf{x}, z) \right)^T,$$

$$\boldsymbol{\mathcal{H}} f(\mathbf{x}, z) = \left(\begin{array}{c|c} \left(\frac{\partial^2 f(\mathbf{x}, z)}{\partial x_i \partial x_j} \right) & \frac{\partial^2 f(\mathbf{x}, z)}{\partial \mathbf{x} \partial z} \\ \hline \frac{\partial^2 f(\mathbf{x}, z)}{\partial z \partial \mathbf{x}^T} & \frac{\partial^2 f(\mathbf{x}, z)}{\partial z^2} \end{array} \right) = \left(\begin{array}{c|c} \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z) & \boldsymbol{\mathcal{H}}_{\mathbf{x}, z} f(\mathbf{x}, z) \\ \hline \boldsymbol{\mathcal{H}}_{\mathbf{x}, z} f(\mathbf{x}, z)^T & \mathcal{H}_z f(\mathbf{x}, z) \end{array} \right).$$

6.3 Central limit theorem for the integrated squared error

The main result in this paper, which is the basis for the tests presented in the next sections, is the CLT for the ISE of the kernel density estimator (6.2).

6.3.1 Main result

The following conditions are required to derive the asymptotic distribution for the ISE:

A1. Extend f from $\Omega_q \times \mathbb{R}$ to $\mathbb{R}^{q+2} \setminus \{(\mathbf{0}, z) : z \in \mathbb{R}\}$, by defining $f(\mathbf{x}, z) \equiv f\left(\mathbf{x}/\left|\left|\mathbf{x}\right|\right|, z\right)$ for all $\mathbf{x} \neq \mathbf{0}$ and $z \in \mathbb{R}$. Assume that f and its first three derivatives are bounded and uniformly continuous with respect to the product Euclidean norm in $\Omega_q \times \mathbb{R}$, $||(\mathbf{x}, z)|| = \sqrt{||\mathbf{x}||^2 + |z|^2}$.

A2. Assume that $L : [0, \infty) \to [0, \infty)$ and $K : \mathbb{R} \to [0, \infty)$ are continuous and bounded. Assume that L is nonincreasing and such that $0 < \lambda_q(L)$, $\lambda_q(L^2) < \infty$, $\forall q \geq 1$ and that K is a linear density, symmetric around zero and with $\mu_2(K) < \infty$.

A3. Assume that $h = h_n$ and $g = g_n$ are sequences of positive numbers such that $h_n \to 0$, $g_n \to 0$ and $n h_n^q g_n \to \infty$ as $n \to \infty$.

The uniform continuity and boundedness up to the second derivatives of f is a common assumption that appears, among others, in Hall (1984) and Rosenblatt and Wahlen (1992), while the assumption on the third derivatives is needed for the uniformity convergence. The assumption of compact support for the directional kernel L, stated in Zhao and Wu (2001), is replaced by the nonincreasing requirement ($L(r) \leq L(s)$, for $r \geq s$) and the finiteness of $\lambda_q(L)$ and $\lambda_q(L^2)$. These two conditions are less restrictive and allow for the consideration of the von Mises kernel. The next theorem provides the limit distribution of the ISE for (6.2). Its proof is based on a generalization of Theorem 1 in Hall (1984), stated as Lemma A.1 in Appendix 6.A.

Theorem 6.1 (CLT for the directional-linear ISE)**.** *Denote the ISE of $\hat{f}_{h,g}$ by I_n. Then, under conditions **A1**–**A3**, it holds that*

i. $n^{\frac{1}{2}} \phi(h, g)^{-\frac{1}{2}} \left(I_n - \mathbb{E}\left[I_n\right] \right) \xrightarrow{d} \mathcal{N}(0, 1)$, *if* $n \phi(h, g) h^q g \to \infty$,

ii. $n (h^q g)^{\frac{1}{2}} \left(I_n - \mathbb{E}\left[I_n\right] \right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma^2\right)$, *if* $n \phi(h, g) h^q g \to 0$,

iii. $n(h^q g)^{\frac{1}{2}} \left(I_n - \mathbb{E}\left[I_n\right]\right) \xrightarrow{d} \mathcal{N}\left(0, \delta + 2\sigma^2\right)$, *if* $n\phi(h,g)h^q g \to \delta$,

where $0 < \delta < \infty$ *and*

$$\phi(h,g) = \frac{4b_q(L)^2}{q^2}\sigma^2_{\mathbf{X}} h^4 + \mu_2(K)^2\sigma^2_Z g^4 + \frac{4b_q(L)\mu_2(K)}{q}\sigma_{\mathbf{X},Z} h^2 g^2,$$

with $\sigma_{\mathbf{X},Z} = \mathbb{C}\mathrm{ov}\left[\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}}(f,\mathbf{X},Z)\right], \mathcal{H}_z f(\mathbf{X},Z)\right]$, $\sigma^2_{\mathbf{X}} = \mathbb{V}\mathrm{ar}\left[\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}}(f,\mathbf{X},Z)\right]\right]$ *and* $\sigma^2_Z = \mathbb{V}\mathrm{ar}\left[\mathcal{H}_z f(\mathbf{X},Z)\right]$. *The remaining constants are given by:*

$$\begin{aligned}
\sigma^2 =& R(f) \times \gamma_q \lambda_q(L)^{-4} \int_0^\infty r^{\frac{q}{2}-1} \left\{\int_0^\infty \rho^{\frac{q}{2}-1} L(\rho)\varphi_q(r,\rho)\, d\rho\right\}^2 dr \\
&\times \int_{\mathbb{R}} \left\{\int_{\mathbb{R}} K(u)K(u+v)\, du\right\}^2 dv,
\end{aligned}$$

$$\varphi_q(r,\rho) = \begin{cases} L\left(r+\rho-2(r\rho)^{\frac{1}{2}}\right) + L\left(r+\rho+2(r\rho)^{\frac{1}{2}}\right), & q = 1, \\ \int_{-1}^1 \left(1-\theta^2\right)^{\frac{q-3}{2}} L\left(r+\rho-2\theta(r\rho)^{\frac{1}{2}}\right) d\theta, & q \geq 2, \end{cases}$$

$$\gamma_q = \begin{cases} 2^{-\frac{1}{2}}, & q = 1, \\ \omega_{q-1}\omega^2_{q-2} 2^{\frac{3q}{2}-3}, & q \geq 2. \end{cases}$$

The same limit distributions hold in i–iii if $\mathbb{E}\left[I_n\right]$ *is replaced by*

$$\int_{\Omega_q \times \mathbb{R}} \left(\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x},z)\right] - f(\mathbf{x},z)\right)^2 dz\, \omega_q(d\mathbf{x}) + \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g}.$$

Bearing in mind the CLT result in Hall (1984) for the linear case, a bandwidth-free rate of convergence should be expected in *iii*. Nevertheless, when $n\phi(h,g)h^q g \to \delta$, the analytical difficulty of joining the two rates of convergence of the dominant terms forces the normalizing rate to be $n(h^q g)^{\frac{1}{2}}$, although the sequence of bandwidths is restricted to satisfy the constraint $n\phi(h,g)h^q g \to \delta$. To clarify this point, the next corollary presents a special case with proportional bandwidth sequences where the rate of convergence can be analytically stated in a bandwidth-free form.

Corollary 6.1. *Under conditions* ***A1****–****A3****, and assuming* $g_n = \beta h_n$ *for a fixed* $\beta > 0$ *and* $0 < \delta < \infty$,

i. $n^{\frac{1}{2}} h^{-2} \left(I_n - \mathbb{E}\left[I_n\right]\right) \xrightarrow{d} \mathcal{N}(0, \phi(1,\beta))$, *if* $nh^{q+5} \to \infty$,

ii. $nh^{\frac{q+1}{2}} \left(I_n - \mathbb{E}\left[I_n\right]\right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma^2\right)$, *if* $nh^{q+5} \to 0$,

iii. $n^{\frac{q+9}{2(q+5)}} \left(I_n - \mathbb{E}\left[I_n\right]\right) \xrightarrow{d} \mathcal{N}\left(0, \phi(1,\beta)\delta^{\frac{4}{q+5}} + 2\sigma^2\delta^{-\frac{q+1}{q+5}}\right)$, *if* $nh^{q+5} \to \delta$.

6.3.2 Extensions of Theorem 6.1

The previous results can be adapted to other contexts involving directional variables, such as directional-directional or directional-multivariate random vectors. Once the common structure and the effects of each component are determined, it is easy to reproduce the computations duplicating a certain component or modifying it. This will be used to derive the directional-directional versions of the most relevant results along the paper. By considering a single bandwidth for the

estimator defined in $\mathbb{R}^p$ (as in Hall (1984), for example), Theorem 6.1 can be easily adapted to account for the multivariate component.

Considering the directional-directional estimator $\hat{f}_{h_1,h_2}$, the corresponding analogues of conditions **A1**–**A3** are obtained (extending f from $\Omega_{q_1} \times \Omega_{q_2}$ to $\{(\mathbf{x},\mathbf{y}) \in \mathbb{R}^{q_1+q_2+2} : \mathbf{x} \neq \mathbf{0}, \mathbf{y} \neq \mathbf{0}\}$ and assuming $nh_{1,n}^{q_1}h_{2,n}^{q_2} \to \infty$). Then, it is possible to derive a directional-directional version of Theorem 6.1.

Corollary 6.2 (CLT for the directional-directional ISE)**.** *Denote the ISE of* $\hat{f}_{h_1,h_2}$ *by* $I_n = \int_{\Omega_{q_1}\times\Omega_{q_2}} (\hat{f}_{h_1,h_2}(\mathbf{x},\mathbf{y}) - f(\mathbf{x},\mathbf{y}))^2 \, \omega_{q_2}(\mathbf{y}) \, \omega_{q_1}(\mathbf{x})$. *Then, under the directional-directional analogues of conditions* ***A1***–***A3****,*

i. $n^{\frac{1}{2}}\phi(h_1,h_2)^{-\frac{1}{2}} \left(I_n - \mathbb{E}\left[I_n\right]\right) \xrightarrow{d} Z$, *if* $n\phi(h_1,h_2)h_1^{q_1}h_2^{q_2} \to \infty$,

ii. $n(h_1^{q_1}h_2^{q_2})^{\frac{1}{2}} \left(I_n - \mathbb{E}\left[I_n\right]\right) \xrightarrow{d} 2^{\frac{1}{2}}\sigma Z$, *if* $n\phi(h_1,h_2)h_1^{q_1}h_2^{q_2} \to 0$,

iii. $n(h_1^{q_1}h_2^{q_2})^{\frac{1}{2}} \left(I_n - \mathbb{E}\left[I_n\right]\right) \xrightarrow{d} \left(\delta + 2\sigma^2\right)^{\frac{1}{2}} Z$, *if* $n\phi(h_1,h_2)h_1^{q_1}h_2^{q_2} \to \delta$,

where $0 < \delta < \infty$ *and*

$$\phi(h_1,h_2) = \frac{4b_{q_1}(L_1)^2}{q_1^2}\sigma_{\mathbf{X}}^2 h_1^4 + \frac{4b_{q_2}(L_2)^2}{q_2^2}\sigma_{\mathbf{Y}}^2 h_2^4 + \frac{8b_{q_1}(L_1)b_{q_2}(L_2)}{q_1 q_2}\sigma_{\mathbf{X},\mathbf{Y}} h_1^2 h_2^2.$$

$$\sigma^2 = R(f) \times \gamma_{q_1}\lambda_{q_1}(L_1)^{-4} \int_0^\infty r^{\frac{q_1}{2}-1} \left\{\int_0^\infty \rho^{\frac{q_1}{2}-1} L_1(\rho)\varphi_{q_1}(r,\rho)\, d\rho\right\}^2 dr$$
$$\times \gamma_{q_2}\lambda_{q_2}(L_2)^{-4} \int_0^\infty r^{\frac{q_2}{2}-1} \left\{\int_0^\infty \rho^{\frac{q_2}{2}-1} L_2(\rho)\varphi_{q_2}(r,\rho)\, d\rho\right\}^2 dr,$$

with $\sigma_{\mathbf{X},\mathbf{Y}} = \mathbb{C}\mathrm{ov}\left[\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}}(f,\mathbf{X},\mathbf{Y})\right], \mathrm{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{y}}(f,\mathbf{X},\mathbf{Y})\right]\right]$, $\sigma_{\mathbf{X}}^2 = \mathbb{V}\mathrm{ar}\left[\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}}(f,\mathbf{X},\mathbf{Y})\right]\right]$ *and* $\sigma_{\mathbf{Y}}^2 = \mathbb{V}\mathrm{ar}\left[\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{y}}(f,\mathbf{X},\mathbf{Y})\right]\right]$. *The same limit distributions hold in i–iii if* $\mathbb{E}\left[I_n\right]$ *is replaced by*

$$\int_{\Omega_{q_1}\times\Omega_{q_2}} \left(\mathbb{E}\left[\hat{f}_{h_1,h_2}(\mathbf{x},\mathbf{y})\right] - f(\mathbf{x},\mathbf{y})\right)^2 \omega_{q_1}(\mathbf{x})\, \omega_{q_2}(\mathbf{y}) + \frac{\lambda_{q_1}(L_1^2)\lambda_{q_2}(L_2^2)}{\lambda_{q_1}(L_1)^2\lambda_{q_2}(L_2)^2 n h_1^{q_1}h_2^{q_2}}.$$

6.4 Testing independence with directional random variables

Given a random sample $(\mathbf{X}_1, Z_1), \ldots, (\mathbf{X}_n, Z_n)$ from a directional-linear variable $(\mathbf{X}, Z)$, one may be interested in the assessment of independence between both components. If such a hypothesis is rejected, the joint kernel density estimator may give an idea of the dependence structure between them.

Let denote by $f_{(\mathbf{X},Z)}$ the directional-linear density of $(\mathbf{X}, Z)$ and by $f_{\mathbf{X}}$ and f_Z the directional and linear marginal densities, respectively. In this setting, the null hypothesis of independence is stated as $H_0 : f_{(\mathbf{X},Z)}(\mathbf{x},z) = f_{\mathbf{X}}(\mathbf{x})f_Z(z)$, $\forall(\mathbf{x},z) \in \Omega_q \times \mathbb{R}$ and the alternative hypothesis as $H_1 : f_{(\mathbf{X},Z)}(\mathbf{x},z) \neq f_{\mathbf{X}}(\mathbf{x})f_Z(z)$, for some $(\mathbf{x},z) \in \Omega_q \times \mathbb{R}$. A statistic to test H_0 can be constructed considering the squared distance between the nonparametric estimator of joint density, denoted in this setting by $\hat{f}_{(\mathbf{X},Z);h,g}$, and the product of the corresponding marginal kernel estimators, denoted by $\hat{f}_{\mathbf{X},h}$ and $\hat{f}_{Z,g}$. That is:

$$T_n = \int_{\Omega_q\times\mathbb{R}} \left(\hat{f}_{(\mathbf{X},Z);h,g}(\mathbf{x},z) - \hat{f}_{\mathbf{X};h}(\mathbf{x})\hat{f}_{Z;g}(z)\right)^2 dz\, \omega_q(d\mathbf{x}).$$

This kind of test was firstly introduced by Rosenblatt (1975) and Rosenblatt and Wahlen (1992) for real bivariate random variables, considering the same bandwidths for smoothing both components. The directional-linear context will require the following assumption on the degree of smoothness in each component.

A4. Assume that $h_n^q g_n^{-1} \to c$, with $0 < c < \infty$, as $n \to \infty$.

Theorem 6.2 (Directional-linear independence test)**.** *Under conditions **A1**–**A4** and under the null hypothesis of independence,*

$$n(h^q g)^{\frac{1}{2}} \left(T_n - A_n\right) \xrightarrow{d} \mathcal{N}(0, 2\sigma_I^2),$$

where

$$A_n = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g} - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(f_Z)}{nh^q} - \frac{R(K)R(f_{\mathbf{X}})}{ng},$$

and σ_I^2 is defined as σ^2 in Theorem 6.1 but with $R(f) = R(f_{\mathbf{X}})R(f_Z)$.

Since the leading term is the same as in Theorem 6.1 for $n\phi(h,g)h^q g \to 0$, the asymptotic variance is also the same. As in the CLT for the ISE, the effect of both components can be clearly disentangled in the asymptotic variance and in the bias term. The a priori complex contribution of the directional part in Theorems 6.1 and 6.4 is explained for a particular scenario in the supplementary material, together with some numerical experiments for the illustration of Theorem 6.2.

6.5 Goodness-of-fit test with directional random variables

Testing methods for a specific parametric directional-linear density (simple H_0) or for a parametric family (composite H_0) will be presented in this section.

6.5.1 Testing a simple null hypothesis

Given a random sample $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$ from an unknown directional-linear density f, the simple null hypothesis testing problem is stated as $H_0 : f = f_{\boldsymbol{\theta}_0}$, $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, where $f_{\boldsymbol{\theta}_0}$ is a certain parametric density with known parameter $\boldsymbol{\theta}_0$ belonging to the parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^p$, with $p \geq 1$. The alternative hypothesis will be taken as $H_1 : f(\mathbf{x}, z) \neq f_{\boldsymbol{\theta}_0}(\mathbf{x}, z)$, for some $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$ in a set of positive measure. The proposed test statistic is given by

$$R_n = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - LK_{h,g} f_{\boldsymbol{\theta}_0}(\mathbf{x}, z)\right)^2 dz\, \omega_q(d\mathbf{x}), \tag{6.3}$$

where $LK_{h,g} f_{\boldsymbol{\theta}_0}(\mathbf{x}, z)$ represents the expected value of $\hat{f}_{h,g}(\mathbf{x}, z)$ under H_0. In general, for a function f, this expected value is given by

$$LK_{h,g} f(\mathbf{x}, z) = \frac{c_{h,g}(L)}{g} \int_{\Omega_q \times \mathbb{R}} LK\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}, \frac{z - t}{g}\right) f(\mathbf{y}, t)\, dt\, \omega_q(d\mathbf{y}). \tag{6.4}$$

Smoothing the parametric density was considered by Fan (1994), in the linear setting, to avoid the bias effects in the integrand of the square error between the nonparametric estimator under the alternative and the parametric estimate under the null. A modification of the smoothing proposal by Fan (1994) was used by Boente et al. (2014) for the directional case.

Theorem 6.3. *Under conditions **A1**–**A3** and under the simple null hypothesis $H_0 : f = f_{\boldsymbol{\theta}_0}$, with $\boldsymbol{\theta}_0 \in \Theta$ known,*

$$n(h^q g)^{\frac{1}{2}} \left(R_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g} \right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma^2_{\boldsymbol{\theta}_0}\right),$$

where $\sigma^2_{\boldsymbol{\theta}_0}$ follows from replacing $f = f_{\boldsymbol{\theta}_0}$ in σ^2 from Theorem 6.1.

6.5.2 Composite null hypothesis

Consider the testing problem $H_0 : f \in \mathcal{F}_{\boldsymbol{\Theta}} = \{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, where $\mathcal{F}_{\boldsymbol{\Theta}}$ is a class of parametric densities indexed by the p-dimensional parameter $\boldsymbol{\theta}$, vs. $H_1 : f \notin \mathcal{F}_{\boldsymbol{\Theta}}$. Under H_0 (which is equivalent to $H_0 : f = f_{\boldsymbol{\theta}_0}$, $\boldsymbol{\theta}_0 \in \Theta$, where $\boldsymbol{\theta}_0$ is unknown), a parametric density estimator $f_{\hat{\boldsymbol{\theta}}}$ can be obtained by Maximum Likelihood (ML). The following conditions are required:

A5. The function $f_{\boldsymbol{\theta}}$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$ and the derivatives are bounded and uniformly continuous for $(\mathbf{x}, z)$.

A6. There exists $\boldsymbol{\theta}_1 \in \Theta$ such that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_1 = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$ and if $H_0 : f = f_{\boldsymbol{\theta}_0}$ holds for a certain $\boldsymbol{\theta}_0 \in \Theta$, then $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$.

A5 is a regularity assumption on the parametric density, whereas **A6** states that the estimation of the unknown parameter must be $\sqrt{n}$-consistent in order to ensure that the effects of parametric estimation can be neglected. Note that the $\sqrt{n}$-consistency is required both under H_0 (for Theorem 6.4) and H_1 (for Theorem 6.6), which is satisfied by the ML estimator. The test statistic is an adaptation of (6.3), but plugging-in the estimator of the unknown parameter $\boldsymbol{\theta}_0$ under H_0 in the test statistic expression:

$$R_n = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - LK_{h,g} f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) \right)^2 dz\, \omega_q(d\mathbf{x}). \tag{6.5}$$

Theorem 6.4 (Goodness-of-fit test for directional-linear densities)**.** *Under conditions **A1**–**A3**, **A5**–**A6** and under the composite null hypothesis $H_0 : f = f_{\boldsymbol{\theta}_0}$, with $\boldsymbol{\theta}_0 \in \Theta$ unknown,*

$$n(h^q g)^{\frac{1}{2}} \left(R_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g} \right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma^2_{\boldsymbol{\theta}_0}\right).$$

Families of local alternatives, also known as Pitman alternatives, are a common way to measure power for tests based on kernel smoothers (e.g. Fan (1994)). For the directional-linear case, these alternatives can be written as

$$H_{1P} : f(\mathbf{x}, z) = f_{\boldsymbol{\theta}_0}(\mathbf{x}, z) + \left(nh^{\frac{q}{2}} g^{\frac{1}{2}}\right)^{-\frac{1}{2}} \Delta(\mathbf{x}, z), \tag{6.6}$$

where $\Delta(\mathbf{x}, z) : \Omega_q \times \mathbb{R} \to \mathbb{R}$ is such that $\int_{\Omega_q \times \mathbb{R}} \Delta(\mathbf{x}, z)\, dz\, \omega_q(d\mathbf{x}) = 0$. A necessary condition to derive the limit distribution of R_n under H_{1P} is that the estimator $\hat{\boldsymbol{\theta}}$ is a $\sqrt{n}$-consistent estimator for $\boldsymbol{\theta}_0$:

A7. For the family of alternatives (6.6), it holds that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$.

Theorem 6.5 (Local power under Pitman alternatives)**.** *Under conditions **A1**–**A3**, **A5**–**A7** and under the alternative hypothesis (6.6),*

$$n(h^q g)^{\frac{1}{2}} \left(R_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g} \right) \xrightarrow{d} \mathcal{N}\left(R\left(\Delta\right), 2\sigma^2_{\boldsymbol{\theta}_0}\right).$$

6.5.3 Calibration in practise

In order to effectively calibrate the proposed test in practise, a parametric bootstrap procedure is investigated. The bootstrap statistic is defined as

$$R_n^* = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}^*(\mathbf{x}, z) - LK_{h,g} f_{\hat{\boldsymbol{\theta}}^*}(\mathbf{x}, z) \right)^2 dz\, \omega_q(d\mathbf{x}),$$

where the superscript * indicates that the estimators are computed from the bootstrap sample $\{(\mathbf{X}_i^*, Z_i^*)\}_{i=1}^n$ obtained from the density $f_{\hat{\boldsymbol{\theta}}}$, with $\hat{\boldsymbol{\theta}}$ computed from the original sample. The bootstrap procedure, considering the composite null hypothesis testing problem, is detailed in the following algorithm. Calibration for the simple null hypothesis test can be done replacing $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^*$ by $\boldsymbol{\theta}_0$.

Algorithm 6.1 (Testing procedure)**.** *Let $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$ be a random sample from f. To test $H_0 : f = f_{\boldsymbol{\theta}_0}$, with $\boldsymbol{\theta}_0 \in \Theta$ unknown, proceed as follows:*

i. Obtain $\hat{\boldsymbol{\theta}}$, a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}_0$.

ii. Compute $R_n = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - LK_{h,g} f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) \right)^2 dz\, \omega_q(d\mathbf{x})$.

iii. Bootstrap strategy. For $b = 1, \ldots, B$:

(a) Obtain a random sample $\{(\mathbf{X}_i^*, Z_i^*)\}_{i=1}^n$ from $f_{\hat{\boldsymbol{\theta}}}$.

(b) Compute $\hat{\boldsymbol{\theta}}^*$ as in step i, from the bootstrap sample in (a).

(c) Compute $R_n^{*b} = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}^*(\mathbf{x}, z) - LK_{h,g} f_{\hat{\boldsymbol{\theta}}^*}(\mathbf{x}, z) \right)^2 dz\, \omega_q(d\mathbf{x})$, where $\hat{f}_{h,g}^*$ is obtained from the bootstrap sample in (a).

iv. Approximate the p-value of the test as p-value $\approx \#\{R_n^{*b} \leq R_n\}/B$.

The consistency of this testing procedure is proved in the next result, using the bootstrap analogue of assumption **A6**.

A8. $\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}} = \mathcal{O}_{\mathbb{P}^*}(n^{-\frac{1}{2}})$, where $\mathbb{P}^*$ represents the probability of $(\mathbf{X}^*, Z^*)$ conditioned on the sample $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$.

Theorem 6.6 (Bootstrap consistency)**.** *Under conditions **A1**–**A3**, **A5**–**A6** and **A8**, for a certain $\boldsymbol{\theta}_1 \in \Theta$,*

$$n(h^q g)^{\frac{1}{2}} \left(R_n^* - \frac{\lambda_q(L^2)\lambda_q(L)^{-2} R(K)}{nh^q g} \right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma_{\boldsymbol{\theta}_1}^2\right) \text{ in probability.}$$

Then, the probability distribution function (pdf) of R_n^* conditioned on the sample converges in probability to a Gaussian pdf, regardless H_0 holds or not. The asymptotic distribution coincides with the one of R_n if H_0 holds ($\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$).

6.5.4 Extensions to directional-directional models

The directional-directional versions of the previous results follow under analogous assumptions (modifying **A5**, (6.4) and (6.6) accordingly). The directional-directional test statistic for the composite hypothesis testing problem is

$$R_n = \int_{\Omega_{q_1} \times \Omega_{q_2}} \left(\hat{f}_{h_1,h_2}(\mathbf{x}, \mathbf{y}) - L_1 L_{2,h_1,h_2} f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, \mathbf{y}) \right)^2 \omega_{q_2}(d\mathbf{y})\, \omega_{q_1}(d\mathbf{x}).$$

Corollary 6.3 (Goodness-of-fit test for directional-directional densities)**.** *Under the directional-directional analogues of conditions* ***A1–A3, A5–A6*** *and under the composite null hypothesis* $H_0 : f = f_{\boldsymbol{\theta}_0}$*, with* $\boldsymbol{\theta}_0 \in \Theta$ *unknown,*

$$n(h_1^{q_1} h_2^{q_2})^{\frac{1}{2}} \left(R_n - \frac{\lambda_{q_1}(L_1^2)\lambda_{q_1}(L_1)^{-2}\lambda_{q_2}(L_2^2)\lambda_{q_2}(L_2)^{-2}}{n h_1^{q_1} h_2^{q_2}} \right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma^2_{\boldsymbol{\theta}_0}\right).$$

6.6 Simulation study

The finite sample performance of the directional-linear and directional-directional goodness-of-fit tests is illustrated in this section for a variety of models, sample sizes and bandwidth choices. The study considers circular-linear and circular-circular scenarios, although these tests can be easily applied in higher dimensions, such as spherical-linear or spherical-circular, due to their general definition and resampling procedure. Details on simulated models and further results are included as supplementary material.

Circular-Linear (CL) and Circular-Circular (CC) parametric scenarios are considered. Figures 6.1 and 6.2 show the density contours in the cylinder (CL) and in the torus (CC) for the different models. The detailed description of each model is given in the supplementary material. Deviations from the composite null hypothesis $H_0 : f \in \mathcal{F}_\Theta$ are obtained by mixing the true density $f_{\boldsymbol{\theta}_0}$ with a density Δ such that the resulting density does not belong to $\mathcal{F}_\Theta$: $H_\delta : f = (1-\delta)f_{\boldsymbol{\theta}_0} + \delta\Delta$, $0 \leq \delta \leq 1$. The goodness-of-fit tests are applied using the bootstrap strategy, for the whole collection of models, sample sizes $n = 100, 500, 1000$ and deviations $\delta = 0, 0.10, 0.15$ ($\delta = 0$ for the null hypothesis). The number of bootstrap and Monte Carlo replicates is 1000.

In each case (model, sample size and deviation), the performance of the goodness-of-fit test is shown for a fixed pair of bandwidths, obtained from the median of 1000 simulated Likelihood Cross Validation (LCV) bandwidths:

$$\begin{aligned} (h, g)_{\text{LCV}} &= \arg\max_{h,g>0} \textstyle\sum_{i=1}^n \log f^{-i}_{h,g}(\mathbf{X}_i, Z_i), \\ (h_1, h_2)_{\text{LCV}} &= \arg\max_{h_1,h_2>0} \textstyle\sum_{i=1}^n \log f^{-i}_{h_1,h_2}(\mathbf{X}_i, \mathbf{Y}_i), \end{aligned} \tag{6.7}$$

where $f^{-i}_{\cdots}$ denotes the kernel estimator computed without the i-th datum. A deeper insight on the bandwidth effect is provided for some scenarios, where percentage of rejections are plotted for a grid of bandwidths (see Figure 6.3 for two cases, and supplementary material for extended results). The kernels considered are the von Mises and the normal ones.

Table 6.1 collects the results of the simulation study for each combination of model (CL or CC), deviation (δ) and sample size (n). When the null hypothesis holds, significance levels are correctly attained for $\alpha = 0.05$ (see supplementary material for $\alpha = 0.10, 0.01$), for all sample sizes, models and deviations. When the null hypothesis does not hold, the tests perform satisfactorily, having in both cases a quick detection of the alternative when only a 10% and a 15% of the data come from a density out of the parametric family. As expected, the rejection rates grow as the sample size and the deviation from the alternative do.

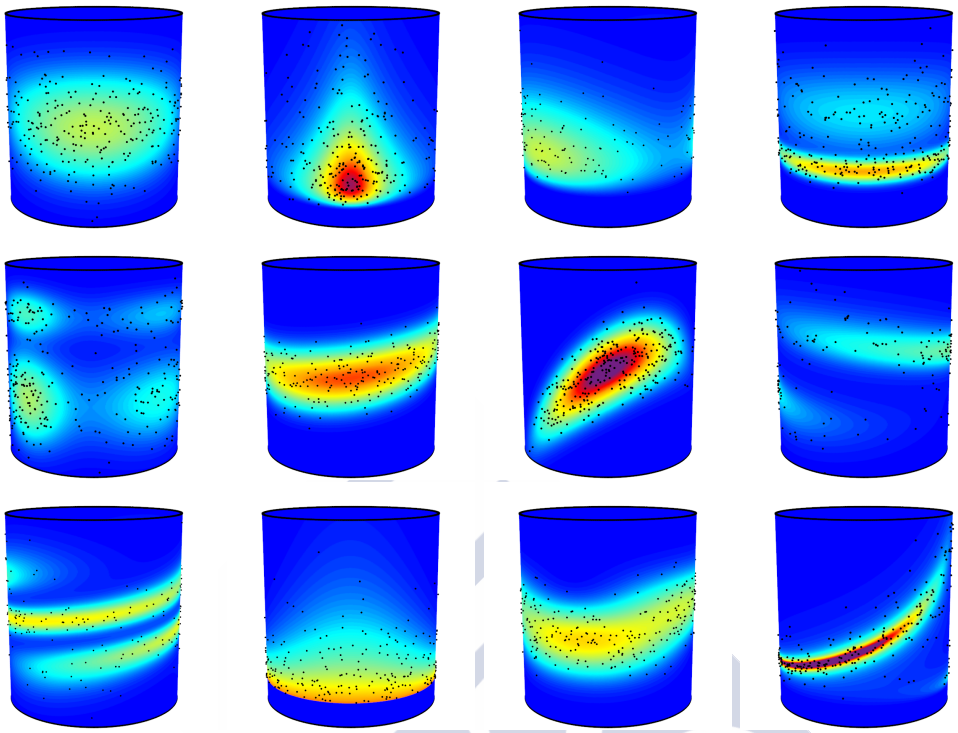


Figure 6.1: Density models for the simulation study in the circular-linear case. From left to right and up to down, models CL1 to CL12.

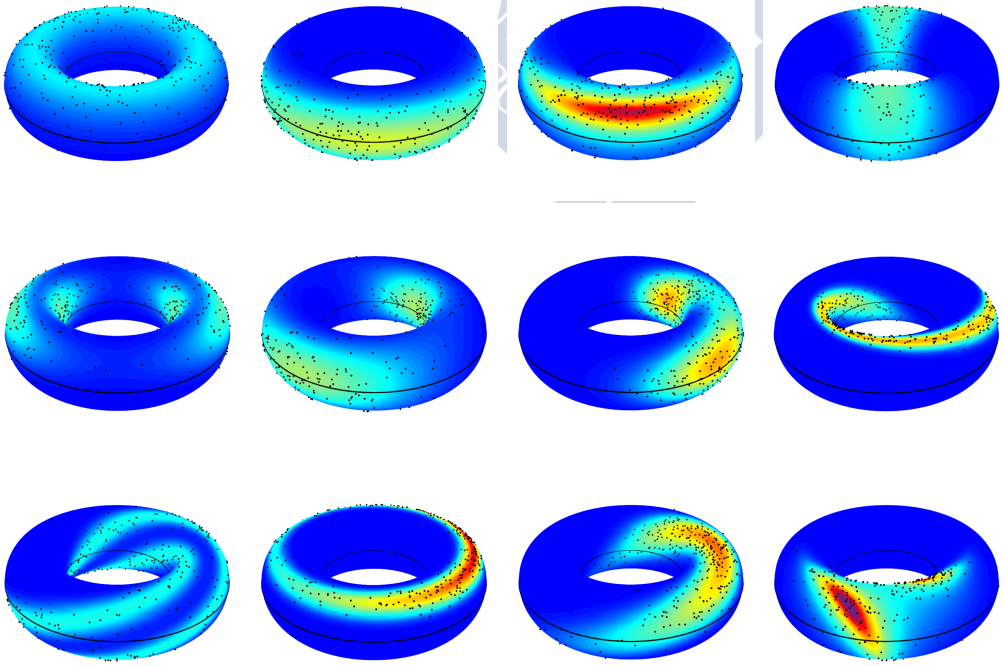


Figure 6.2: Density models for the simulation study in the circular-circular case. From left to right and up to down, models CC1 to CC12.

Model	Sample size n and deviation δ								
	$n = 100$			$n = 500$			$n = 1000$		
	δ=0	δ=0.10	δ=0.15	δ=0	δ=0.10	δ=0.15	δ=0	δ=0.10	δ=0.15
CL1	0.051	0.552	0.997	0.052	0.822	1.000	0.048	1.000	1.000
CL2	0.051	0.244	0.805	0.049	0.525	0.997	0.050	0.982	1.000
CL3	0.048	0.107	0.362	0.046	0.163	0.682	0.050	0.659	0.940
CL4	0.045	0.172	0.568	0.039	0.297	0.869	0.045	0.868	0.993
CL5	0.049	0.272	0.972	0.049	0.514	0.999	0.041	1.000	1.000
CL6	0.039	0.996	1.000	0.043	1.000	1.000	0.050	1.000	1.000
CL7	0.042	1.000	1.000	0.043	1.000	1.000	0.049	1.000	1.000
CL8	0.049	0.204	0.893	0.050	0.379	0.997	0.044	1.000	1.000
CL9	0.062	0.914	1.000	0.043	0.989	1.000	0.064	1.000	1.000
CL10	0.045	0.218	0.723	0.056	0.378	0.975	0.045	0.944	1.000
CL11	0.059	0.510	0.993	0.056	0.763	1.000	0.056	1.000	1.000
CL12	0.073	0.152	0.655	0.054	0.254	0.967	0.051	0.969	1.000
CC1	0.061	0.456	0.751	0.047	0.995	1.000	0.048	1.000	1.000
CC2	0.054	0.506	0.798	0.043	0.994	1.000	0.056	1.000	1.000
CC3	0.061	0.706	0.932	0.042	1.000	1.000	0.058	1.000	1.000
CC4	0.049	0.837	0.958	0.048	1.000	1.000	0.052	1.000	1.000
CC5	0.059	0.431	0.720	0.050	1.000	1.000	0.051	1.000	1.000
CC6	0.069	0.123	0.270	0.045	0.759	0.960	0.034	0.958	0.993
CC7	0.048	0.112	0.201	0.059	0.724	0.976	0.044	0.989	1.000
CC8	0.043	0.693	0.945	0.054	1.000	1.000	0.050	1.000	1.000
CC9	0.043	0.325	0.600	0.057	1.000	1.000	0.042	1.000	1.000
CC10	0.047	1.000	1.000	0.041	1.000	1.000	0.042	1.000	1.000
CC11	0.041	0.973	1.000	0.047	1.000	1.000	0.053	1.000	1.000
CC12	0.062	0.899	0.993	0.058	1.000	1.000	0.048	1.000	1.000

Table 6.1: Empirical size and power of the circular-linear and circular-circular goodness-of-fit tests for models CL1–CL12 and CC1–CC12 (respectively) with significance level $\alpha = 0.05$ and different sample sizes and deviations.

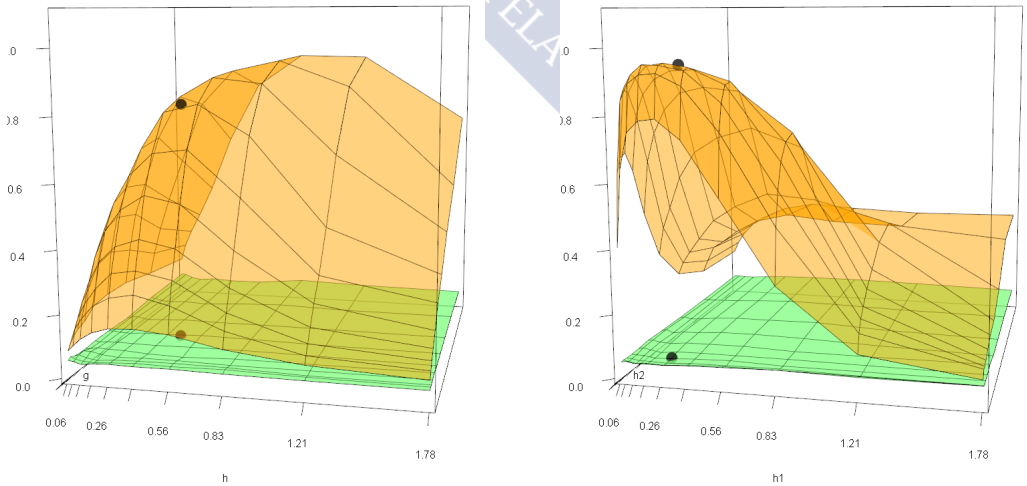


Figure 6.3: Empirical size and power of the circular-linear (left, model CL1) and circular-circular (right, model CC8) goodness-of-fit tests for a 10×10 logarithmic spaced grid. Lower surface represents the empirical rejection rate under $H_{0.00}$ and upper surface under $H_{0.15}$. Green colour indicates that the percentage of rejections is in the 95% confidence interval of $\alpha = 0.05$, blue that is smaller and orange that is larger. Black points represent the empirical size and power obtained with the median of the LCV bandwidths.

Finally, the effect of the bandwidths is explored in Figure 6.3. For models CL1 and CC8, the empirical size and power of the tests are computed on a bivariate grid of bandwidths, for sample size $n = 100$ and deviations $\delta = 0$ (green surface, null hypothesis) and $\delta = 0.15$ (orange surface). As it can be seen, the tests are correctly calibrated regardless the bandwidths value. However, the power is notably affected by the choice of the bandwidths, with different behaviours depending on the model and the alternative. Reasonable choices of the bandwidths, such as the median of the LCV bandwidths (6.7), present a competitive power. Further results supporting the same conclusions are available on supplement.

6.7 Real data application

The proposed goodness-of-fit tests are applied to study two real datasets (see supplementary material for further details). The first dataset comes from forestry and contains orientations and log-burnt areas of 26870 wildfires occurred in Portugal between 1985 and 2005. Data was aggregated in watersheds, giving 102 observations of the circular mean orientation and mean log-burnt area for each watershed (circular-linear example). Further details on the data acquisition procedure, measurement of fires orientation and watershed delimitation can be seen in Barros et al. (2012) and García-Portugués et al. (2014). The model proposed by Mardia and Sutton (1978) was tested for this dataset (Figure 6.4, left) using the LCV bandwidths and $B = 1000$ bootstrap replicates, resulting a p-value of 0.156, showing no evidence against the null hypothesis.

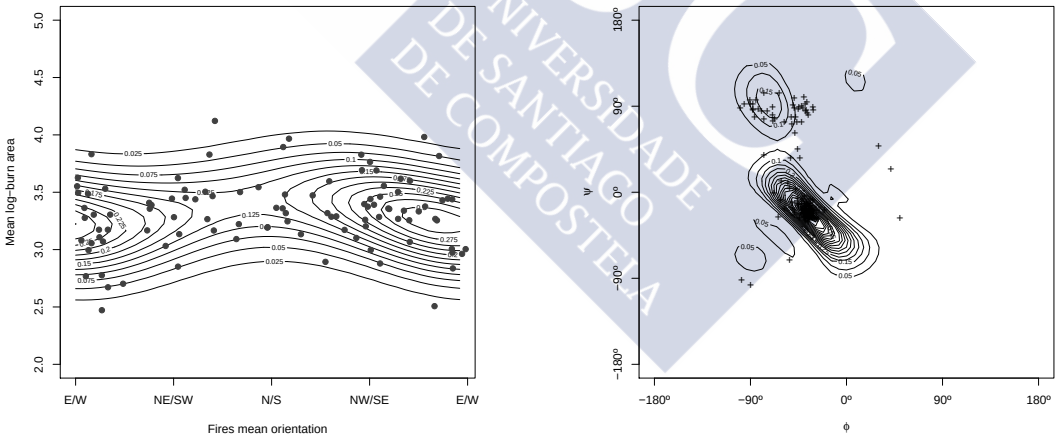


Figure 6.4: Left: parametric fit (model from Mardia and Sutton (1978)) to the circular mean orientation and mean log-burnt area of the fires in each of the 102 watersheds of Portugal. Right: parametric fit (model from Fernández-Durán (2007)) for the dihedral angles of the alanine-alanine-alanine segments.

The second dataset contains pairs of dihedral angles of segments of the type alanine-alanine-alanine in alanine amino acids in 1932 proteins. The dataset, formed by 233 pairs of angles (circular-circular), was studied by Fernández-Durán (2007) using Nonnegative Trigonometric Sums (NTSS) for the marginal and link function of the model of Wehrly and Johnson (1979). The best model in terms of BIC described in Fernández-Durán (2007) was implemented using two-step Maximum Likelihood Estimation (MLE) procedure and the tools of the `CircNNTSR` package Fernández-Durán and Gregorio-Domínguez (2013) for fitting the NTSS parametric densities (Figure 6.4, right). The resulting p-value with the LCV bandwidths is 0.000, indicating that the dependence model of Wehrly and Johnson (1979) is not flexible enough to capture the

dependence structure between the two angles. The reason for this lack of fit may be explained by a poor fit in a secondary cluster of data around $\Psi = 90°$, as it can be seen in the contour plot shown in Figure 6.4.

Acknowledgements

This research has been supported by project MTM2008–03010 from the Spanish Ministry of Science and StuDyS network, from the Interuniversity Attraction Poles Programme (IAP–network P7/06), Belgian Science Policy Office. First author's work has been supported by FPU grant AP2010–0957 from the Spanish Ministry of Education. Authors acknowledge the computational resources used at the SVG cluster of the CESGA Supercomputing Center. The editors and two anonymous referees are acknowledged for their contributions.

6.A Sketches of the main proofs

This section contains the sketches of the main proofs. Proofs for technical lemmas, complete numerical experiments and simulation results and further details on data analysis are given in supplement.

6.A.1 CLT for the integrated squared error

Proof of Theorem 6.1. The ISE can be decomposed into four addends, $I_n = I_{n,1}+I_{n,2}+I_{n,3}+I_{n,4}$:

$$
\begin{aligned}
I_{n,1} &= 2\frac{c_{h,q}(L)}{ng}\sum_{i=1}^{n}\int_{\Omega_q\times\mathbb{R}} LK_n\left((\mathbf{x},z),(\mathbf{X}_i,Z_i)\right)\left(\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x},z)\right]-f(\mathbf{x},z)\right)\,dz\,\omega_q(d\mathbf{x}),\\
I_{n,2} &= \frac{c_{h,q}(L)^2}{n^2g^2}\sum_{i=1}^{n}\int_{\Omega_q\times\mathbb{R}} LK_n^2\left((\mathbf{x},z),(\mathbf{X}_i,Z_i)\right)\,dz\,\omega_q(d\mathbf{x}),\\
I_{n,3} &= \frac{c_{h,q}(L)^2}{n^2g^2}\sum_{1\le i<j\le n}\int_{\Omega_q\times\mathbb{R}} LK_n\left((\mathbf{x},z),(\mathbf{X}_i,Z_i)\right)LK_n\left((\mathbf{x},z),(\mathbf{X}_j,Z_j)\right)\,dz\,\omega_q(d\mathbf{x}),\\
I_{n,4} &= \int_{\Omega_q\times\mathbb{R}}\left(\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x},z)\right]-f(\mathbf{x},z)\right)^2\,dz\,\omega_q(d\mathbf{x}),
\end{aligned}
$$

where $LK_n\left((\mathbf{x},z),(\mathbf{y},t)\right) = LK\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2},\frac{z-t}{g}\right) - \mathbb{E}\left[LK\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2},\frac{z-Z}{g}\right)\right]$.

Except for the fourth term, which is deterministic, the CLT for the ISE is derived by examining the asymptotic behaviour of each addend. The first two ones can be written as $I_{n,1} = \sum_{i=1}^{n} I_{n,1}^{(i)}$ and $I_{n,2} = \frac{c_{h,q}(L)^2}{n^2g^2}\sum_{i=1}^{n} I_{n,2}^{(i)}$, where $I_{n,1}^{(i)}$ and $I_{n,2}^{(i)}$ can be directly extracted from the previous expressions. Then, by Lemma A.2,

$$
n^{\frac{1}{2}}\phi(h,g)^{-\frac{1}{2}} I_{n,1} \xrightarrow{d} \mathcal{N}(0,1) \tag{6.8}
$$

and by Lemma A.3,

$$
I_{n,2} = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^qg} + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{3}{2}}h^{-q}g^{-1}\right). \tag{6.9}
$$

The third term can be written as follows:

$$I_{n,3} = 2\frac{c_{h,q}(L)^2}{n^2 g^2} \sum_{1\leq i<j\leq n} H_n\left((\mathbf{X}_i, Z_i), (\mathbf{X}_j, Z_j)\right) = 2\frac{c_{h,q}(L)^2}{n^2 g^2} U_n, \tag{6.10}$$

where U_n is an U-statistic with kernel function H_n given in Lemma A.4. U_n is degenerate since $\mathbb{E}\left[LK_n\left((\mathbf{x}, z), (\mathbf{X}, Z)\right)\right] = 0$.

In order to properly apply Lemma A.1 for obtaining the asymptotic distribution of U_n in (6.10), Lemma A.4 provides the explicit expressions for the required elements. Then, considering $\varphi_n \equiv 0$ in Lemma A.1, condition $A_n B_n^{-2} \to 0$ is satisfied by assumption **A3** and as a consequence $B_n^{-\frac{1}{2}} U_n \xrightarrow{d} \mathcal{N}(0,1)$. Since the variance of $I_{n,3}$ is

$$\mathbb{V}\mathrm{ar}\left[I_{n,3}\right] = 4\frac{c_{h,q}(L)^4}{n^4 g^4}\mathbb{V}\mathrm{ar}\left[U_n\right] = 2\frac{\sigma^2}{n^2 h^q g}(1 + o(1)), \tag{6.11}$$

by Slutsky's theorem, results (6.10) and (6.1),

$$n\left(h^q g\right)^{\frac{1}{2}} I_{n,3} \xrightarrow{d} \mathcal{N}\left(0, 2\sigma^2\right). \tag{6.12}$$

From (6.8), (6.9) and (6.12), it follows that:

$$I_n - \mathbb{E}\left[I_n\right] = n^{-\frac{1}{2}}\phi(h,g)^{\frac{1}{2}} N_{n,1} + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{3}{2}} h^{-q} g^{-1}\right) + 2^{\frac{1}{2}}\sigma n^{-1}(h^q g)^{-\frac{1}{2}} N_{n,3}, \tag{6.13}$$

where $N_{n,1}, N_{n,3} \xrightarrow{d} \mathcal{N}(0,1)$. By **A3**, $\left(n^{\frac{3}{2}} h^q g\right)^{-1} = o((nh^{\frac{q}{2}} g^{\frac{1}{2}})^{-1})$ and the second addend $I_{n,2}$ is asymptotically negligible compared with $I_{n,3}$. In order to determine the dominant term between $I_{n,1}$ and $I_{n,3}$, the squared quotient between the coefficients of both of them is examined, being of order $n\phi(h,g)h^q g$. Then if $n\phi(h,g)h^q g \to \infty$ the last term on (6.13) is asymptotically negligible in comparison with the first, while if $n\phi(h,g)h^q g \to 0$, the first term is negligible in comparison with the last. Note also that by (6.9), expression (6.13) can also be stated as

$$\begin{aligned} I_n - \left(\int_{\Omega_q \times \mathbb{R}} \left(\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x}, z)\right] - f(\mathbf{x}, z)\right)^2 dz\, \omega_q(d\mathbf{x}) + \frac{\lambda_q(L^2)\lambda_q(L)^{-2} R(K)}{nh^q g} \right) \\ = n^{-\frac{1}{2}}\phi(h,g)^{\frac{1}{2}} N_{n,1} + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{3}{2}} h^{-q} g^{-1}\right) + 2^{\frac{1}{2}}\sigma n^{-1}(h^q g)^{-\frac{1}{2}} N_{n,3}. \end{aligned}$$

The case where $n\phi(h,g)h^q g \to \delta$, $0 < \delta < \infty$, needs a special treatment because none of the terms can be neglected. In this case,

$$\begin{aligned} I_n - \mathbb{E}\left[I_n\right] &= n^{-\frac{1}{2}}\phi(h,g)^{\frac{1}{2}} N_{n,1} + 2^{\frac{1}{2}}\sigma n^{-1}(h^q g)^{-\frac{1}{2}} N_{n,3} + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{3}{2}} h^{-q} g^{-1}\right) \\ &= n^{-1}(h^q g)^{-\frac{1}{2}}\left(\delta^{\frac{1}{2}} N_{n,1} + 2^{\frac{1}{2}}\sigma N_{n,3}\right) + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{3}{2}} h^{-q} g^{-1}\right). \end{aligned}$$

In order to apply Lemma A.1, set $\widetilde{U}_n = I_{n,1} + I_{n,3}$ with

$$\widetilde{U}_n = \sum_{i=1}^{n} \varphi_n(\mathbf{X}_i, Z_i) + \sum_{1\leq i<j\leq n} \widetilde{H}_n\left((\mathbf{X}_i, Z_i), (\mathbf{X}_j, Z_j)\right),$$

where $\varphi_n(\mathbf{X}_1, Z_1) = I_{n,1}^{(1)}$, $\widetilde{H}_n\left((\mathbf{x}, z), (\mathbf{y}, t)\right) = 2\frac{c_{h,q}(L)^2}{n^2 g^2} H_n\left((\mathbf{x}, z), (\mathbf{y}, t)\right)$ and $\widetilde{G}_n\left((\mathbf{x}, z), (\mathbf{y}, t)\right) = \mathbb{E}\left[\widetilde{H}_n\left((\mathbf{X}, Z), (\mathbf{x}, z)\right)\widetilde{H}_n\left((\mathbf{X}, Z), (\mathbf{y}, t)\right)\right]$.

By Lemma A.4 and the definitions of $\widetilde{H}_n$, $\widetilde{G}_n$, φ_n and M_n,

$$\begin{aligned}
\mathbb{E}\left[\widetilde{H}_n^2\left((\mathbf{X}_1,Z_1),(\mathbf{X}_2,Z_2)\right)\right] &= 4n^{-4}h^{-q}g^{-1}\sigma^2\left(1+o\left(1\right)\right),\\
\mathbb{E}\left[\widetilde{H}_n^4\left((\mathbf{X}_1,Z_1),(\mathbf{X}_2,Z_2)\right)\right] &= \mathcal{O}\left(n^{-8}h^{-3q}g^{-3}\right),\\
\mathbb{E}\left[\widetilde{G}_n^2\left((\mathbf{X}_1,Z_1),(\mathbf{X}_2,Z_2)\right)\right] &= \mathcal{O}\left(n^{-8}h^{-q}g^{-1}\right),\\
\mathbb{E}\left[\varphi_n^2(\mathbf{X}_1,Z_1)\right] &= n^{-2}\phi(h,g)\left(1+o\left(1\right)\right),\\
\mathbb{E}\left[\varphi_n^4(\mathbf{X}_1,Z_1)\right] &= \mathcal{O}\left(n^{-4}(h^8+g^8)\right),\\
\mathbb{E}\left[M_n^2(\mathbf{X}_1,Z_1)\right] &= \mathcal{O}\left(n^{-6}(h^4+g^4)h^{-\frac{3q}{2}}g^{-\frac{3}{2}}\right).
\end{aligned}$$

Applying these orders and using that $n\phi(h,g)h^qg\to\delta$, it follows

$$\frac{A_n}{B_n^2}=\mathcal{O}\left(n^{-1}\right)+\mathcal{O}\left((nh^qg)^{-1}h^{\frac{q}{2}}g^{\frac{1}{2}}\right)+\mathcal{O}\left((nh^qg)^{-1}\right)+\mathcal{O}\left(h^qg\right).$$

Then, by **A3**, the four previous orders tend to zero and therefore $B_n^{-\frac{1}{2}}\widetilde{U}_n\xrightarrow{d}\mathcal{N}(0,1)$, where $B_n\sim n^{-1}\phi(h,g)+2n^{-2}(h^qg)^{-1}\sigma^2\sim n^{-2}(h^qg)^{-1}\left(\delta+2\sigma^2\right)$. Finally, $n\left(h^qg\right)^{\frac{1}{2}}2^{\frac{1}{2}}\left(\delta+2\sigma^2\right)^{-\frac{1}{2}}(I_{n,1}+I_{n,3})\xrightarrow{d}\mathcal{N}(0,1)$ by Slutsky's theorem. □

Proof of Corollary 6.1. As $g=\beta h$, for a fixed $\beta>0$, then $n\phi(h,g)h^qg=\mathcal{O}\left(nh^{q+5}\right)$ and the cases in Theorem 6.1 are given by the asymptotic behaviour of this sequence. When $nh^{q+5}\to\infty$ and $nh^{q+5}\to 0$, the result is obtained immediately, whereas for $nh^{q+5}\to\delta$, $0<\delta<\infty$ Lemma A.1 gives

$$B_n\sim\phi(1,\beta)n^{-1}h^4+2\sigma^2n^{-2}h^{-(q+1)}\sim n^{-\frac{q+9}{q+5}}\left(\phi(1,\beta)\delta^{\frac{4}{q+5}}+2\sigma^2\delta^{-\frac{q+1}{q+5}}\right).$$

Therefore, $n^{\frac{q+9}{2(q+5)}}\left(\phi(1,\beta)\delta^{\frac{4}{q+5}}+2\sigma^2\delta^{-\frac{q+1}{q+5}}\right)^{-\frac{1}{2}}(I_{n,1}+I_{n,3})\xrightarrow{d}\mathcal{N}(0,1)$. □

Proof of Corollary 6.2. The proof follows from an adaptation of the proof of Theorem 6.1 to the directional-directional context. □

6.A.2 Testing independence with directional data

Proof of Theorem 6.2. The test statistic is decomposed as $T_n=T_{n,1}+T_{n,2}+T_{n,3}$ taking into account that, under independence, $\mathbb{E}[\hat{f}_{h,g}(\mathbf{x},z)]=\mathbb{E}[\hat{f}_h(\mathbf{x})]\mathbb{E}[\hat{f}_g(z)]$:

$$\begin{aligned}
T_{n,1}&=\int_{\Omega_q\times\mathbb{R}}\left(\hat{f}_{h,g}(\mathbf{x},z)-\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x},z)\right]\right)^2dz\,\omega_q(d\mathbf{x}),\\
T_{n,2}&=\int_{\Omega_q\times\mathbb{R}}\left(\hat{f}_h(\mathbf{x})\hat{f}_g(z)-\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]\mathbb{E}\left[\hat{f}_g(z)\right]\right)^2dz\,\omega_q(d\mathbf{x}),\\
T_{n,3}&=-2\int_{\Omega_q\times\mathbb{R}}\left(\hat{f}_{h,g}(\mathbf{x},z)-\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x},z)\right]\right)\left(\hat{f}_h(\mathbf{x})\hat{f}_g(z)-\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]\mathbb{E}\left[\hat{f}_g(z)\right]\right)dz\,\omega_q(d\mathbf{x}).
\end{aligned}$$

By Chebychev's inequality and Lemmas A.6 and A.7, the sum of the second and third addends is $-\mathbb{E}\left[T_{n,2}\right]+\mathcal{O}_{\mathbb{P}}(n^{-1}(h^{-q}+g^{-1})^{\frac{1}{2}})$. Considering the test statistic decomposition and using Lemma A.5 yields

$$T_n=T_{n,1}-\mathbb{E}\left[T_{n,2}\right]+\mathcal{O}_{\mathbb{P}}\left(n^{-1}(h^{-q}+g^{-1})^{\frac{1}{2}}\right)$$

$$= \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g} + 2^{\frac{1}{2}}\sigma n^{-1}(h^q g)^{-\frac{1}{2}} N_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(f_Z)}{nh^q},$$
$$- \frac{R(K)R(f_{\mathbf{X}})}{ng} + o\left(n^{-1}(h^{-q}+g^{-1})\right) + \mathcal{O}_{\mathbb{P}}\left(n^{-1}(h^{-q}+g^{-1})^{\frac{1}{2}}\right).$$

Now, $\mathcal{O}_{\mathbb{P}}(n^{-1}(h^{-q}+g^{-1})^{\frac{1}{2}})$ is negligible in comparison with the second addend by condition **A3** and the deterministic order $o(n^{-1}(h^{-q}+g^{-1}))$ is also negligible by **A3** and **A4**. Therefore, $n(h^q g)^{\frac{1}{2}}(T_n - A_n) \xrightarrow{d} \mathcal{N}(0, 2\sigma_I^2)$. □

6.A.3 Goodness-of-fit test for models with directional data

Proof of Theorem 6.3. Under $H_0 : f = f_{\boldsymbol{\theta}_0}$, the test statistic $R_n = I_{n,2} + I_{n,3}$, where $I_{n,2}$ and $I_{n,3}$ are given by (6.9) and (6.12) in the proof of Theorem 6.1, so

$$n(h^q g)^{\frac{1}{2}}\left(I_{n,2} + I_{n,3} - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g}\right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma^2_{\boldsymbol{\theta}_0}\right). \tag{6.14}$$

□

Proof of Theorem 6.4. The test statistic is decomposed as $R_n = R_{n,1} + I_{n,2} + I_{n,3} + R_{n,4}$ by adding and subtracting $\mathbb{E}[\hat{f}_{h,g}(\mathbf{x}, z)] = LK_{h,g}f(\mathbf{x}, z)$, with

$$R_{n,1} = 2\int_{\Omega_q\times\mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - LK_{h,g}f(\mathbf{x}, z)\right) LK_{h,g}\left(f(\mathbf{x}, z) - f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z)\right) dz\, \omega_q(d\mathbf{x}),$$
$$R_{n,4} = \int_{\Omega_q\times\mathbb{R}} \left(LK_{h,g}\left(f(\mathbf{x}, z) - f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z)\right)\right)^2 dz\, \omega_q(d\mathbf{x}).$$

The limit of $I_{n,2} + I_{n,3}$ is given by (6.14) whereas by Lemma A.8 $R_{n,1}$ and $R_{n,4}$ are negligible in probability. Then, the limit distribution of R_n is determined by $I_{n,2} + I_{n,3}$. □

Proof of Theorem 6.5. As in the proof of Theorem 6.4, $R_n = \widetilde{R}_{n,1} + I_{n,2} + I_{n,3} + \widetilde{R}_{n,4}$, where $I_{n,2} + I_{n,3}$ behaves as (6.14). The asymptotic variance remains $\sigma^2_{\boldsymbol{\theta}_0}$ since

$$R(f) = R(f_{\boldsymbol{\theta}_0}) + \frac{R(\Delta)}{nh^{\frac{q}{2}}g^{\frac{1}{2}}} + \frac{\int_{\Omega_q\times\mathbb{R}} f(\mathbf{x}, z)\Delta(\mathbf{x}, z)\, dz\, \omega_q(d\mathbf{x})}{n^{\frac{1}{2}}h^{\frac{q}{4}}g^{\frac{1}{4}}}$$

and then the second and third addends are negligible with respect to the first one by assumption **A3**, remaining the same asymptotic variance. The terms $\widetilde{R}_{n,1} = R_{n,1} + \widetilde{R}^{(1)}_{n,1}$ and $\widetilde{R}_{n,4} = R_{n,4} + \widetilde{R}^{(1)}_{n,4} + \widetilde{R}^{(2)}_{n,4}$ are decomposed as follows:

$$\widetilde{R}^{(1)}_{n,1} = \frac{2}{\sqrt{nh^{\frac{q}{2}}g^{\frac{1}{2}}}} \int_{\Omega_q\times\mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - LK_{h,g}f(\mathbf{x}, z)\right) LK_{h,g}\Delta(\mathbf{x}, z)\, dz\, \omega_q(d\mathbf{x}),$$
$$\widetilde{R}^{(1)}_{n,4} = \frac{1}{nh^{\frac{q}{2}}g^{\frac{1}{2}}} \int_{\Omega_q\times\mathbb{R}} \left(LK_{h,g}\Delta(\mathbf{x}, z)\right)^2 dz\, \omega_q(d\mathbf{x}),$$
$$\widetilde{R}^{(2)}_{n,4} = \frac{2}{\sqrt{nh^{\frac{q}{2}}g^{\frac{1}{2}}}} \int_{\Omega_q\times\mathbb{R}} LK_{h,g}\left(f(\mathbf{x}, z) - f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z)\right) LK_{h,g}\Delta(\mathbf{x}, z)\, dz\, \omega_q(d\mathbf{x}).$$

The remaining terms follow from Lemma A.9. □

Proof of Theorem 6.6. Similarly to the proof of Theorem 6.4, $R_n^* = R_{n,1}^* + I_{n,2}^* + I_{n,3}^* + R_{n,4}^*$, where the terms involved are the bootstrap versions of the ones defined in the aforementioned proof:

$$
\begin{aligned}
R_{n,1}^* &= 2\int_{\Omega_q\times\mathbb{R}} \left(\hat{f}_{h,g}^*(\mathbf{x},z) - LK_{h,g}f_{\hat{\boldsymbol{\theta}}}(\mathbf{x},z)\right) LK_{h,g}\left(f_{\hat{\boldsymbol{\theta}}}(\mathbf{x},z) - f_{\hat{\boldsymbol{\theta}}^*}(\mathbf{x},z)\right)\, dz\,\omega_q(d\mathbf{x}),\\
I_{n,2}^* &= \frac{c_{h,q}(L)^2}{n^2g^2}\sum_{i=1}^{n}\int_{\Omega_q\times\mathbb{R}} \left(LK_n^*\left((\mathbf{x},z),(\mathbf{X}_i^*,Z_i^*)\right)\right)^2\, dz\,\omega_q(d\mathbf{x}),\\
I_{n,3}^* &= \frac{c_{h,q}(L)^2}{n^2g^2}\sum_{1\le i<j\le n}\int_{\Omega_q\times\mathbb{R}} LK_n^*\left((\mathbf{x},z),(\mathbf{X}_i^*,Z_i^*)\right) LK_n^*((\mathbf{x},z),(\mathbf{X}_j^*,Z_j^*))\, dz\,\omega_q(d\mathbf{x}),\\
R_{n,4}^* &= \int_{\Omega_q\times\mathbb{R}} \left(LK_{h,g}\left(f_{\hat{\boldsymbol{\theta}}}(\mathbf{x},z) - f_{\hat{\boldsymbol{\theta}}^*}(\mathbf{x},z)\right)\right)^2\, dz\,\omega_q(d\mathbf{x}),
\end{aligned}
$$

with $LK_n^*((\mathbf{x},z),(\mathbf{y},t)) = LK\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2},\frac{z-t}{g}\right) - \mathbb{E}^*\left[LK\left(\frac{1-\mathbf{x}^T\mathbf{X}^*}{h^2},\frac{z-Z^*}{g}\right)\right]$ and where $\mathbb{E}^*$ represents the expectation with respect to $f_{\hat{\boldsymbol{\theta}}}$, which is obtained from the original sample.

Using the same arguments as in Lemma A.8, but replacing assumption **A6** by **A8**, it follows that $n(h^qg)^{\frac{1}{2}}R_{n,1}^*$ and $n(h^qg)^{\frac{1}{2}}R_{n,4}^*$ converge to zero conditionally on the sample, that is, in probability $\mathbb{P}^*$. On the other hand, the terms $I_{n,2}^*$ and $I_{n,3}^*$ follow from considering similar arguments to the ones used for deriving (6.9) and (6.12), but conditionally on the sample. Specifically, it follows that $I_{n,2}^* = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^qg} + \mathcal{O}_{\mathbb{P}^*}\left(n^{-\frac{3}{2}}h^{-q}g^{-1}\right)$ and, for a certain $\boldsymbol{\theta}_1\in\Theta$, $(nh^qg)^{\frac{1}{2}}I_{n,3}^* \xrightarrow{d} \mathcal{N}(0,2\sigma_{\boldsymbol{\theta}_1}^2)$. The main difference with the proof of Theorem 6.4 concerns the asymptotic variance given by $n(h^qg)^{\frac{1}{2}}I_{n,3}^*$: $\mathbb{V}\mathrm{ar}^*\left[n(h^qg)^{\frac{1}{2}}I_{n,3}^*\right] \xrightarrow{p} 2\sigma_{\boldsymbol{\theta}_1}^2$, since by assumption **A5**, $R(f_{\hat{\boldsymbol{\theta}}}) = R(f_{\boldsymbol{\theta}_1}) + \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$. Hence,

$$
n(h^qg)^{\frac{1}{2}}\left(R_n^* - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^qg}\right) = o_{\mathbb{P}^*}(1) + \mathcal{O}_{\mathbb{P}^*}\left((nh^qg)^{-\frac{1}{2}}\right) + 2^{\frac{1}{2}}\sigma_{\boldsymbol{\theta}_1}N_n + o_{\mathbb{P}^*}(1)
$$

and the bootstrap consistency follows. □

Proof of Corollary 6.3. The proof follows by adapting the proof of Theorem 6.4. □

References

Bai, Z. D., Rao, C. R., and Zhao, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.*, 27(1):24–39.

Barros, A. M. G., Pereira, J. M. C., and Lund, U. J. (2012). Identifying geographical patterns of wildfire orientation: a watershed-based analysis. *Forest. Ecol. Manag.*, 264:98–107.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.*, 1(6):1071–1095.

Boente, G., Rodríguez, D., and González-Manteiga, W. (2014). Goodness-of-fit test for directional data. *Scand. J. Stat.*, 41(1):259–275.

Fan, Y. (1994). Testing the goodness of fit of a parametric density function by kernel method. *Economet. Theor.*, 10(2):316–356.

Fernández-Durán, J. J. (2007). Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics*, 63(2):579–585.

Fernández-Durán, J. J. and Gregorio-Domínguez, M. M. (2013). *CircNNTSR: an R package for the statistical analysis of circular data using NonNegative Trigonometric Sums (NNTS) models.* R package version 2.1.

Fisher, N. I. and Lee, A. J. (1981). Nonparametric measures of angular-linear association. *Biometrika*, 68(3):629–636.

García-Portugués, E. (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron. J. Stat.*, 7:1655–1685.

García-Portugués, E., Barros, A. M. G., Crujeiras, R. M., González-Manteiga, W., and Pereira, J. (2014). A test for directional-linear independence, with applications to wildfire orientation and size. *Stoch. Environ. Res. Risk Assess.*, 28(5):1261–1275.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013a). Exploring wind direction and SO_2 concentration by circular-linear density estimation. *Stoch. Environ. Res. Risk Assess.*, 27(5):1055–1067.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013b). Kernel density estimation for directional-linear data. *J. Multivariate Anal.*, 121:152–175.

González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *Test*, 22(3):361–411.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.*, 14(1):1–16.

Hall, P., Watson, G. S., and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762.

Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *J. Amer. Statist. Assoc.*, 73(363):602–606.

Klemelä, J. (2000). Estimation of densities and derivatives of densities with directional data. *J. Multivariate Anal.*, 73(1):18–40.

Mardia, K. V. (1976). Linear-circular correlation coefficients and rhythmometry. *Biometrika*, 63(2):403–405.

Mardia, K. V. and Jupp, P. E. (2000). *Directional statistics.* Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, second edition.

Mardia, K. V. and Sutton, T. W. (1978). A model for cylindrical variables with applications. *J. Roy. Statist. Soc. Ser. B*, 40(2):229–233.

Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2012). A plug-in rule for bandwidth selection in circular density estimation. *Comput. Statist. Data Anal.*, 56(12):3898–3908.

Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.*, 3(1):1–14.

Rosenblatt, M. and Wahlen, B. E. (1992). A nonparametric measure of independence under a hypothesis of independent components. *Statist. Probab. Lett.*, 15(3):245–252.

Scott, D. W. (1992). *Multivariate density estimation*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. John Wiley & Sons, New York.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Srivastava, A. N. and Sahami, M., editors (2009). *Text mining: classification, clustering, and applications*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press, Boca Raton.

Taylor, C. C. (2008). Automatic bandwidth selection for circular density estimation. *Comput. Statist. Data Anal.*, 52(7):3493–3500.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Watson, G. S. (1983). *Statistics on spheres*, volume 6 of *University of Arkansas Lecture Notes in the Mathematical Sciences*. John Wiley & Sons, New York.

Wehrly, T. E. and Johnson, R. A. (1979). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, 67(1):255–256.

Zhao, L. and Wu, C. (2001). Central limit theorem for integrated square error of kernel estimators of spherical density. *Sci. China Ser. A*, 44(4):474–483.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 7

Testing parametric models in linear-directional regression

Abstract

This paper presents a goodness-of-fit test for parametric regression models with scalar response and directional predictor, that is, vectors in a sphere of arbitrary dimension. The testing procedure is based on the weighted squared distance between a smooth and a parametric regression estimator, where the smooth regression estimator is obtained by a projected local approach. Asymptotic behavior of the test statistic under the null hypothesis and local alternatives is provided, jointly with a consistent bootstrap algorithm for application in practice. A simulation study illustrates the performance of the test in finite samples. The procedure is also applied to a real data example from text mining.

Reference

García-Portugués, E., Van Keilegom, I., Crujeiras, R. M. and González-Manteiga, W. (2014). Testing parametric models in linear-directional regression. *arXiv 1409.0506*.

Contents

7.1 Introduction 152
7.2 Nonparametric linear-directional regression 153
7.2.1 The projected local estimator 154
7.2.2 Properties 155
7.3 Goodness-of-fit test for linear-directional regression 156
7.3.1 Bootstrap calibration 158
7.4 Simulation study 160
7.5 Application to text mining 162
7.A Main results 165
7.A.1 Projected local estimator properties 165
7.A.2 Asymptotic results for the goodness-of-fit test 168
References 172

7.1 Introduction

Directional data (data on a general sphere of dimension q) appear in a variety of contexts being the simplest one provided by observations of angles on a circle (circular data), for instance, from wind directions or animal orientation (Mardia and Jupp, 2000). Stars positions could be seen as data on a two-dimensional sphere and quite recently, directional data in higher dimensions have been considered in text mining (Srivastava and Sahami, 2009). In order to identify a statistical pattern within a certain collection of texts, these objects may be represented by a vector on a sphere where each vector component gives the relative frequency of a certain word. For instance, from this vector-space representation, text classification (see Banerjee et al. (2005)) can be performed, but other interesting problems such as popularity prediction could be tackled. Such a characterization can be done, for instance, for articles in news aggregators, where popularity prediction in web texts (news) can be quantified by the number of comments or views (see Tatar et al. (2012)). In addition, in order to model or predict the popularity of a certain web entry based on its contents (in a vector-space form), a linear-directional regression model could be used.

When dealing with directional and linear variables at the same time, the joint behavior could be modeled by considering a flexible density estimator, as the one proposed by García-Portugués et al. (2013). Nevertheless, a regression approach may be more useful, allowing at the same time for explaining a relation between the variables and for making predictions. Nonparametric regression estimation methods for linear-directional models have been proposed by different authors. For instance, Cheng and Wu (2013) introduced a general local linear regression method on manifolds, and quite recently, Di Marzio et al. (2014) presented a local polynomial method for the regression function when both the predictor and the response are defined on spheres.

Despite the fact that these methods provide flexible estimators which may capture the regression shape, in terms of interpretation of the results, purely parametric models may be more convenient. In this context, goodness-of-fit testing methods can be designed, providing a tool for assessing a certain parametric linear-directional regression model. Up to the authors' knowledge, such a problem has not been considered in the statistical literature, with the exception of Deschepper et al. (2008), who propose an exploratory tool and a lack-of-fit test for circular-linear regression. A goodness-of-fit test for parametric regression models will be presented in this paper. The test is based on a squared distance between the parametric fit and a nonparametric one. Specifically, a modified local linear estimator, similar to the proposal of Di Marzio et al. (2014) will be introduced with this purpose. Theoretical properties of the test statistic will be studied, and its effectiveness in practice will be confirmed by simulation results.

The paper is organized as follows. Basic notation is introduced in Section 7.2, where the projected local regression estimator is also analyzed. Section 7.3 includes the main results, regarding the asymptotic behavior of the test statistic. A consistent bootstrap strategy is also presented. The performance of the proposed method is assessed for finite samples in a simulation study, provided in Section 7.4. Section 7.5 shows a real data application on news popularity prediction. An appendix contains the proofs of the main results. Proofs of technical lemmas and further information on the simulation study, jointly with more simulation results, are provided as supplementary material.

7.2 Nonparametric linear-directional regression

Some basic concepts in nonparametric directional density and linear-directional regression estimation will be provided in this section. Basic notation will be introduced, jointly with some motivation for the regression estimator proposal that will be used in the testing procedure.

Let $\Omega_q = \{\mathbf{x} \in \mathbb{R}^{q+1} : x_1^2 + \cdots + x_{q+1}^2 = 1\}$ denote the q-sphere in $\mathbb{R}^{q+1}$, with associated Lebesgue measure denoted by ω_q (when there is no possible confusion, the surface area of Ω_q will be denoted by $\omega_q = 2\pi^{\frac{q+1}{2}}/\Gamma(\frac{q+1}{2})$, $q \geq 0$). A directional density f on Ω_q satisfies $\int_{\Omega_q} f(\mathbf{x})\,\omega_q(d\mathbf{x}) = 1$. Consider a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ in Ω_q, from a directional random variable $\mathbf{X}$ with density f. A kernel density estimator for f was introduced by Hall et al. (1987) and Bai et al. (1988). For a given point $\mathbf{x} \in \Omega_q$, the kernel density estimator is defined as

$$\hat{f}_h(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} L_h(\mathbf{x}, \mathbf{X}_i), \quad L_h(\mathbf{x}, \mathbf{y}) = c_{h,q}(L) L\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}\right), \tag{7.1}$$

where L is a directional kernel, $h > 0$ is the bandwidth parameter and the normalizing constant $c_{h,q}(L)$ is given by

$$c_{h,q}(L)^{-1} = \lambda_{h,q}(L)h^q = \lambda_q(L)h^q + o(1) \tag{7.2}$$

with $\lambda_{h,q}(L) = \omega_{q-1}\int_0^{2h^{-2}} L(r) r^{\frac{q}{2}-1}(2 - rh^2)^{\frac{q}{2}-1}\,dr$ and $\lambda_q(L) = 2^{\frac{q}{2}-1}\omega_{q-1}\int_0^{\infty} L(r) r^{\frac{q}{2}-1}\,dr$.

In many practical situations, the interest lies in the analysis of the directional variable $\mathbf{X}$ jointly with a real random variable Y. The joint behavior of both variables $(\mathbf{X}, Y)$ may be studied by a density approach, as in García-Portugués et al. (2013). However, a regression approach may be more suitable in some situations. Assume that the directional random variable $\mathbf{X}$ with density f may be the covariate in the following regression model

$$Y = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon, \tag{7.3}$$

where Y is a scalar random (response) variable, m is the regression function given by the conditional mean ($m(\mathbf{x}) = \mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$), and σ^2 is the conditional variance ($\sigma^2(\mathbf{x}) = \mathbb{V}\mathrm{ar}[Y|\mathbf{X} = \mathbf{x}]$). Errors are collected by ε, a random variable such that $\mathbb{E}[\varepsilon|\mathbf{X}] = 0$, $\mathbb{E}[\varepsilon^2|\mathbf{X}] = 1$ and $\mathbb{E}[|\varepsilon|^3|\mathbf{X}]$ and $\mathbb{E}[\varepsilon^4|\mathbf{X}]$ are bounded random variables.

The regression and density functions $m, f : \Omega_q \longrightarrow \mathbb{R}$ can be extended from Ω_q to $\mathbb{R}^{q+1}\setminus\{\mathbf{0}\}$ by considering a radial projection (see Zhao and Wu (2001) for the density case). This allows for the consideration of derivatives of these functions and the use of Taylor expansions.

A1. m and f are extended from Ω_q to $\mathbb{R}^{q+1}\setminus\{\mathbf{0}\}$ by $m(\mathbf{x}) \equiv m(\mathbf{x}/||\mathbf{x}||)$ and $f(\mathbf{x}) \equiv f(\mathbf{x}/||\mathbf{x}||)$. m is three times and f is twice continuously differentiable and f is bounded away from zero.

The continuity up to the second derivatives of f and up to the third derivatives of m, together with the q-spherical compact support, guarantees that these functions are in fact uniformly bounded. As a consequence of the radial extension, the directional derivative of m in the direction $\mathbf{x}$ and evaluated at $\mathbf{x}$ is zero, that is, $\mathbf{x}^T\boldsymbol{\nabla} m(\mathbf{x}) = 0$.

Consider, from now on, that a random sample from model (7.3), namely $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ independent and identically distributed (iid) vectors in $\Omega_q \times \mathbb{R}$, is available. Given two points $\mathbf{x}, \mathbf{X}_i \in \Omega_q$, under assumption **A1**, the one-term Taylor expansion of m at $\mathbf{X}_i$, conditionally on $\mathbf{X}_1, \ldots, \mathbf{X}_n$, can be written as:

$$\begin{aligned} m(\mathbf{X}_i) &= m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T (\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(||\mathbf{X}_i - \mathbf{x}||^2\right) \\ &= m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T \left(\mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T\right)(\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(||\mathbf{X}_i - \mathbf{x}||^2\right) \\ &= m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T \mathbf{B}_\mathbf{x} \mathbf{B}_\mathbf{x}^T (\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(||\mathbf{X}_i - \mathbf{x}||^2\right) \\ &\approx \beta_0 + \boldsymbol{\beta}_1^T \mathbf{B}_\mathbf{x}^T (\mathbf{X}_i - \mathbf{x}), \end{aligned}$$

with $\mathbf{B}_\mathbf{x} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$ the *projection matrix*. For a given $\mathbf{x} \in \Omega_q$, let $\{\mathbf{b}_1, \ldots, \mathbf{b}_q\}$ be a collection of q resulting vectors that complete $\mathbf{x}$ to an orthonormal basis $\{\mathbf{x}, \mathbf{b}_1, \ldots, \mathbf{b}_q\}$ of $\mathbb{R}^{q+1}$ (given, for example, by Gram-Schmidt). The projection matrix $\mathbf{B}_\mathbf{x} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)$ is a $(q+1)\times q$ semiorthogonal matrix, *i.e.* $\mathbf{B}_\mathbf{x}^T \mathbf{B}_\mathbf{x} = \mathbf{I}_q$, with $\mathbf{I}_q$ the identity matrix of dimension q. By the spectral decomposition theorem, $\mathbf{B}_\mathbf{x}\mathbf{B}_\mathbf{x}^T = \sum_{i=1}^q \mathbf{b}_i \mathbf{b}_i^T = \mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T$.

With this setting, the first coefficient $\beta_0 \in \mathbb{R}$ captures the constant effect in $m(\mathbf{x})$ while the second one, $\boldsymbol{\beta}_1 \in \mathbb{R}^q$, contains the linear effects of the *projected gradient* of m given by $\mathbf{B}_\mathbf{x}^T \boldsymbol{\nabla} m(\mathbf{x})$. It should be noted that $\boldsymbol{\beta}_1$ has dimension q (an appropriate dimension in the q-sphere Ω_q, instead of having dimension $q+1$, the one that will arise from an usual Taylor expansion in $\mathbb{R}^{q+1}$). The previous Taylor expansion provides the motivation for the *projected local estimator* of the regression function m at $\mathbf{x} \in \Omega_q$ that will be introduced and analyzed in the next section.

7.2.1 The projected local estimator

As in the previous section, consider $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ a random sample from model (7.3) and recall the expansion of the regression function derived under assumption **A1**:

$$m(\mathbf{X}_i) \approx \beta_0 + \boldsymbol{\beta}_1^T \mathbf{B}_\mathbf{x}^T (\mathbf{X}_i - \mathbf{x}).$$

The projected local estimator proposed in this work is obtained as a local fit by weighting the constants β_0 or the hyperplanes $\beta_0 + \boldsymbol{\beta}_1^T \mathbf{B}_\mathbf{x}^T (\mathbf{X}_i - \mathbf{x})$ according to the influence of $\mathbf{X}_i$ over $\mathbf{x}$. Both situations can be formulated together as a weighted least squares problem:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{q+1}} \sum_{i=1}^n \left(Y_i - \beta_0 - \delta_{p,1} \left(\beta_1, \ldots, \beta_q\right)^T \mathbf{B}_\mathbf{x}^T (\mathbf{X}_i - \mathbf{x})\right)^2 L_h(\mathbf{x}, \mathbf{X}_i), \tag{7.4}$$

where $\delta_{p,q}$ is the Kronecker Delta and is used to control both the local constant ($p = 0$) and local linear ($p = 1$) fits and L_h are the directional kernels, defined as in (7.1). The solution to the minimization problem (7.4) is given by

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_\mathbf{x} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_\mathbf{x} \mathbf{Y}, \tag{7.5}$$

where $\mathbf{Y}$ is the vector of observed responses, $\boldsymbol{\mathcal{W}}_\mathbf{x}$ is the weight matrix and $\boldsymbol{\mathcal{X}}_{\mathbf{x},p}$ is the design matrix. Specifically:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \boldsymbol{\mathcal{W}}_\mathbf{x} = \operatorname{diag}\left(L_h(\mathbf{x}, \mathbf{X}_1), \ldots, L_h(\mathbf{x}, \mathbf{X}_n)\right), \boldsymbol{\mathcal{X}}_{\mathbf{x},1} = \begin{pmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T \mathbf{B}_\mathbf{x} \\ \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T \mathbf{B}_\mathbf{x} \end{pmatrix}$$

and $\boldsymbol{\mathcal{X}}_{\mathbf{x},0} = \mathbf{1}_n$, with $\mathbf{1}_d$ denoting a column vector of length d with all entries equal to one and whose dimension will be omitted and determined by the context. The projected local (constant or linear) estimator at $\mathbf{x}$ is given by the estimated coefficient $\hat{\beta}_0 = \hat{m}_{h,p}(\mathbf{x})$ and is a weighted linear combination of the responses:

$$\hat{m}_{h,p}(\mathbf{x}) = \hat{\beta}_0 = \mathbf{e}_1^T \left(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \mathbf{Y} = \sum_{i=1}^{n} W_n^p(\mathbf{x}, \mathbf{X}_i) Y_i, \tag{7.6}$$

where $W_n^p(\mathbf{x}, \mathbf{X}_i) = \mathbf{e}_1^T \left(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \mathbf{e}_i$ ($\mathbf{e}_i$ denotes a unit canonical vector).

It should be noted that, for $p = 0$ (local constant), $W_n^0(\mathbf{x}, \mathbf{X}_i) = \frac{L_h(\mathbf{x},\mathbf{X}_i)}{\sum_{j=1}^n L_h(\mathbf{x},\mathbf{X}_j)}$. This corresponds to the Nadaraya-Watson estimator for directional predictor and scalar response, introduced by Wang et al. (2000).

Remark 7.1. *Di Marzio et al. (2014) have recently proposed a local linear estimator for model (7.3), but from a different approach. Specifically, these authors consider another alternative for developing the Taylor expansions of m based on the tangent-normal decomposition: $\mathbf{X}_i = \mathbf{x}\cos\eta_i + \boldsymbol{\xi}_i \sin\eta_i$, with $\eta_i \in [0, 2\pi)$ and $\boldsymbol{\xi}_i \in \Omega_q$ satisfying that $\boldsymbol{\xi}_i^T \mathbf{x} = 0$. This approach leads to an overparametrized design matrix of $q+2$ columns which makes $\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}$ exactly singular. This problem can be avoided in practice by computing a pseudo-inverse. Although both approaches come from different motivations, it can be seen that their asymptotic behavior is the same (conditional bias, variance and asymptotic normality). This can be checked by considering the different parametrization used in Di Marzio et al. (2014), where the smoothing parameter is $\kappa \equiv 1/h^2$. It should be also noted that, for the circular case ($q = 1$), the projected local estimator corresponds to the proposal by Di Marzio et al. (2009). See supplement for a detailed discussion of particular cases.*

7.2.2 Properties

Asymptotic bias and variance for the estimator (7.5), jointly with its asymptotic normality, will be derived in this section. Some further assumptions will be required:

A2. σ^2 is uniformly continuous and bounded away from zero.

A3. The directional kernel L is a bounded function $L : [0, \infty) \to [0, \infty)$ with exponential decay: $L(r) \leq Me^{-\alpha r}$, $\forall r \in [0, \infty)$, with $M, \alpha > 0$.

A4. The sequence of bandwidths $h = h_n$ is positive and satisfies $h \to 0$ and $nh^q \to \infty$.

Conditions **A2** and **A4** are the usual ones for the multivariate local linear estimator (see Ruppert and Wand (1994)). Condition **A3** allows for the use of non-compactly supported kernels, such as the *von Mises kernel* $L(r) = e^{-r}$, and implies condition **A2** in García-Portugués et al. (2014).

Theorem 7.1 (Conditional bias and variance)**.** *Under assumptions **A1**–**A4**, the conditional bias and variance for the projected local estimator with $p = 0, 1$ are given by*

$$\mathbb{E}\left[\hat{m}_{h,p}(\mathbf{x}) - m(\mathbf{x}) | \mathbf{X}_1, \ldots, \mathbf{X}_n\right] = \frac{b_q(L)}{q} B_p(\mathbf{x}) h^2 + o_{\mathbb{P}}\left(h^2\right) + \delta_{p,0} \mathcal{O}_{\mathbb{P}}\left(\frac{h}{\sqrt{nh^q}}\right),$$

$$\mathbb{V}\mathrm{ar}\left[\hat{m}_{h,p}(\mathbf{x}) | \mathbf{X}_1, \ldots, \mathbf{X}_n\right] = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q f(\mathbf{x})} \sigma^2(\mathbf{x}) + o_{\mathbb{P}}\left((nh^q)^{-1}\right),$$

uniformly for all $\mathbf{x} \in \Omega_q$*, where the terms in the bias are given by (*tr *stands for the trace):*

$$B_p(\mathbf{x}) = \begin{cases} 2\frac{\nabla f(\mathbf{x})^T \nabla m(\mathbf{x})}{f(\mathbf{x})} + \operatorname{tr}\left[\boldsymbol{\mathcal{H}}_m(\mathbf{x})\right], & p = 0, \\ \operatorname{tr}\left[\boldsymbol{\mathcal{H}}_m(\mathbf{x})\right], & p = 1, \end{cases} \qquad b_q(L) = \frac{\int_0^\infty L(r) r^{\frac{q}{2}}\, dr}{\int_0^\infty L(r) r^{\frac{q}{2}-1}\, dr}.$$

Remark 7.2. *As it happens in the Euclidean setting, the conditional bias is reduced from the local constant fit to the local linear one, whereas the variance remains the same for both estimators. The expressions and residual terms obtained in this setting agree with their Euclidean analogues (see Fan et al. (1996)), with the role of the kernel's second moment played by* $q^{-1}b_q(L)$ *and the integral of the squared kernel by* $\lambda_q(L^2)\lambda_q(L)^{-2}$*.*

From Theorem 7.1, an *equivalent kernel* expression can be obtained. Such a result allows for an explicit form of the weights $W_n^p(\mathbf{x}, \mathbf{X}_i)$, resulting an estimator asymptotically equivalent (in probability) to (7.6). This formulation, for $p = 1$, provides a simpler form for the weighting kernel than the one given in (7.6). In addition, the asymptotic expression will only depend on the datum $(\mathbf{X}_i, Y_i)$ and not on the whole data sample. This feature will be crucial for developing the goodness-of-fit test in Section 7.3 and the asymptotic normality of the estimators, collected in the next result.

Corollary 7.1 (Equivalent kernel)**.** *Under assumptions* ***A1****–****A4****, the projected local estimator* $\hat{m}_{h,p}(\mathbf{x}) = \sum_{i=1}^n W_n^p(\mathbf{x}, \mathbf{X}_i) Y_i$ *for* $p = 0, 1$ *satisfies uniformly in* $\mathbf{x} \in \Omega_q$*:*

$$\hat{m}_{h,p}(\mathbf{x}) = \sum_{i=1}^n L_h^*(\mathbf{x}, \mathbf{X}_i) Y_i \left(1 + o_{\mathbb{P}}(1)\right), \quad L_h^*(\mathbf{x}, \mathbf{X}_i) = \frac{1}{nh^q \lambda_q(L) f(\mathbf{x})} L\left(\frac{1 - \mathbf{x}^T \mathbf{X}_i}{h^2}\right).$$

Note that the equivalent kernel is the same for $p = 0$ and $p = 1$, as it happens in the Euclidean case when estimating the regression function (see pages 64–66 of Fan and Gijbels (1996) with $\nu = 0$ and $p = 0, 1$).

Theorem 7.2 (Asymptotic normality)**.** *Under assumptions* ***A1****–****A4*** *and for* $p = 0, 1$*, for every fixed point* $\mathbf{x} \in \Omega_q$ *such that* $\mathbb{E}\left[(Y - m(\mathbf{x}))^{2+\delta} | \mathbf{X} = \mathbf{x}\right] < \infty$*, for some* $\delta > 0$*,*

$$\sqrt{nh^q}\left(\hat{m}_{h,p}(\mathbf{x}) - m(\mathbf{x}) - \frac{b_q(L)}{q} B_p(\mathbf{x}) h^2\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{f(\mathbf{x})} \sigma^2(\mathbf{x})\right).$$

7.3 Goodness-of-fit test for linear-directional regression

In this section, a test statistic for assessing if the regression function m belongs to a class of parametric functions $\mathcal{M}_\Theta = \{m_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^s\}$ will be introduced. Assuming that model (7.3) holds, the goal is to test the null hypothesis

$$H_0 : m(\mathbf{x}) = m_{\boldsymbol{\theta}_0}(\mathbf{x}), \text{ for all } \mathbf{x} \in \Omega_q, \text{ versus } H_1 : m(\mathbf{x}) \neq m_{\boldsymbol{\theta}_0}(\mathbf{x}), \text{ for some } \mathbf{x} \in \Omega_q,$$

with $\boldsymbol{\theta}_0 \in \Theta$ known (simple hypothesis) or unknown (composite hypothesis) and where the statement *for all* holds except for a set of probability zero and *for some* holds for a set of positive probability.

The proposed statistic to test H_0 compares the projected local estimator introduced in Subsection 7.2.1 with a parametric estimator in $\mathcal{M}_\Theta$ throughout a squared weighted norm:

$$T_n = \int_{\Omega_q} \left(\hat{m}_{h,p}(\mathbf{x}) - \mathcal{L}_{h,p} m_{\hat{\boldsymbol{\theta}}}(\mathbf{x})\right)^2 \hat{f}_h(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x})$$

where $\mathcal{L}_{h,p}m(\mathbf{x}) = \sum_{i=1}^{n} W_n^p(\mathbf{x}, \mathbf{X}_i)\, m(\mathbf{X}_i)$ represents the local smoothing of the function m from measurements $\{\mathbf{X}_i\}_{i=1}^n$ and $\hat{\boldsymbol{\theta}}$ denotes either the known parameter $\boldsymbol{\theta}_0$ (simple null hypothesis) or a consistent estimator (composite null hypothesis; see condition **A6** below). This smoothing of the (possibly estimated) parametric regression function is included to reduce the asymptotic bias (Härdle and Mammen, 1993). In order to mitigate the effect of the difference between $\hat{m}_{h,p}$ and $m_{\hat{\boldsymbol{\theta}}}$ in sparse areas of the covariate, the squared difference is weighted by a kernel density estimate of $\mathbf{X}$, namely $\hat{f}_h$. Furthermore, by the inclusion of $\hat{f}_h$, the effects of the unknown density both on the asymptotic bias and variance are removed. Optionally, a weight function $w : \Omega_q \longrightarrow [0, \infty)$ can also be considered.

Two additional assumptions regarding the smoothness of the parametric regression function and the estimation of $\boldsymbol{\theta}_0$ in the composite hypothesis are required for deriving the distribution of T_n under the null hypothesis:

A5. $m_{\boldsymbol{\theta}}$ is continuously differentiable as a function of $\boldsymbol{\theta}$, and this derivative is also continuous for $\mathbf{x} \in \Omega_q$.

A6. Under H_0, there exists an $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$, *i.e.* $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$ and such that, under H_1, $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_1 = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$ for a certain $\boldsymbol{\theta}_1$.

Theorem 7.3 (Limit distribution of T_n)**.** *Under conditions **A1**–**A6** and under the null hypothesis $H_0 : m \in \mathcal{M}_\Theta$ (that is, $m(\mathbf{x}) = m_{\boldsymbol{\theta}_0}(\mathbf{x})$, for all $\mathbf{x} \in \Omega_q$),*

$$nh^{\frac{q}{2}} \left(T_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q} \int_{\Omega_q} \sigma^2_{\boldsymbol{\theta}_0}(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \right) \xrightarrow{d} \mathcal{N}\left(0, 2\nu^2_{\boldsymbol{\theta}_0}\right),$$

where $\sigma^2_{\boldsymbol{\theta}_0}(\mathbf{x}) = \mathbb{E}\left[(Y - m_{\boldsymbol{\theta}_0}(\mathbf{X}))^2 | \mathbf{X} = \mathbf{x}\right]$ is the conditional variance under H_0 and

$$\nu^2_{\boldsymbol{\theta}_0} = \int_{\Omega_q} \sigma^4_{\boldsymbol{\theta}_0}(\mathbf{x}) w(\mathbf{x})^2\, \omega_q(d\mathbf{x}) \times \gamma_q \lambda_q(L)^{-4} \int_0^\infty r^{\frac{q}{2}-1} \left\{ \int_0^\infty \rho^{\frac{q}{2}-1} L(\rho) \varphi_q(r, \rho)\, d\rho \right\}^2 dr,$$

$$\varphi_q(r,\rho) = \begin{cases} L\left(r + \rho - 2(r\rho)^{\frac{1}{2}}\right) + L\left(r + \rho + 2(r\rho)^{\frac{1}{2}}\right), & q = 1, \\ \int_{-1}^{1} (1-\theta^2)^{\frac{q-3}{2}} L\left(r + \rho - 2\theta(r\rho)^{\frac{1}{2}}\right) d\theta, & q \geq 2, \end{cases}$$

$$\gamma_q = \begin{cases} 2^{-\frac{1}{2}}, & q = 1, \\ \omega_{q-1}\omega_{q-2}^2 2^{\frac{3q}{2}-3}, & q \geq 2. \end{cases}$$

It should be noted that the convergence rate as well as the asymptotic bias and variance agree with the results in the multivariate setting given by Härdle and Mammen (1993) and Alcalá et al. (1999), except for the cancellation of the design density effects in bias and variance, achieved by the inclusion of $\hat{f}_h$ in the test statistic. The use of a local estimator with $p = 0$ or $p = 1$ (local constant or local linear) does not affect the limiting distribution, given that the equivalent kernel is the same (as stated in Corollary 7.1). Finally, the general complex structure of the asymptotic variance (see also Zhao and Wu (2001) and García-Portugués et al. (2014) for the density context) turns much simpler if a von Mises kernel (see Subsection 7.2.2) is used:

$$\nu^2 = \int_{\Omega_q} \sigma^4(\mathbf{x}) w(\mathbf{x})^2\, \omega_q(d\mathbf{x}) \times (8\pi)^{-\frac{q}{2}}.$$

The contribution of this kernel to the asymptotic bias is $\lambda_q(L^2)\lambda_q(L)^{-2} = \left(2\pi^{\frac{1}{2}}\right)^{-q}$.

The power of the proposed test statistic is also investigated for a family of local Pitman alternatives that is asymptotically close to H_0. Denote these local alternatives by H_{1P}:

$$H_{1P} : m(\mathbf{x}) = m_{\boldsymbol{\theta}_0}(\mathbf{x}) + \left(nh^{\frac{q}{2}}\right)^{-\frac{1}{2}} g(\mathbf{x}), \text{ for all } \mathbf{x} \in \Omega_q,$$

where $m_{\boldsymbol{\theta}_0} \in \mathcal{M}_\Theta$, $g : \Omega_q \longrightarrow \mathbb{R}$ and, by assumption **A4**, $\left(nh^{\frac{q}{2}}\right)^{-\frac{1}{2}} \to 0$. With this notation, H_{1P} becomes H_0 when the component g is such that $m_{\boldsymbol{\theta}_0} + \left(nh^{\frac{q}{2}}\right)^{-\frac{1}{2}} g \in \mathcal{M}_\Theta$ ($g \equiv 0$, for example) and H_1 when the previous statement does not hold for a set of positive probability. The following conditions are required for deriving the limiting distribution of T_n under H_{1P}:

A7. The function g is continuous.

A8. Under H_{1P} , the $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\theta}}$ also satisfies $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$.

Theorem 7.4 (Power under local alternatives)**.** *Under conditions **A1**–**A5**, **A7**–**A8** and under the hypothesis H_{1P},*

$$nh^{\frac{q}{2}} \left(T_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q} \int_{\Omega_q} \sigma^2_{\boldsymbol{\theta}_0}(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \right) \xrightarrow{d} \mathcal{N}\left(\int_{\Omega_q} g(\mathbf{x})^2 f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}), 2\nu^2_{\boldsymbol{\theta}_0} \right).$$

Under local alternatives, the effect of g in the limiting distribution of the test statistic is clearly seen in the asymptotic bias. Specifically, the shift is given by the squared norm of g, weighted with respect to the product of f and w and therefore the test asymptotically detects all kinds of local alternatives from H_0 whose component g has a positive squared weighted norm.

To illustrate the effective convergence of the statistic to the asymptotic distribution, a simple numerical experiment is provided. The regression setting is the model $Y = c + \varepsilon$, with $c = 1$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 = \frac{1}{2}$ and $\mathbf{X}$ uniformly distributed on the circle ($q = 1$). The composite hypothesis $H_0 : m \equiv c$, for $c \in \mathbb{R}$ unknown (test for no effect), is checked using the local constant estimator ($p = 0$) with von Mises kernel and considering the weight function $w \equiv 1$. Figure 7.1 presents two QQ-plots computed from samples $\left\{nh^{\frac{1}{2}}(T_n^j - \frac{\sqrt{\pi}}{4}nh)\right\}_{j=1}^{500}$ obtained for different sample sizes n. Two bandwidth sequences $h_n = \frac{1}{2} \times n^{-r}$, $r = \frac{1}{3}, \frac{1}{5}$ are chosen to illustrate the effect of the bandwidths in the convergence to the asymptotic distribution, and, specifically, that the effect of undersmoothing boosts the convergence since the bias is mitigated. The Kolmogorov-Smirnov (K-S) and Shapiro-Wilk (S-W) tests are applied on to measure the closeness of the empirical distribution of the statistic to a $\mathcal{N}(0, 2\nu^2_{\boldsymbol{\theta}_0})$ and to normality, respectively.

7.3.1 Bootstrap calibration

As it usually happens in smoothed tests (see Härdle and Mammen (1993) or García-Portugués et al. (2014)), the asymptotic distribution cannot be used to calibrate the test statistic for small or moderate sample sizes due to the slow convergence rate and due to the presence of unknown quantities depending on the design density and the error structure. In this situation, bootstrap calibration is an alternative.

The main idea is to approximate the distribution of T_n under H_0 by one of its bootstrapped version T_n^*, which can be arbitrarily well approximated by Monte Carlo by generating bootstrap samples $\{(\mathbf{X}_i, Y_i^*)\}_{i=1}^n$. Under H_0, the bootstrap responses are obtained from the parametric

fit and bootstrap errors that imitate the conditional variance by a wild bootstrap procedure: $Y_i^* = m_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i) + \hat{\varepsilon}_i V_i^*$, where $\hat{\varepsilon}_i = Y_i - m_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i)$ and the variables $V_1^*, \ldots, V_n^*$ are independent from the observed sample and iid with $\mathbb{E}[V_i^*] = 0$, $\mathbb{V}\mathrm{ar}[V_i^*] = 1$ and finite third and fourth moments. A common choice is considering a binary variable with probabilities $\mathbb{P}\{V_i^* = (1-\sqrt{5})/2\} = (5+\sqrt{5})/10$ and $\mathbb{P}\{V_i^* = (1+\sqrt{5})/2\} = (5-\sqrt{5})/10$, which corresponds to the *golden section* bootstrap. The bootstrap test statistic is

$$T_n^* = \int_{\Omega_q} \left(\hat{m}_{h,p}^*(\mathbf{x}) - \mathcal{L}_{h,p} m_{\hat{\boldsymbol{\theta}}^*}(\mathbf{x})\right)^2 \hat{f}_h(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}),$$

where $\hat{m}_{h,p}^*$ and $\hat{\boldsymbol{\theta}}^*$ are the analogues of $\hat{m}_{h,p}$ and $\hat{\boldsymbol{\theta}}$, respectively, obtained from the bootstrapped sample $\{(\mathbf{X}_i, Y_i^*)\}_{i=1}^n$.

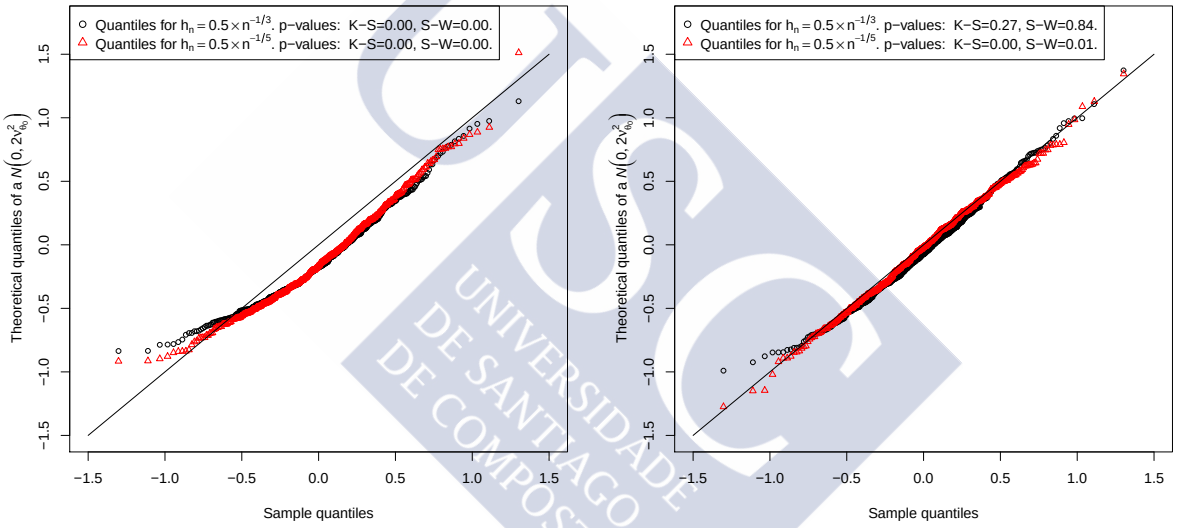


Figure 7.1: QQ-plot comparing the quantiles of the asymptotic distribution given by Theorem 7.3 with the sample quantiles for $\left\{nh^{\frac{1}{2}}\left(T_n^j - \frac{\sqrt{\pi}}{4}nh\right)\right\}_{j=1}^{500}$ with $n = 10^2$ (left) and $n = 5 \times 10^5$ (right).

The testing procedure for calibrating the test is summarized in the next algorithm, stated for the composite hypothesis. If the simple hypothesis is considered, then set $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^*$.

Algorithm 7.1 (Test in practice)**.** *Consider $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$ a random sample from model (7.3). To test $H_0 : m \in \mathcal{M}_{\boldsymbol{\Theta}}$, set a bandwidth h and a weight function w and proceed as follows:*

i. Compute $\hat{\boldsymbol{\theta}}$ and obtain the fitted residuals $\hat{\varepsilon}_i = Y_i - m_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i)$, $i = 1, \ldots, n$.

ii. Compute $T_n = \int_{\Omega_q} \left(\hat{m}_{h,p}(\mathbf{x}) - \mathcal{L}_{h,p} m_{\hat{\boldsymbol{\theta}}}(\mathbf{x})\right)^2 \hat{f}_h(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x})$.

iii. Bootstrap resampling. For $b = 1, \ldots, B$:

(a) Obtain a bootstrap random sample $\{(\mathbf{X}_i, Y_i^*)\}_{i=1}^n$, where $Y_i^* = m_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i) + \hat{\varepsilon}_i V_i^*$ and V_i^* are iid golden section binary variables, $i = 1, \ldots, n$.

(b) Compute $\hat{\boldsymbol{\theta}}^*$ as in i), but now from the bootstrap sample from a).

(c) Compute $T_n^{*b} = \int_{\Omega_q} \left(\hat{m}_{h,p}^*(\mathbf{x}) - \mathcal{L}_{h,p} m_{\hat{\boldsymbol{\theta}}^*}(\mathbf{x})\right)^2 \hat{f}_h(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x})$.

iv. Approximate the p-value by $\frac{1}{B}\sum_{b=1}^{B}\mathbf{1}_{\{T_n^{*b}\leq T_n\}}$.

Bootstrap strategies may be computational expensive. In this case, it should be noted that the test statistic can be written as $T_n = \int_{\Omega_q}\left(\sum_{i=1}^n W_n^p(\mathbf{x},\mathbf{X}_i)\hat{\varepsilon}_i\right)^2\hat{f}_h(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})$, using the equivalent kernel notation. The bootstrap test statistic T_n^* is the same, just taking $\hat{\varepsilon_i}^* = Y_i^* - m_{\hat{\boldsymbol{\theta}}^*}(\mathbf{X}_i)$ instead of $\hat{\varepsilon}_i$, so there is no need to recompute the other elements of T_n in the bootstrap resampling.

In order to prove the consistency of the resampling mechanism detailed in Algorithm 7.1, that is, that the bootstrapped statistic T_n^* has the same asymptotic distribution as the original statistic T_n, a bootstrap analogue of assumption **A6** is required:

A9. The estimator $\hat{\boldsymbol{\theta}}^*$ computed from $\{(\mathbf{X}_i, Y_i^*)\}_{i=1}^n$ is such that $\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}} = \mathcal{O}_{\mathbb{P}^*}(n^{-\frac{1}{2}})$, where $\mathbb{P}^*$ is the probability law conditional on $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$.

Based on this assumption and on the previous ones, it is proved from Theorem 7.3 that the probability distribution function (pdf) of T_n^* conditional on the sample converges always in probability to a Gaussian pdf, which is the same asymptotic pdf of T_n if H_0 holds.

Theorem 7.5 (Bootstrap consistency)**.** *Under conditions* ***A1***–***A6*** *and* ***A9*** *and conditionally on* $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$,

$$nh^{\frac{q}{2}}\left(T_n^* - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2_{\boldsymbol{\theta}_1}(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right)\xrightarrow{d}\mathcal{N}\left(0, 2\nu^2_{\boldsymbol{\theta}_1}\right)$$

in probability. If the null hypothesis holds, then $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$.

7.4 Simulation study

The finite sample performance of the goodness-of-fit test is explored in four regression models, considering different sample sizes, dimensions and bandwidths. Given the regression model (7.3) and taking T_n as test statistic, the following components must be specified: the density of the predictor $\mathbf{X}$, the regression function m, the noise $\sigma(\mathbf{X})\varepsilon$ and the deviations from H_0.

The parametric regression functions with directional covariate and scalar response are shown in Figure 7.2, with the following codification: the radius from the origin represents the response $m(\mathbf{x})$ for a $\mathbf{x}$ direction, resulting in a distortion from a perfect circle or sphere. The noise considered is $\varepsilon\sim\mathcal{N}(0,1)$, with two different conditional standard deviations given by $\sigma_1(\mathbf{x}) = \frac{1}{2}$ (homocedastic) and $\sigma_2(\mathbf{x}) = \frac{1}{4} + 3f_{\text{M16}}(\mathbf{x})$ (heteroskedastic), with f_{M16} being the density of the M16 model in García-Portugués (2013). In order to define the design densities, some models introduced by García-Portugués (2013) have been considered: M1 (uniform), M4, M12 and M20 are used as single densities or as part of mixture distributions, as in S2 and S3 (see Table 7.1). The alternative hypothesis H_1 is obtained by adding the deviations $\Delta_1(\mathbf{x}) = \cos(2\pi x_1)(x_{q+1}^3 - 1)/\log(2+|x_{q+1}|)$ and $\Delta_2(\mathbf{x}) = \cos(2\pi x_1^2 x_2)\exp\{x_{q+1}\}$ to the true regression function $m_{\boldsymbol{\theta}_0}(\mathbf{x})$. The different combinations considered in S1 to S4 are given in Table 7.1 (see supplementary material also for further details).

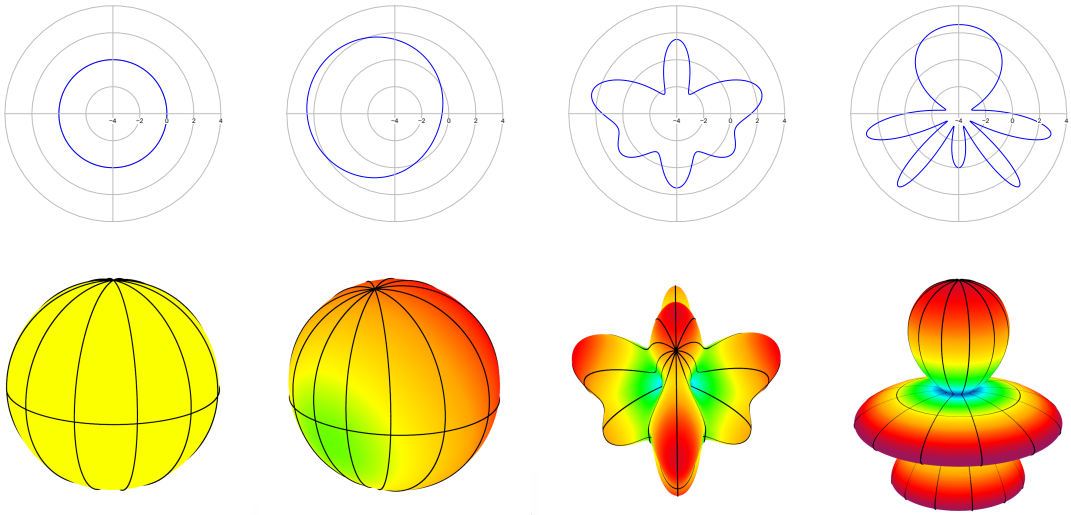


Figure 7.2: Parametric regression models for scenarios S1 to S4, for circular and spherical cases.

Scenario	Regression function	Parameters	Density	Noise	Deviation
S1	$m(\mathbf{x}) = c$	$c = 0$	M1	Het.	$\frac{3}{4}\Delta_1(\mathbf{x})$
S2	$m(\mathbf{x}) = c + \boldsymbol{\eta}^T\mathbf{x}$	$c = 1$, $\boldsymbol{\eta} = \left(-\frac{3}{2}, \frac{\mathbf{1}}{\mathbf{2}_q}\right)$	$\frac{3}{5}$M4 + $\frac{2}{5}$M1	Het.	$-\frac{3}{4}\Delta_1(\mathbf{x})$
S3	$m(\mathbf{x}) = c + a\sin(2\pi x_2) + b\cos(2\pi x_1)$	$c = 0$, $a = 1$, $b = \frac{3}{2}$	$\frac{3}{5}$M12 + $\frac{2}{5}$M1	Hom.	$\frac{3}{4}\Delta_2(\mathbf{x})$
S4	$m(\mathbf{x}) = c + a\sin\left(2\pi b\left(2 + x_{q+1}\right)^{-1}\right)$	$c = 0$, $a = 3$, $b = 4$	M20	Hom.	$\frac{1}{2}\Delta_2(\mathbf{x})$

Table 7.1: Specification of simulation scenarios for model (7.3).

The tests based on the projected local constant and local linear estimators ($p = 0, 1$) are compared in these four scenarios, under H_0 and H_1, for a grid of bandwidths, different sample sizes $n = 100, 250, 500$ and dimensions $q = 1, 2, 3$. $M = 1000$ Monte Carlo trials and $B = 1000$ bootstrap replicates are considered. Parametric estimation is done by nonlinear least squares, which is justified by their simplicity and asymptotic normality under certain conditions (Jennrich, 1969), hence satisfying assumption **A6**.

The empirical sizes of the goodness-of-fit tests are shown using the so called *significance trace* (Bowman and Azzalini, 1997), that is, the curve of percentages of empirical rejections for different bandwidths. These empirical rejections are computed from the same generated samples and bootstrap replicates. As it is shown in Figure 7.3, except for very small bandwidths that result in a conservative test, the significance level is stabilized around the 95% confidence band for the nominal level $\alpha = 0.05$, for the different scenarios, dimensions and sample sizes. With respect to the power, given the mild deviations from the null hypotheses (see supplementary material for quantification), the power performance of the proposed tests seems quite competitive. Despite the fact that the test based on the local linear estimator ($p = 1$) provides a better power for large bandwidths in certain scenarios, the overall impression is that the test with $p = 0$ is hard to beat: the powers with $p = 0$ and $p = 1$ are almost the same for low dimensions, whereas as the dimension increases the local constant estimator performs better for a wider range of

bandwidths. This effect could be explained by the spikes that local linear regression tends to show in the boundaries of the support (design densities of S3 and S4), which become more important as the dimension increases. More simulation results for different sample sizes and significance levels are available in the supplement.

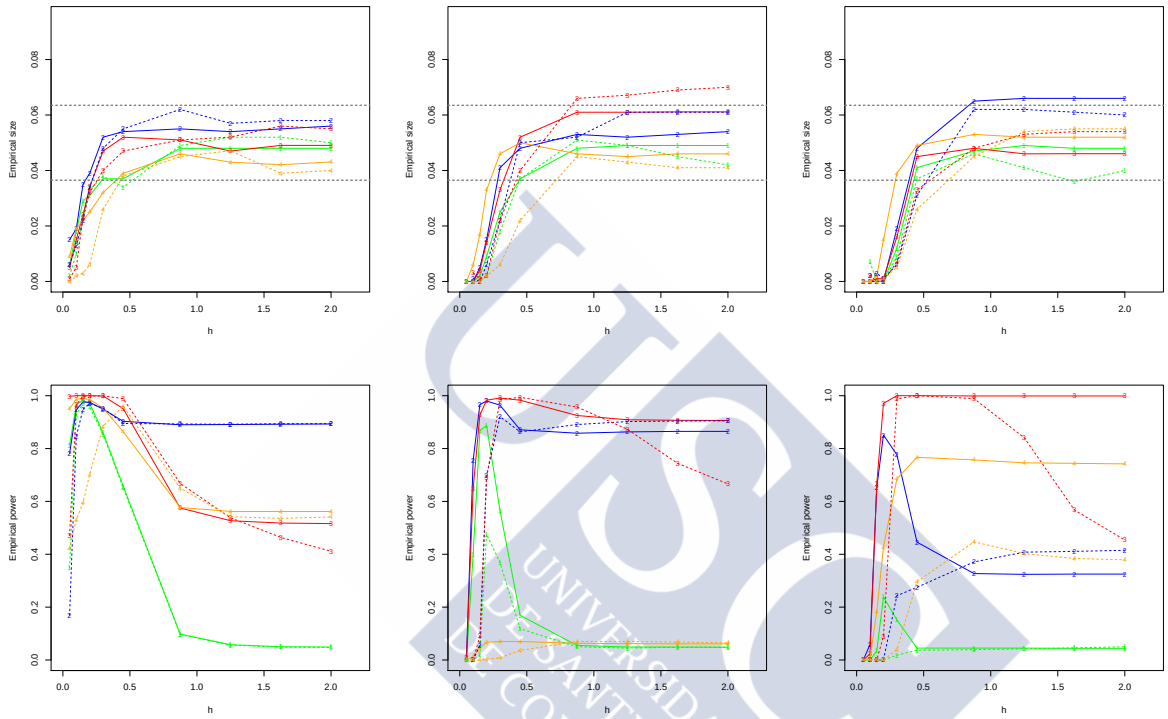


Figure 7.3: Empirical sizes (first row) and powers (second row) for significance level $\alpha = 0.05$ for the different scenarios, with $p = 0$ (solid line) and $p = 1$ (dashed line). From left to right, columns represent dimensions $q = 1, 2, 3$ with sample size $n = 100$.

7.5 Application to text mining

A challenging field where directional data techniques may be applied is text mining. In different applications within this context, it is quite common to consider a *corpus* (collection of documents) and to determine the so-called vector space model: a corpus $\mathbf{d}_1, \ldots, \mathbf{d}_n$ is codified by the set of vectors $\{(d_{i1}, \ldots, d_{iD})\}_{i=1}^n$ (the *document-term matrix*) with respect to a dictionary (or a *bag of words*) $\{w_1, \ldots, w_D\}$, such that d_{ij} represents the frequency of the dictionary's j-th word in the document d_i. Since large documents are expected to have higher word frequencies, a normalization is required. For instance, if the Euclidean norm is used, $\mathbf{d}_i / ||\mathbf{d}_i||$, then the documents can be regarded as points in Ω_{D-1} providing therefore a set of directional data. Some recent references using directional statistics in text mining are Banerjee et al. (2005), Buchta et al. (2012) and Surian and Chawla (2013).

In this example, the corpus that is analyzed was acquired from the news aggregator *Slashdot* (`wwww.slashdot.org`). This website publishes summaries of news about technology and science that are submitted and evaluated by users. Each news entry includes a title, a summary with

links to other related news and a discussion thread gathering users comments. Obviously, not all the news have the same impact in terms of popularity (or participation), measuring this variable as the number of comments for each entry. The goal of this application is to test a linear model that takes as a predictor the topic of the news (a directional variable) and as a response the log-number of comments. The consideration of simple linear models seems frequent in this context. For instance, Tatar et al. (2012) consider linear regression models for providing a ranking on online news based on the number of comments at two different time moments after the news publication. Asur and Huberman (2010) present simple linear models for predicting the box-office of a film from tweets information, and show that a simple linear regression performs better in predicting the box-office than artificial money markets. Also in favor of linear models, it can be also argued that in text classifications, it has been checked that non-linear classifiers hardly provide any advantage with respect to linear ones (Joachims, 2002).

Titles, summaries and number of comments in each news appeared in 2013 were downloaded, resulting in a collection of $n = 8121$ documents. After that, the next steps were performed with the help of the text mining `R` library `tm` (Meyer et al., 2008): 1) merge titles and summaries in the same document, omitting user submission details; 2) deletion of HTML codes; 3) conversion to lowercase; 4) deletion of stop words (defined in `tm` and `MySQL`), punctuation, white spaces and numbers; 5) stemming of words. The distribution of the *document frequency* (df, number of documents containing a particular word) is highly right skewed and more than 50% of the processed words only appeared in a single document, while in contrast a few words are repeated in many documents. To overcome this problem, a pruning was done such that only the words with df between quantiles 95% and 99.95% were considered (words appearing within 58 and 1096 documents). After this process, the documents are represented in a document term matrix formed by the $D = 1508$ words.

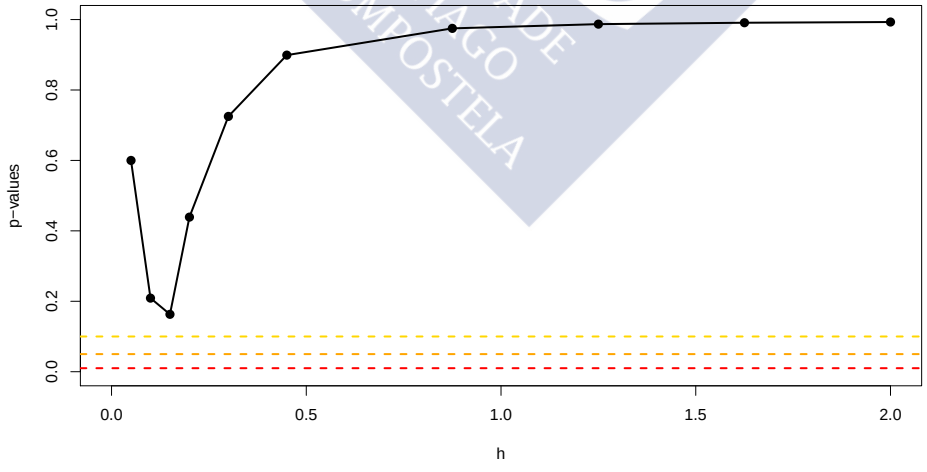


Figure 7.4: Significance trace of the local constant goodness-of-fit test for the constrained linear model.

In order to construct a plausible linear model, a preliminary variable selection was performed using LASSO regression with (tuning) parameter λ selected by an overpenalized *three* standard error rule (see Hastie et al. (2009)). After removing some extra variables by using a backward stepwise method with BIC, we obtain a fitted vector $\hat{\boldsymbol{\eta}} \in \mathbb{R}^D$ with $d = 77$ non-zero entries. The test is applied to check the null hypothesis of a candidate linear model with coefficient $\boldsymbol{\eta}$ constrained to be zero except in these previously selected d words, that is $H_0 : m(\mathbf{x}) = c + \boldsymbol{\eta}^T \mathbf{x}$,

with $\boldsymbol{\eta}$ subject to $\mathbf{A}\boldsymbol{\eta} = \mathbf{0}$ for an adequate choice of the matrix $\mathbf{A}_{(D-d)\times D}$. The significance trace in Figure 7.4 shows no evidence to reject the linear model for a wide grid of bandwidths, using a local constant approach (local linear was not implemented due to its higher cost and computational limitations). Table 7.2 shows the fitted linear model under the null hypothesis. As it can be seen, news where stemmed words like "kill", "climat", "polit" appear have a strong positive impact on the number of comments, since these news are likely more controversial and generate broader discussions. On the other hand, scientific related words like "mission", "abstract" or "lab" have a negative impact, since they tend to raise more objective and higher specific discussions.

(int)	conclud	gun	kill	refus	averag	lose	obama	declin	climat	snowden	stop	wrong
4.97	2.56	2.13	1.86	1.77	1.74	1.72	1.68	1.63	1.53	1.44	1.43	1.35
war	polit	senat	tesla	violat	concern	slashdot	ban	reason	health	pay	window	american
1.34	1.31	1.27	1.26	1.25	1.22	1.22	1.19	1.15	1.14	1.14	1.12	1.10
told	worker	man	comment	state	think	movi	ask	job	drive	know	problem	employe
1.10	1.09	1.09	1.04	1.00	0.97	0.96	0.95	0.94	0.91	0.87	0.87	0.87
nsa	charg	feder	money	sale	need	microsoft	project	network	cell	imag	avail	video
0.84	0.80	0.80	0.80	0.78	0.76	0.52	−0.46	−0.51	−0.69	−0.70	−0.73	−0.78
process	data	materi	nasa	launch	electron	robot	satellit	detect	planet	help	cloud	hack
−0.81	−0.82	−0.88	−0.92	−0.92	−0.94	−0.95	−0.96	−1.04	−1.06	−1.06	−1.08	−1.10
open	lab	mobil	techniqu	vulner	mission	team	supercomput	abstract	simul	demo	guid	
−1.15	−1.15	−1.16	−1.17	−1.21	−1.23	−1.50	−1.89	−1.97	−1.99	−2.01	−2.02	

Table 7.2: Fitted constrained linear model on the Slashdot dataset, with $R^2 = 0.25$. The significances of each coefficient are lower than 0.002.

Acknowledgments

We would like to thank professor David E. Losada, from CiTIUS at the University of Santiago de Compostela, for his guidance in the data application. This research was supported by Project MTM2008–03010 from the Spanish Ministry of Science and Innovation, by Project 10MDS207015PR from Dirección Xeral de I+D of the Xunta de Galicia, by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy), by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013) / ERC Grant agreement No. 203650, and by the contract "Projet d'Actions de Recherche Concertées" (ARC) 11/16–039 of the "Communauté française de Belgique" (granted by the "Académie universitaire Louvain"). Work of E. García-Portugués has been supported by a grant from Fundación Barrié and FPU grant AP2010–0957 from the Spanish Ministry of Education. Authors gratefully acknowledge the computational resources used at the SVG cluster of the CESGA Supercomputing Center.

Supplement

Three extra appendices are included as supplementary material, containing particular cases of the projected local estimator, the technical lemmas and further results for the simulation study.

7.A Main results

7.A.1 Projected local estimator properties

Proof of Theorem 7.1. The proof is divided in three sections: the conditional bias is first obtained for $p = 1$, then the result for $p = 0$ follows by restricting the computations to the first column of $\boldsymbol{\mathcal{X}}_{\mathbf{x},p}$ and the variance is proved to be common to both estimators.

Bias of $\hat{m}_{h,1}$. Working conditionally, by (7.3) and (7.5),

$$\mathbb{E}\left[\hat{m}_{h,p}(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n\right] = \mathbf{e}_1^T\left(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\mathbf{m}, \tag{7.7}$$

where $\mathbf{m} = (m(\mathbf{X}_1),\ldots,m(\mathbf{X}_n))^T$. The proof is based on Theorem 2.1 in Ruppert and Wand (1994) but adapted to the projected local estimator. First of all, consider the Taylor expansion of $m(\mathbf{X}_i)$ of second order around the point $\mathbf{x} \in \Omega_q$, which follows naturally by extending the one given in Section 7.2 (since $\mathbf{x}^T\boldsymbol{\mathcal{H}}_m(\mathbf{x})\mathbf{x} = 0$, where $\boldsymbol{\mathcal{H}}_m(\mathbf{x})$ is the Hessian of m):

$$\begin{aligned} m(\mathbf{X}_i) =& m(\mathbf{x}) + \boldsymbol{\nabla}m(\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x}) + (\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\mathcal{H}}_m(\mathbf{x})\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x}) \\ &+ o\left(||\mathbf{X}_i-\mathbf{x}||^2\right). \end{aligned}$$

The Taylor expansion can be expressed componentwise (also for the orders) as

$$\mathbf{m} = \boldsymbol{\mathcal{X}}_{\mathbf{x},1}\left(m(\mathbf{x}),\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\nabla}m(\mathbf{x})\right)^T + \frac{1}{2}\mathbf{Q}_m(\mathbf{x}) + o\left(\mathbf{R}_m(\mathbf{x})\right),$$

with $\mathbf{Q}_m(\mathbf{x})$ the vector with i-th entry given by $(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\mathcal{H}}_m(\mathbf{x})\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})$ and remainder term of order $\mathbf{R}_m(\mathbf{x}) = \left(||\mathbf{X}_1-\mathbf{x}||^2,\ldots,||\mathbf{X}_n-\mathbf{x}||^2\right)^T$, uniform in $\mathbf{x} \in \Omega_q$ since the third derivative of m is bounded by assumption **A1**. Then, by (7.7), the first term in the Taylor expansion is

$$\mathbf{e}_1^T\left(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},1}\left(m(\mathbf{x}),\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\nabla}m(\mathbf{x})\right)^T, \tag{7.8}$$

which for $p = 1$ equals $m(\mathbf{x})$ and hence the conditional bias is given by $\mathbf{e}_1^T(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p})^{-1}$ $\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T$ times the remaining vector. By using the results *i*, *ii* and *iv* of Lemma B.1, it follows that, componentwise,

$$\begin{aligned} &n^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p} \\ &\quad= \frac{1}{n}\sum_{i=1}^n\begin{pmatrix} L_h(\mathbf{x},\mathbf{X}_i) & L_h(\mathbf{x},\mathbf{X}_i)(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}} \\ L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x}) & L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}} \end{pmatrix} \\ &\quad= \begin{pmatrix} f(\mathbf{x}) & \frac{2b_q(L)}{q}\boldsymbol{\nabla}f(\mathbf{x})^T\mathbf{B}_{\mathbf{x}}h^2 \\ \frac{2b_q(L)}{q}\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\nabla}f(\mathbf{x})h^2 & \frac{2b_q(L)}{q}\mathbf{I}_q f(\mathbf{x})h^2 \end{pmatrix} + o_{\mathbb{P}}\left(\mathbf{1}\mathbf{1}^T\right). \end{aligned} \tag{7.9}$$

This matrix can be inverted by the inversion formula of a block matrix, resulting in

$$\left(n^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1} = \begin{pmatrix} f(\mathbf{x})^{-1} & -f(\mathbf{x})^{-2}\boldsymbol{\nabla}f(\mathbf{x})^T\mathbf{B}_{\mathbf{x}} \\ -f(\mathbf{x})^{-2}\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\nabla}f(\mathbf{x}) & \left(\frac{2b_q(L)}{q}f(\mathbf{x})h^2\right)^{-1}\mathbf{I}_q \end{pmatrix} + o_{\mathbb{P}}\left(\mathbf{1}\mathbf{1}^T\right). \tag{7.10}$$

Now the quadratic term of the Taylor expansion yields by results *v–vi* of Lemma B.1:

$$\begin{aligned} & n^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\mathbf{Q}_m(\mathbf{x}) \\ &\quad = \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} L_h(\mathbf{x},\mathbf{X}_i)(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\mathcal{H}}_m(\mathbf{x})\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x}) \\ L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\mathcal{H}}_m(\mathbf{x})\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x}) \end{pmatrix} \\ &\quad = \begin{pmatrix} \frac{2b_q(L)}{q}\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_m(\mathbf{x})\right]f(\mathbf{x})h^2 + o_{\mathbb{P}}\left(h^2\right) \\ o_{\mathbb{P}}\left(h^3\mathbf{1}\right) \end{pmatrix}. \end{aligned} \tag{7.11}$$

Finally, the remaining order is $o_{\mathbb{P}}\left(h^2\right)$, because

$$\begin{aligned} & \mathbf{e}_1^T\left(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\mathbf{R}_m(\mathbf{x}) \\ &\quad = \left(f(\mathbf{x})^{-1} + o_{\mathbb{P}}\left(1\right), -f(\mathbf{x})^{-2}\boldsymbol{\nabla} f(\mathbf{x})^T\mathbf{B}_{\mathbf{x}} + o_{\mathbb{P}}\left(\mathbf{1}^T\right)\right)\left(\mathcal{O}_{\mathbb{P}}\left(h^2\right), o_{\mathbb{P}}\left(h^3\mathbf{1}^T\right)\right)^T, \end{aligned}$$

by setting $\boldsymbol{\mathcal{H}}_m(\mathbf{x}) \equiv \mathbf{I}_{q+1}$ and using *v–vi* from Lemma B.1. Joining (7.10) and (7.11), then

$$\mathbb{E}\left[\hat{m}_{h,1}(\mathbf{x}) - m(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n\right] = \frac{b_q(L)}{q}\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_m(\mathbf{x})\right]h^2 + o_{\mathbb{P}}\left(h^2\right).$$

Bias of $\hat{m}_{h,0}$. For the case $p=0$ the product in (7.8) is not $m(\mathbf{x})$ but slightly different. By (7.10), $\mathbf{e}_1^T\left(n^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},0}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},0}\right)^{-1} = \left(f(\mathbf{x}) + o_{\mathbb{P}}\left(1\right)\right)^{-1}$ and also by (7.9) and *i–ii* in Lemma B.1,

$$n^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},0}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},1} = \left(f(\mathbf{x}) + o_{\mathbb{P}}\left(1\right), \frac{2b_q(L)}{q}\boldsymbol{\nabla} f(\mathbf{x})^T\mathbf{B}_{\mathbf{x}}h^2 + o_{\mathbb{P}}\left(h^2\mathbf{1}^T\right) + \mathcal{O}_{\mathbb{P}}\left(\frac{h}{\sqrt{nh^q}}\mathbf{1}^T\right)\right).$$

Then, (7.8) turns into $m(\mathbf{x}) + \frac{2b_q(L)}{q}\frac{\boldsymbol{\nabla} f(\mathbf{x})^T\boldsymbol{\nabla} m(\mathbf{x})}{f(\mathbf{x})}h^2 + o_{\mathbb{P}}(h^2\mathbf{1}^T) + \mathcal{O}_{\mathbb{P}}\left(\frac{h}{\sqrt{nh^q}}\mathbf{1}^T\right)$ because the coefficient in $m(\mathbf{x})$ is *exactly* one. Adding this to the bias of $\hat{m}_{h,1}$, the result follows since the contribution of the linear part in (7.11) and in the remaining order is negligible.

Variance of $\hat{m}_{h,p}$. By the variance property for linear combinations,

$$\mathbb{V}\mathrm{ar}\left[\hat{m}_{h,p}(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n\right] = \mathbf{e}_1^T\left(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\mathbf{V}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}\left(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1}\mathbf{e}_1,$$

where $\mathbf{V} = \mathrm{diag}\left(\sigma^2(\mathbf{X}_1),\ldots,\sigma^2(\mathbf{X}_n)\right)^T$. By results *vii–ix* of Lemma B.1,

$$\begin{aligned} & n^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\mathbf{V}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p} \\ &\quad = \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} L_h^2(\mathbf{x},\mathbf{X}_i)\sigma^2(\mathbf{X}_i) & L_h^2(\mathbf{x},\mathbf{X}_i)(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\sigma^2(\mathbf{X}_i) \\ L_h^2(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})\sigma^2(\mathbf{X}_i) & L_h^2(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\sigma^2(\mathbf{X}_i) \end{pmatrix} \\ &\quad = \begin{pmatrix} \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{h^q}\sigma^2(\mathbf{x})f(\mathbf{x}) & \mathbf{0}^T \\ \mathbf{0} & \mathbf{0}\mathbf{0}^T \end{pmatrix} + o_{\mathbb{P}}\left(h^{-q}\mathbf{1}\mathbf{1}^T\right). \end{aligned} \tag{7.12}$$

Therefore, by (7.10) and (7.12), the common variance expression follows. □

Proof of Corollary 7.1. Note that $W_n^p(\mathbf{x},\mathbf{X}_i) = \mathbf{e}_1^T(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p})^{-1}(1,\delta_{p,1}(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}})^T L_h(\mathbf{x},\mathbf{X}_i)$. Then, by expression (7.10), uniformly in $\mathbf{x}\in\Omega_q$ it follows that

$$\hat{m}_{h,p}(\mathbf{x}) = \frac{1}{nf(\mathbf{x})}\sum_{i=1}^n L_h(\mathbf{x},\mathbf{X}_i)Y_i(1+o_{\mathbb{P}}(1))$$

$$+\frac{\delta_{p,1}\boldsymbol{\nabla} f(\mathbf{x})^T\mathbf{B}_{\mathbf{x}}}{f(\mathbf{x})^2}\frac{1}{n}\sum_{i=1}^{n}L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})Y_i(1+o_{\mathbb{P}}(1)).$$

By (7.1) and (7.2), the first addend is $\frac{1}{nh^q\lambda_q(L)f(\mathbf{x})}\sum_{i=1}^{n}L_h(\mathbf{x},\mathbf{X}_i)Y_i(1+o_{\mathbb{P}}(1)$. The second term is $o_{\mathbb{P}}(1)$ (see *iii* in Lemma B.1) and negligible in comparison with the first one, which is $\mathcal{O}_{\mathbb{P}}(1)$. Then, it can be absorbed inside the factor $(1+o_{\mathbb{P}}(1))$, proving the corollary. □

Proof of Theorem 7.2. For a fixed $\mathbf{x}\in\Omega_q$, the next decomposition is studied:

$$\begin{aligned}\sqrt{nh^q}(\hat{m}_{h,p}(\mathbf{x})-m(\mathbf{x})) &= \sqrt{nh^q}\left(\hat{m}_{h,p}(\mathbf{x})-\mathbb{E}\left[\hat{m}_{h,p}(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n\right]\right)\\ &\quad+\sqrt{nh^q}\left(\mathbb{E}\left[\hat{m}_{h,p}(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n\right]-m(\mathbf{x})\right)\\ &= N_1+N_2.\end{aligned}$$

Term N_1. From the proof of Theorem 7.1, $N_1=\sqrt{nh^q}\mathbf{e}_1^T\left(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}(\mathbf{Y}-\mathbf{m})$. By the Cramér-Wold device, if $n^{-1}\mathbf{a}^T\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}(\mathbf{Y}-\mathbf{m})=n^{-1}\sum_{i=1}^{n}V_{n,i}=\bar{V}_n$ is asymptotically normal for any $\mathbf{a}\in\mathbb{R}^{pq+1}$, then $n^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}(\mathbf{Y}-\mathbf{m})$ is also asymptotically normal. To obtain the asymptotic normality of $\bar{V}_n$ the Lyapunov's Central Limit Theorem (CLT) for triangular arrays $\{V_{n,i}\}_{i=1}^n$ is employed, this is: if for $\delta>0$

$$\lim_{n\to\infty}\left(n^{\frac{\delta}{2}}\mathbb{V}\mathrm{ar}\left[V_n\right]^{1+\frac{\delta}{2}}\right)^{-1}\mathbb{E}\left[\left|V_n-\mathbb{E}\left[V_n\right]\right|^{2+\delta}\right]=0,$$

where $V_n=L_h(\mathbf{x},\mathbf{X})(\mathbf{Y}-m(\mathbf{X}))\mathbf{a}^T(1,\delta_{p,1}\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}-\mathbf{x}))$, then $\sqrt{n}\frac{\bar{V}_n-\mathbb{E}[V_n]}{\sqrt{\mathbb{V}\mathrm{ar}[V_n]}}\xrightarrow{d}\mathcal{N}(0,1)$. From $\mathbb{E}\left[\,|V_n-\mathbb{E}\left[V_n\right]|^{2+\delta}\,\right]=\mathcal{O}(\mathbb{E}\left[\,|V_n|^{2+\delta}\,\right])$ and the use of Lemma B.6, it holds that:

$$\begin{aligned}\mathbb{E}\left[|V_n|^{2+\delta}\right] &= \frac{c_{h,q}(L)^{2+\delta}}{c_{h,q}(L^{2+\delta})}\int_{\mathbb{R}}\left((y-m(\mathbf{x}))\mathbf{a}^T\left(1,\delta_{p,1}\mathbf{B}_{\mathbf{x}}^T(\mathbf{x}-\mathbf{x})\right)\right)^{2+\delta}f_{\mathbf{X},Y}(\mathbf{x},y)\,dy\,(1+o(1))\\ &= \frac{\lambda_q(L^{2+\delta})a_1^{2+\delta}}{\lambda_q(L)^{2+\delta}h^{(1+\delta)q}}f(\mathbf{x})\mathbb{E}\left[(Y-m(\mathbf{x}))^{2+\delta}|\mathbf{X}=\mathbf{x}\right](1+o(1))\\ &= \mathcal{O}\left(h^{-(1+\delta)q}\right).\end{aligned}$$

Note that $\mathbb{E}\left[(Y-m(\mathbf{x}))^{2+\delta}|\mathbf{X}=\mathbf{x}\right]<\infty$ is required for a $\delta>0$ and that by **A3** the kernel $L^{2+\delta}$ plays the same role as L. By using this result with $\delta=0$, it follows that $\mathbb{V}\mathrm{ar}\left[V_n\right]\leq\mathbb{E}\left[V_n^2\right]=\mathcal{O}\left(h^{-q}\right)$. Therefore,

$$\frac{\mathbb{E}\left[\left|V_n-\mathbb{E}\left[V_n\right]\right|^{2+\delta}\right]}{n^{\frac{\delta}{2}}\mathbb{V}\mathrm{ar}\left[V_n\right]^{1+\frac{\delta}{2}}}=\mathcal{O}\left(\frac{h^{-(1+\delta)q}}{n^{\frac{\delta}{2}}h^{-(1+\frac{\delta}{2})q}}\right)=\mathcal{O}\left((nh^q)^{-\frac{\delta}{2}}\right),$$

so by **A4** and the Cramér-Wold device $\sqrt{nh^q}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}(\mathbf{Y}-\mathbf{m})\xrightarrow{d}\mathcal{N}_{pq+1}(\mathbf{0},\boldsymbol{\Sigma})$, with the covariance matrix arising from (7.12):

$$\boldsymbol{\Sigma}=\begin{pmatrix}\lambda_q(L^2)\lambda_q(L)^{-2}\sigma^2(\mathbf{x})f(\mathbf{x}) & \mathbf{0}^T\\ \mathbf{0} & \mathbf{0}\mathbf{0}^T\end{pmatrix}.$$

On the other hand, by (7.10), $\mathbf{e}_1^T\left(n^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1}$ converges in probability to $(f(\mathbf{x})^{-1},-f(\mathbf{x})^{-2}\boldsymbol{\nabla}f(\mathbf{x})^T\mathbf{B}_{\mathbf{x}})$ if $p=1$ and to $f(\mathbf{x})$ if $p=0$. The desired result then follows by the use of

Slutsky's theorem:

$$N_1 = \sqrt{nh^q}\mathbf{e}_1^T \left(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}\right)^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \left(\mathbf{Y} - \mathbf{m}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{f(\mathbf{x})}\sigma^2(\mathbf{x})\right).$$

Term N_2. By the conditional bias expansion of Theorem 7.1 N_2 converges in probability as

$$N_2 = \sqrt{nh^q}\frac{b_q(L)}{q}B_q(\mathbf{x})h^2(1 + o_{\mathbb{P}}(1)) + \delta_{p,0}\, o_{\mathbb{P}}(1),$$

so adding this bias to N_1 the asymptotic normality is proved by Slutsky's theorem. □

7.A.2 Asymptotic results for the goodness-of-fit test

Proof of Theorems 7.3 and 7.4. Both theorems are proved at the same time by assuming that H_{1P} holds and considering H_0 a particular case with $g \equiv 0$. The proof follows the steps of Härdle and Mammen (1993) and Alcalá et al. (1999) and makes use of the equivalent kernel representation for simplifying the computations and applying de Jong (1987)'s CLT. The test statistic T_n can be separated into three addends by adding and subtracting the true smoothed regression function:

$$\begin{aligned} T_n &= \int_{\Omega_q} \left(\hat{m}_{h,p}(\mathbf{x}) - \mathcal{L}_{h,p} m_{\hat{\boldsymbol{\theta}}}(\mathbf{x})\right)^2 \hat{f}_h(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \\ &= (T_{n,1} + T_{n,2} - 2T_{n,3})(1 + o_{\mathbb{P}}(1)), \end{aligned} \tag{7.13}$$

where, thanks to result i from Lemma B.1, the addends are:

$$\begin{aligned} T_{n,1} &= \int_{\Omega_q} \left(\sum_{i=1}^n W_n^p(\mathbf{x}, \mathbf{X}_i)\left(Y_i - m_{\boldsymbol{\theta}_0}(\mathbf{X}_i)\right)\right)^2 f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}), \\ T_{n,2} &= \int_{\Omega_q} \left(\mathcal{L}_{h,p}\left(m_{\boldsymbol{\theta}_0} - m_{\hat{\boldsymbol{\theta}}}\right)(\mathbf{x})\right)^2 f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}), \\ T_{n,3} &= \int_{\Omega_q} \left(\hat{m}_{h,p}(\mathbf{x}) - \mathcal{L}_{h,p} m_{\boldsymbol{\theta}_0}(\mathbf{x})\right) \mathcal{L}_{h,p}\left(m_{\boldsymbol{\theta}_0} - m_{\hat{\boldsymbol{\theta}}}\right)(\mathbf{x}) f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}). \end{aligned}$$

By Slutsky's theorem, the asymptotic distribution of T_n will be the one of $T_{n,1} + T_{n,2} - 2T_{n,3}$, so the proof is divided in the examination of each addend.

Terms $T_{n,2}$ and $T_{n,3}$. By a Taylor expansion on $m_{\boldsymbol{\theta}}(\mathbf{x})$ as a function of $\boldsymbol{\theta}$ (see **A5**),

$$\begin{aligned} T_{n,2} &= \int_{\Omega_q} \left(\mathcal{L}_{h,p}\left(\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^T \frac{\partial m_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n}\right)(\mathbf{x})\right)^2 f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \\ &= \int_{\Omega_q} \left(\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^T \mathcal{L}_{h,p}\left(\mathcal{O}_{\mathbb{P}}(\mathbf{1})\right)(\mathbf{x})\right)^2 f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \\ &= \mathcal{O}_{\mathbb{P}}\left(n^{-1}\right), \end{aligned}$$

with $\boldsymbol{\theta}_n \in \Theta$. The second equality holds by the boundedness of $\frac{\partial m_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \boldsymbol{\theta}}$ for $\mathbf{x} \in \Omega_q$, where the last holds by **A6** and **A8**. On the other hand,

$$T_{n,3} = (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})^T \int_{\Omega_q} \left(\hat{m}_{h,p}(\mathbf{x}) - \mathcal{L}_{h,p} m_{\boldsymbol{\theta}_0}(\mathbf{x})\right) \mathcal{L}_{h,p}\left(\frac{\partial m_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n}\right)(\mathbf{x}) f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x})$$

$$
\begin{aligned}
&= \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}}) \int_{\Omega_q} \left(\hat{m}_{h,p}(\mathbf{x}) - \mathcal{L}_{h,p} m_{\boldsymbol{\theta}_0}(\mathbf{x})\right) f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \\
&= \mathcal{O}_{\mathbb{P}}(n^{-1}),
\end{aligned}
$$

because of the previous considerations used and i from Lemma B.2. As a consequence, $nh^{\frac{q}{2}} T_{n,2} = \mathcal{O}_{\mathbb{P}}(h^{\frac{q}{2}})$ and $nh^{\frac{q}{2}} T_{n,3} = \mathcal{O}_{\mathbb{P}}(h^{\frac{q}{2}})$, so by **A3** it happens that

$$
nh^{\frac{q}{2}} T_{n,3} \xrightarrow{p} 0 \text{ and } nh^{\frac{q}{2}} T_{n,2} \xrightarrow{p} 0. \tag{7.14}
$$

Term $T_{n,1}$. Now, $T_{n,1}$ can be dealt with the equivalent kernel of Corollary 7.1:

$$
\begin{aligned}
T_{n,1} &= \int_{\Omega_q} \left(\sum_{i=1}^{n} L_h^*(\mathbf{x}, \mathbf{X}_i) \left(1 + o_{\mathbb{P}}(1)\right) \left(Y_i - m_{\boldsymbol{\theta}_0}(\mathbf{X}_i)\right) \right)^2 f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \\
&= \widetilde{T}_{n,1} \left(1 + o_{\mathbb{P}}(1)\right).
\end{aligned}
$$

Using again Slutsky's theorem, the asymptotic distribution of $T_{n,1}$, and hence of T_n, will be the one of $\widetilde{T}_{n,1}$. Now it is possible to split

$$
\widetilde{T}_{n,1} = \widetilde{T}_{n,1}^{(1)} + \widetilde{T}_{n,1}^{(2)} + 2\widetilde{T}_{n,1}^{(3)} \tag{7.15}
$$

by recalling that $Y_i - m_{\boldsymbol{\theta}_0}(\mathbf{X}_i) = \sigma(\mathbf{X}_i)\varepsilon_i + \left(nh^{\frac{q}{2}}\right)^{-\frac{1}{2}} g(\mathbf{X}_i)$ by the model definition (7.3) and hypothesis H_{1P}. Specifically, under H_{1P} the conditional variance can be expressed as $\sigma^2(\mathbf{x}) = \mathbb{E}\left[\left(Y - m_{\boldsymbol{\theta}_0}(\mathbf{X}) - (nh^{\frac{q}{2}})^{-\frac{1}{2}} g(\mathbf{X})\right)^2 | \mathbf{X} = \mathbf{x}\right] = \sigma^2_{\boldsymbol{\theta}_0}(\mathbf{x})(1 + o(1))$, uniformly in $\mathbf{x} \in \Omega_q$ since g and $\sigma_{\boldsymbol{\theta}_0}$ are continuous and bounded by **A2** and **A7**. Therefore,

$$
\begin{aligned}
\widetilde{T}_{n,1}^{(1)} &= \int_{\Omega_q} \left(\sum_{i=1}^{n} L_h^*(\mathbf{x}, \mathbf{X}_i)\, \sigma(\mathbf{X}_i)\varepsilon_i \right)^2 f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}), \\
\widetilde{T}_{n,1}^{(2)} &= \left(nh^{\frac{q}{2}}\right)^{-1} \int_{\Omega_q} \left(\sum_{i=1}^{n} L_h^*(\mathbf{x}, \mathbf{X}_i)\, g(\mathbf{X}_i) \right)^2 f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}), \\
\widetilde{T}_{n,1}^{(3)} &= \left(nh^{\frac{q}{2}}\right)^{-\frac{1}{2}} \int_{\Omega_q} \sum_{i=1}^{n} \sum_{j=1}^{n} L_h^*(\mathbf{x}, \mathbf{X}_i)\, L_h^*(\mathbf{x}, \mathbf{X}_j)\, \sigma(\mathbf{X}_i)\varepsilon_i g(\mathbf{X}_j) f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}).
\end{aligned}
$$

By results *ii* and *iii* of Lemma B.2, the behavior of the two last terms is

$$
nh^{\frac{q}{2}} \widetilde{T}_{n,1}^{(2)} = \int_{\Omega_q} g(\mathbf{x})^2 f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) + o_{\mathbb{P}}(1) \text{ and } nh^{\frac{q}{2}} \widetilde{T}_{n,1}^{(3)} = o_{\mathbb{P}}(1). \tag{7.16}
$$

For the first addend, consider now

$$
\begin{aligned}
\widetilde{T}_{n,1}^{(1)} &= \int_{\Omega_q} \sum_{i=1}^{n} \left(L_h^*(\mathbf{x}, \mathbf{X}_i)\, \sigma(\mathbf{X}_i)\varepsilon_i\right)^2 f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \\
&\quad + \int_{\Omega_q} \sum_{i \neq j} L_h^*(\mathbf{x}, \mathbf{X}_i)\, L_h^*(\mathbf{x}, \mathbf{X}_j)\, \sigma(\mathbf{X}_i)\sigma(\mathbf{X}_j)\varepsilon_i\varepsilon_j f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \\
&= \widetilde{T}_{n,1}^{(1a)} + \widetilde{T}_{n,1}^{(1b)}.
\end{aligned}
$$

From result *iv* of Lemma B.2 and because $\sigma^2(\mathbf{x}) = \sigma^2_{\boldsymbol{\theta}_0}(\mathbf{x})(1 + o(1))$ uniformly,

$$
nh^{\frac{q}{2}} \widetilde{T}_{n,1}^{(1a)} = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{h^{\frac{q}{2}}} \int_{\Omega_q} \sigma^2_{\boldsymbol{\theta}_0}(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x})(1 + o(1)) + o_{\mathbb{P}}(1). \tag{7.17}
$$

The asymptotic behavior of $\widetilde{T}_{n,1}^{(1b)}$ is obtained using Theorem 2.1 in de Jong (1987). This result states that the sum $W_n = \sum_{i,j=1}^{n} W_{ijn}$, with W_{ijn} random variables depending on the sample size and on independent variables X_i and X_j, converges as $W_n \xrightarrow{d} \mathcal{N}(0, v^2)$ under the following conditions:

a) the random variables W_{ijn} are *clean*, *i.e.* $\mathbb{E}\left[W_{ijn} + W_{jin} | X_i\right] = 0$ for $1 \leq i < j \leq n$,

b) $\mathbb{V}\mathrm{ar}\left[W_n\right] \to v^2$,

c) $\left(\max_{1\leq i\leq n} \sum_{j=1}^{n} \mathbb{V}\mathrm{ar}\left[W_{ijn}\right]\right) v^{-2} \to 0$,

d) $\mathbb{E}\left[W_n^4\right] v^{-4} \to 3$.

In order to apply this result, let us denote first

$$W_{ijn} = \begin{cases} nh^{\frac{q}{2}} \int_{\Omega_q} L_h^*\left(\mathbf{x}, \mathbf{X}_i\right) L_h^*\left(\mathbf{x}, \mathbf{X}_j\right) \sigma(\mathbf{X}_i)\sigma(\mathbf{X}_j)\varepsilon_i\varepsilon_j f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}), & i \neq j, \\ 0, & i = j. \end{cases}$$

Then, $nh^{\frac{q}{2}}\widetilde{T}_{n,1}^{(1b)} = W_n = \sum_{i\neq j} W_{ijn}$ and the random variables on which W_{ijn} depends are $(\mathbf{X}_i, \varepsilon_i)$ and $(\mathbf{X}_j, \varepsilon_j)$. Condition *a)* is easily seen to hold by $\mathbb{E}\left[\varepsilon|\mathbf{X}\right] = 0$ and the tower property, which implies that $\mathbb{E}\left[W_{ijn}\right] = 0$. Because of this, the fact that $W_{ijn} = W_{jin}$ and Lemma 2.1 in de Jong (1987), we have for condition *b)*:

$$\mathbb{V}\mathrm{ar}\left[W_n\right] = \mathbb{E}\left[\Big(\sum_{i\neq j} W_{ijn}\Big)^2\right] = 2\mathbb{E}\left[\sum_{i\neq j} W_{ijn}^2\right] = 2n(n-1)\mathbb{E}\left[W_{ijn}^2\right]. \tag{7.18}$$

Then, by v in Lemma B.2 and the fact that $\sigma^2(\mathbf{x}) = \sigma_{\boldsymbol{\theta}_0}^2(\mathbf{x})(1+o(1))$, $\mathbb{E}[W_{ijn}^2] = n^{-2}\nu_{\boldsymbol{\theta}_0}^2\left(1 + o\left(1\right)\right)$ and as a consequence $\mathbb{V}\mathrm{ar}\left[W_n\right] \to 2\nu_{\boldsymbol{\theta}_0}^2$. Condition *c)* follows easily from the previous computation:

$$\left(\max_{1\leq i\leq n} \sum_{j=1}^{n} \mathbb{V}\mathrm{ar}\left[W_{ijn}\right]\right) v^{-2} \leq \left(\max_{1\leq i\leq n} n^{-1}\nu_{\boldsymbol{\theta}_0}^2\left(1 + o\left(1\right)\right)\right)(2\nu_{\boldsymbol{\theta}_0}^2)^{-1} = (2n)^{-1}(1 + o\left(1\right)) \to 0.$$

To check condition *d)*, note that $\mathbb{E}\left[W_n^4\right]$ can be split in the following form in virtue of Lemma 2.1 in de Jong (1987), as Härdle and Mammen (1993) stated:

$$\begin{aligned} \mathbb{E}\left[W_n^4\right] &= \mathbb{E}\Big[\sum_{i_1\neq j_1}\sum_{i_2\neq j_2}\sum_{i_3\neq j_3}\sum_{i_4\neq j_4} W_{i_1j_1n}W_{i_2j_2n}W_{i_3j_3n}W_{i_4j_4n}\Big] \\ &= 8\sum_{i,j}^{\neq} \mathbb{E}\left[W_{ijn}^4\right] + 12\sum_{i,j,k,l}^{\neq} \mathbb{E}\left[W_{ijn}^2 W_{kln}^2\right] + 48\sum_{i,j,k}^{\neq} \mathbb{E}\left[W_{ijn}W_{ikn}^2W_{jkn}\right] \\ &\quad + 192\sum_{i,j,k,l}^{\neq} \mathbb{E}\left[W_{ijn}W_{jkn}W_{kln}W_{lin}\right], \end{aligned} \tag{7.19}$$

where the notation $\sum^{\neq}$ stands for the summation over all *pairwise different* indexes (*i.e.*, indexes that satisfy $i \neq j$ for their associated W_{ijn}). By the results given in v of Lemma B.2, $\mathbb{E}[W_{ijn}^4] = \mathcal{O}\left((n^4h^q)^{-1}\right)$, $\mathbb{E}\left[W_{ijn}W_{jkn}W_{kln}W_{lin}\right] = \mathcal{O}\left(n^{-4}h^{2q}\right)$ and $\mathbb{E}\left[W_{ijn}W_{ikn}^2W_{jkn}\right] = \mathcal{O}\left(n^{-4}\right)$. Therefore, by (7.18) and (7.19),

$$\mathbb{E}\left[W_n^4\right] = 12\sum_{i\neq j}\sum_{k\neq l} \mathbb{E}\left[W_{ijn}^2W_{kln}^2\right] + o\left(1\right) = 3\Big(2\sum_{i\neq j} \mathbb{E}\left[W_{ijn}^2\right]\Big)^2 + o\left(1\right) = 3\mathbb{V}\mathrm{ar}\left[W_n\right]^2 + o\left(1\right)$$

and by **A4**, $\mathbb{E}\left[W_n^4\right] = 3\mathbb{V}\mathrm{ar}\left[W_n\right]^2 + o\left(1\right)$, so condition *d)* is satisfied, having that

$$nh^{\frac{q}{2}}\widetilde{T}_{n,1}^{(1b)} \xrightarrow{d} \mathcal{N}\left(0, 2\nu_{\boldsymbol{\theta}_0}^2\right). \tag{7.20}$$

Finally, using decompositions (7.13) and (7.15) and results (7.14) and (7.16), it holds

$$\begin{aligned} nh^{\frac{q}{2}}T_n =& nh^{\frac{q}{2}}\left(\left[\widetilde{T}_{n,1}^{(1a)} + \widetilde{T}_{n,1}^{(1b)} + \widetilde{T}_{n,1}^{(2)} - 2\widetilde{T}_{n,1}^{(3)}\right]\left(1 + o_{\mathbb{P}}\left(1\right)\right) + T_{n,2} + 2T_{n,3}\right) \\ =& \left(\frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{h^{\frac{q}{2}}}\int_{\Omega_q}\sigma_{\boldsymbol{\theta}_0}^2(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}) + nh^{\frac{q}{2}}\widetilde{T}_{n,1}^{(1b)}\right. \\ &\left. + \int_{\Omega_q} g(\mathbf{x})^2 f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right)\left(1 + o_{\mathbb{P}}\left(1\right)\right) \end{aligned}$$

and the limit distribution follows by Slutsky's theorem and result (7.20). □

Proof of Theorem 7.5. The proof mimics the steps of the proof of Theorem 7.3. First of all, the bootstrap test statistic T_n^* can be separated as

$$T_n^* = T_{n,1}^* + T_{n,2}^* - 2T_{n,3}^*, \tag{7.21}$$

where

$$\begin{aligned} T_{n,1}^* &= \int_{\Omega_q}\left(\sum_{i=1}^n W_n^p\left(\mathbf{x}, \mathbf{X}_i\right)\left(Y_i^* - m_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i)\right)\right)^2 \hat{f}_h(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}), \\ T_{n,2}^* &= \int_{\Omega_q}\left(\mathcal{L}_{h,p}(m_{\hat{\boldsymbol{\theta}}} - m_{\hat{\boldsymbol{\theta}}^*})(\mathbf{x})\right)^2 \hat{f}_h(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}), \\ T_{n,3}^* &= \int_{\Omega_q}\left(\hat{m}_{h,p}^*(\mathbf{x}) - \mathcal{L}_{h,p}m_{\hat{\boldsymbol{\theta}}}(\mathbf{x})\right)\mathcal{L}_{h,p}(m_{\hat{\boldsymbol{\theta}}} - m_{\hat{\boldsymbol{\theta}}^*})(\mathbf{x})\hat{f}_h(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}). \end{aligned}$$

Terms $T_{n,2}^*$ and $T_{n,3}^*$. By assumption **A9** and analogous computations to the ones in the proof of Theorem 7.3, it follows that $nh^{\frac{q}{2}}T_{n,2}^* \xrightarrow{p^*} 0$ and $nh^{\frac{q}{2}}T_{n,3}^* \xrightarrow{p^*} 0$, where the convergence is stated in the probability law $\mathbb{P}^*$ that is conditional on the sample $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$.

Term $T_{n,1}^*$. By the resampling procedure of Algorithm 7.1, $\hat{\varepsilon}_i V_i^* = (Y_i - m_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i))V_i^*$ and the dominant term can be split into

$$\begin{aligned} T_{n,1}^* =& \int_{\Omega_q}\sum_{i=1}^n \left(W_n^p\left(\mathbf{x}, \mathbf{X}_i\right)\hat{\varepsilon}_i V_i^*\right)^2 \hat{f}_h(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}) \\ &+ \int_{\Omega_q}\sum_{i\neq j} W_n^p\left(\mathbf{x}, \mathbf{X}_i\right)W_n^p\left(\mathbf{x}, \mathbf{X}_j\right)\hat{\varepsilon}_i V_i^*\hat{\varepsilon}_j V_j^*\hat{f}_h(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}) \\ =& T_{n,1}^{*(1)} + T_{n,1}^{*(2)}. \end{aligned}$$

From result *i* of Lemma B.3, the first term is

$$nh^{\frac{q}{2}}T_{n,1}^{*(1)} = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{h^{\frac{q}{2}}}\int_{\Omega_q}\sigma_{\boldsymbol{\theta}_1}^2(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})(1 + o_{\mathbb{P}}\left(1\right)) + o_{\mathbb{P}^*}(1), \tag{7.22}$$

so the dominant term is $T_{n,1}^{*(2)}$, whose asymptotic behavior is obtained using Theorem 2.1 in de Jong (1987) conditionally on the sample. Let us denote now

$$W_{ijn}^{*} = \begin{cases} nh^{\frac{q}{2}} \int_{\Omega_q} W_n^p(\mathbf{x}, \mathbf{X}_i) W_n^p(\mathbf{x}, \mathbf{X}_j) \hat{\varepsilon}_i V_i^* \hat{\varepsilon}_j V_j^* \hat{f}_h(\mathbf{x}) w(\mathbf{x}) \omega_q(d\mathbf{x}), & i \neq j, \\ 0, & i = j. \end{cases}$$

Then, $nh^{\frac{q}{2}} T_{n,1}^{*(2)} = W_n^* = \sum_{i \neq j} W_{ijn}^*$ and the random variables on which W_{ijn}^* depends are now V_i^* and V_j^*. Condition *a)* of the theorem follows immediately by the properties of the V_i^*'s: $\mathbb{E}^*[W_{ijn}^* + W_{jin}^* | V_i^*] = 0$. On the other hand, analogously to (7.18),

$$\mathbb{V}\mathrm{ar}^*[W_n^*] = 2 \sum_{i \neq j} \mathbb{E}^*\left[W_{ijn}^{*2}\right] = 2n^2 h^q \sum_{i \neq j} \left[\int_{\Omega_q} W_n^p(\mathbf{x}, \mathbf{X}_i) W_n^p(\mathbf{x}, \mathbf{X}_j) \hat{\varepsilon}_i \hat{\varepsilon}_j \hat{f}_h(\mathbf{x}) w(\mathbf{x}) \omega_q(d\mathbf{x}) \right]^2$$

and by result *ii* of Lemma B.3, $\mathbb{V}\mathrm{ar}^*[W_n^*] \xrightarrow{p} 2\nu_{\boldsymbol{\theta}_1}^2$, resulting in the verification of condition *c)* in probability. Condition *d)* is checked using the same decomposition for $\mathbb{E}^*[W_n^{*4}]$ and the results collected in *ii* of Lemma B.3. Hence $\mathbb{E}^*[W_n^{*4}] = 3\mathbb{V}\mathrm{ar}^*[W_n^*]^2 + o_{\mathbb{P}}(1)$ and *d)* is satisfied in probability, from which it follows that, conditionally on $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$ the pdf of $nh^{\frac{q}{2}} T_{n,1}^{*(2)}$ converges in probability to the pdf of $\mathcal{N}(0, 2\nu_{\boldsymbol{\theta}_1}^2)$, that is:

$$nh^{\frac{q}{2}} T_{n,1}^{*(2)} \xrightarrow{d} \mathcal{N}\left(0, 2\nu_{\boldsymbol{\theta}_1}^2\right) \text{ in probability.} \tag{7.23}$$

Joining (7.21) and (7.23) and applying Slutsky's theorem conditionally on the sample, the theorem is proved:

$$\begin{aligned} nh^{\frac{q}{2}} T_n^* &= nh^{\frac{q}{2}} \left(T_{n,1}^{*(1)} + T_{n,1}^{*(2)} + T_{n,2}^* + T_{n,3}^* \right) \\ &= \left(\frac{\lambda_q(L^2) \lambda_q(L)^{-2}}{h^{\frac{q}{2}}} \int_{\Omega_q} \sigma_{\boldsymbol{\theta}_1}^2(\mathbf{x}) w(\mathbf{x}) \omega_q(d\mathbf{x}) + nh^{\frac{q}{2}} T_{n,1}^{*(2)} \right) (1 + o_{\mathbb{P}}(1)) + o_{\mathbb{P}^*}(1). \end{aligned}$$

□

References

Alcalá, J. T., Cristóbal, J. A., and González-Manteiga, W. (1999). Goodness-of-fit test for linear models based on local polynomials. *Statist. Probab. Lett.*, 42(1):39–46.

Asur, S. and Huberman, B. A. (2010). Predicting the future with social media. In Huang, J. X., King, I., Raghavan, V., and Rueger, S., editors, *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pages 492–499. IEEE.

Bai, Z. D., Rao, C. R., and Zhao, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.*, 27(1):24–39.

Banerjee, A., Dhillon, I. S., Ghosh, J., and Sra, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions. *J. Mach. Learn. Res.*, 6:1345–1382.

Bowman, A. W. and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*. Oxford Statistical Science Series. Clarendon Press, Oxford.

Buchta, C., Kober, M., Feinerer, I., and Hornik, K. (2012). Spherical k-means clustering. *J. Stat. Softw.*, 50(10):1–22.

Cheng, M.-Y. and Wu, H.-T. (2013). Local linear regression on manifolds and its geometric interpretation. *J. Amer. Statist. Assoc.*, 108(504):1421–1434.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields*, 75(2):261–277.

Deschepper, E., Thas, O., and Ottoy, J. P. (2008). Tests and diagnostic plots for detecting lack-of-fit for circular-linear regression models. *Biometrics*, 64(3):912–920.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.*, 79(19):2066–2075.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2014). Nonparametric regression for spherical data. *J. Amer. Statist. Assoc.*, 109(506):748–763.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Fan, J., Gijbels, I., Hu, T.-C., and Huang, L.-S. (1996). A study of variable bandwidth selection for local polynomial regression. *Statist. Sinica*, 6(1):113–127.

García-Portugués, E. (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron. J. Stat.*, 7:1655–1685.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013). Kernel density estimation for directional-linear data. *J. Multivariate Anal.*, 121:152–175.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2014). Central limit theorems for directional and linear data with applications. *Statist. Sinica*, to appear.

Hall, P., Watson, G. S., and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, 21(4):1926–1947.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning*. Springer Series in Statistics. Springer, New York, second edition.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.*, 40(2):633–643.

Joachims, T. (2002). *Learning to classify text using support vector machines: Methods, theory and algorithms*, volume 668 of *Kluwer International Series in Engineering and Computer Science*. Kluwer Academic Publishers, Boston.

Mardia, K. V. and Jupp, P. E. (2000). *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, second edition.

Meyer, D., Hornik, K., and Feinerer, I. (2008). Text mining infrastructure in R. *J. Stat. Softw.*, 25(5):1–54.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.*, 22(3):1346–1370.

Srivastava, A. N. and Sahami, M., editors (2009). *Text mining: classification, clustering, and applications*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press, Boca Raton.

Surian, D. and Chawla, S. (2013). Mining outlier participants: Insights using directional distributions in latent models. In Blockeel, H., Kersting, K., Nijssen, S., and Žtelezný, F., editors, *Machine learning and knowledge discovery in databases*, volume 8188 of *Lecture Notes in Artificial Intelligence*, pages 337–352. Springer, Heidelberg.

Tatar, A., Antoniadis, P., De Amorim, M. D., and Fdida, S. (2012). Ranking news articles based on popularity prediction. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, pages 106–110. IEEE.

Wang, X., Zhao, L., and Wu, Y. (2000). Distribution free laws of the iterated logarithm for kernel estimator of regression function based on directional data. *Chinese Ann. Math. Ser. B*, 21(4):489–498.

Zhao, L. and Wu, C. (2001). Central limit theorem for integrated square error of kernel estimators of spherical density. *Sci. China Ser. A*, 44(4):474–483.

Chapter 8

Future research

This chapter presents some ideas for future contributions on different areas related with the scope of the thesis. The first sections are focused on further contributions on kernel smoothing with directional data: Section 8.1 studies new bandwidth selectors for the projected local linear estimator presented in Chapter 7 and Section 8.2 introduces a new kernel density estimator under rotational symmetry motivated by the results on Chapters 3 and 4. Section 8.3 describes the structure of an `R` package containing all the code developed for the implementation of the methods presented in the thesis. Finally, Section 8.4 looks back to Chapter 2 to present a test for the copula structure of Wehrly and Johnson (1979) and Johnson and Wehrly (1978) models using empirical processes.

Contents

8.1 Bandwidth selection in nonparametric linear-directional regression 175

8.2 Kernel density estimation under rotational symmetry 180

8.3 `R` package `DirStats` 185

8.4 A goodness-of-fit test for the Johnson and Wehrly copula structure 187

References 189

8.1 Bandwidth selection in nonparametric linear-directional regression

In Chapter 7 a projected local estimator was proposed to estimate the regression function of a linear response on a directional predictor. From the results derived on the bias and variance (see Theorem 7.1) it is possible to derive bandwidth selection rules for the proposed estimator. The next corollary provides a starting point, stating the conditional Mean Integrated Squared Error (MISE) and the optimal bandwidth expression that minimizes the Asymptotic MISE (AMISE) of the projected local estimator. A weight function w is included in the conditional MISE definition to allow an easy adaptation of the global error criteria to the case where the interest is the estimation of m on specific areas of the support of the predictor (for example, the areas with larger density).

Corollary 8.1 (AMISE optimal bandwidth)**.** *Under assumptions **A1**–**A4** from Chapter 7, for a given weight function $w : \Omega_q \longrightarrow \mathbb{R}$, the conditional weighted MISE of the projected local*

estimator is

$$\begin{aligned}\operatorname{MISE}\left[\hat{m}_{h,p}|\mathbf{X}_1,\ldots,\mathbf{X}_n\right] = &\frac{b_q(L)^2}{q^2}\int_{\Omega_q} B_p(\mathbf{x})^2 w(\mathbf{x})\,\omega_q(d\mathbf{x}) h^4 \\ &+ \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\frac{\sigma^2(\mathbf{x})}{f(\mathbf{x})} w(\mathbf{x})\,\omega_q(d\mathbf{x}) \\ &+ o_{\mathbb{P}}\left(h^4+(nh^q)^{-1}\right)+\delta_{p,0}\mathcal{O}_{\mathbb{P}}\left(\frac{h^2}{nh^q}\right).\end{aligned}$$

The bandwidth that minimizes the weighted AMISE of the local estimator is

$$h_{\mathrm{AMISE}} = \left[\frac{q^3\lambda_q(L^2)\lambda_q(L)^{-2}\int_{\Omega_q}\frac{\sigma^2(\mathbf{x})}{f(\mathbf{x})} w(\mathbf{x})\,\omega_q(d\mathbf{x})}{4b_q(L)^2\int_{\Omega_q} B_p(\mathbf{x})^2 w(\mathbf{x})\,\omega_q(d\mathbf{x}) n}\right]^{\frac{1}{4+q}}.$$

If the von Mises kernel is used, then

$$h_{\mathrm{AMISE}} = \left[\frac{q^3\int_{\Omega_q}\frac{\sigma^2(\mathbf{x})}{f(\mathbf{x})} w(\mathbf{x})\,\omega_q(d\mathbf{x})}{(2\pi^{\frac{1}{2}})^q\int_{\Omega_q} B_p(\mathbf{x})^2 w(\mathbf{x})\,\omega_q(d\mathbf{x}) n}\right]^{\frac{1}{4+q}}.$$

The order of the resulting weighted AMISE for the h_{AMISE} bandwidth is $n^{-\frac{4}{4+q}}$.

However, this bandwidth depends on the unknown functions $f(\mathbf{x})$, $B_p(\mathbf{x})$ (see Theorem 7.1) and $\sigma^2(\mathbf{x})$. A first idea to improve its usability is to consider the weight function proportional to the density, $w(\mathbf{x}) = w_0(\mathbf{x})f(\mathbf{x})$, with $w_0 : \Omega_q \longrightarrow \mathbb{R}$. With this weight function (which penalizes the estimation error on the most likely predictor values), the density in the denominator cancels out while it appears on the integral of the numerator, enabling a Monte Carlo estimation. In addition, assuming homocedasticity and a pilot parametric model for the regression function, it is possible to approximate the unknown terms using their corresponding parametric estimates. These two considerations lead to the following *rule of thumb bandwidth selector*:

$$h_{\mathrm{ROT}} = \left[\frac{q^3\lambda_q(L^2)\lambda_q(L)^{-2}\int_{\Omega_q}\hat{\sigma}^2 w_0(\mathbf{x})\,\omega_q(d\mathbf{x})}{4b_q(L)^2\sum_{i=1}^n B_{p,\hat{\boldsymbol{\theta}}}(\mathbf{X}_i)^2 w_0(\mathbf{X}_i)}\right]^{\frac{1}{4+q}}.$$

This selector depends on a homocedastic parametric model $Y = m_{\boldsymbol{\theta}}(\mathbf{X}) + \sigma\varepsilon$, which can be for example a quadratic one:

$$m^Q_{\boldsymbol{\theta}}(\mathbf{x}) = m + \mathbf{m}^T\mathbf{x} + \mathbf{q}^T\mathbf{x}^2_{-(q+1)} = \left(m, \mathbf{m}^T, \mathbf{q}^T\right)\begin{pmatrix}1\\ \mathbf{x}\\ \mathbf{x}^2_{-(q+1)}\end{pmatrix}, \tag{8.1}$$

where $\mathbf{x}^2_{-(q+1)}$ stands for the vector with squared components and omitting its last entry. The estimation of the parameters is easily accomplished by $(\hat{m}, \hat{\mathbf{m}}, \hat{\mathbf{q}}) = \left(\boldsymbol{\mathcal{X}}_m^T\boldsymbol{\mathcal{X}}_m\right)^{-1}\boldsymbol{\mathcal{X}}_m^T\mathbf{Y}$ and $\hat{\sigma}^2 = \frac{1}{n-2(q+1)}\sum_{i=1}^n\left(Y_i - m^Q_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i)\right)^2$, where

$$\boldsymbol{\mathcal{X}}_m = \begin{pmatrix}1 & \mathbf{X}_1^T & \left(\mathbf{X}^2_{1,-(q+1)}\right)^T\\ \vdots & \vdots & \vdots\\ 1 & \mathbf{X}_n^T & \left(\mathbf{X}^2_{n,-(q+1)}\right)^T\end{pmatrix}.$$

The choice of this parametric model is justified because it is the simplest model that allows an estimation of the first and second derivatives of m and also allows for the computation of the bias integrand in a closed form:

$$B_{p,\boldsymbol{\theta}}(\mathbf{x}) = \begin{cases} 2\frac{\boldsymbol{\nabla} f(\mathbf{x})^T(\mathbf{m}+2(q_1x_1,\ldots,q_qx_q))}{f(\mathbf{x})} + \sum_{i=1}^q q_i, & p=0, \\ \sum_{i=1}^q q_i, & p=1. \end{cases} \tag{8.2}$$

For the estimation of $\boldsymbol{\nabla} f(\mathbf{x})$ in (8.2) when $p=0$, it is possible to consider an r-mixture of von Mises densities for the unknown density f and fit them using Algorithm 4.3, resulting in

$$\boldsymbol{\nabla} f_r(\mathbf{x}) = \sum_{j=1}^r C_q(\hat{\kappa}_j)\hat{\kappa}_j e^{\hat{\kappa}_j \mathbf{x}^T\hat{\boldsymbol{\mu}}_j}\hat{\boldsymbol{\mu}}_j.$$

The h_{ROT} is a simple bandwidth selector that will work better when the regression function m is close to $m^Q_{\boldsymbol{\theta}}$. Since this assumption may not be realistic in many situations, it can be considered as the beginning for more sophisticated selectors based on plug-in ideas. Under homocedasticity, there are several possibilities described by Ruppert et al. (1995), mimicking in the regression context the ideas given by Sheather and Jones (1991) for the density. A first possibility is to estimate the unknown quantity in the denominator of h_{AMISE} (assuming $w(\mathbf{x}) = w_0(\mathbf{x})f(\mathbf{x})$),

$$B_{22} = \int_{\Omega_q} B_p(\mathbf{x})^2 f(\mathbf{x}) w_0(\mathbf{x})\, \omega_q(d\mathbf{x}),$$

by a blocked quadratic fit (Härdle and Marron, 1995). This estimator could be computed from N estimators (8.1) on N random subsamples obtained from the ordered sample, *i.e.*: $m^Q_{\hat{\boldsymbol{\theta}}_j}$ obtained from $\boldsymbol{\mathcal{X}}_j = \{\mathbf{X}_{((j-1)t+1)}, \ldots, \mathbf{X}_{(jt)}\}$, $t = \lfloor \frac{n}{N} \rfloor$, $j = 1, \ldots, N$. Then, an estimate of B_{22} is

$$\widehat{B}^Q_{22}(N) = \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^N B^Q_{p,\hat{\boldsymbol{\theta}}_j}(\mathbf{X}_i)^2 w_0(\mathbf{X}_i)\mathbf{1}_{\{\mathbf{X}_i \in \boldsymbol{\mathcal{X}}_j\}}$$

and the estimation of the variance will be obtained by

$$\hat{\sigma}^2_Q(N) = \frac{1}{n - 2(q+1)N}\sum_{i=1}^n \sum_{j=1}^N \left(Y_i - m^Q_{\hat{\boldsymbol{\theta}}_j}(\mathbf{X}_i)\right)^2 \mathbf{1}_{\{\mathbf{X}_i \in \boldsymbol{\mathcal{X}}_j\}}.$$

The choice of the tuning parameter N can be done by the Bayesian Information Criteria (BIC). Another possibility to improve h_{ROT} is to focus on the local linear case ($p=1$) and estimate B_{22} by

$$\widehat{B}_{22}(g,\nu) = \frac{1}{n}\sum_{i=1}^n \mathrm{tr}\left[\boldsymbol{\mathcal{H}}_{\hat{m}_{g,\nu}}(\mathbf{X}_i)\right]^2 w_0(\mathbf{X}_i), \text{ with } \nu \geq 2,$$

being tr the trace operator and $\boldsymbol{\mathcal{H}}_{\hat{m}_{g,\nu}}$ the Hessian matrix of a projected local polynomial estimator $\hat{m}_{g,\nu}$. This induces a new problem on choosing a suitable bandwidth g for estimating the functional B_{22}, which can be done by means of the bandwidth g_{AMSE} that minimizes the Asymptotic Mean Squared Error (AMSE) of $\widehat{B}_{22}(g,\nu)$. This expression depends on high order derivatives of m, that should be estimated by a parametric fit of order ν (this is, cubic or quartic), which will be constructed from a blocked cubic or quartic fit. Further research on these two bandwidth selectors will be required and, particularly, on the extension of the projected local estimator of Chapter 7 to higher polynomial orders.

Finally, adaptations of the classical cross-validatory bandwidth selectors can be also defined in this setting. The usual Cross-Validatory (CV) bandwidth is given by

$$h_{\mathrm{CV}} = \arg\min_{h>0} \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{m}_{h,p}^{-i}(\mathbf{X}_i)\right)^2,$$

whose computation can be optimized using that $\hat{m}_{h,p}^{-i}(\mathbf{X}_i) = \sum_{j\neq i} \frac{W_n^p(\mathbf{X}_i,\mathbf{X}_j)}{1-W_n^p(\mathbf{X}_i,\mathbf{X}_i)} Y_j$ by the theory of linear smoothers (Hastie and Tibshirani, 1990, page 46). Then, it is only necessary to compute the hat matrix $\mathbf{H} = (W_n^p(\mathbf{X}_i,\mathbf{X}_j))_{ij}$ just once for each bandwidth h without refitting, yielding

$$h_{\mathrm{CV}} = \arg\min_{h>0} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \hat{m}_{h,p}(\mathbf{X}_i)}{1-H_{ii}}\right)^2.$$

In addition, a bandwidth selector based on the Generalized Cross Validation (GCV) criterion (see Hastie and Tibshirani (1990)) can be defined also as follows:

$$h_{\mathrm{GCV}} = \arg\min_{h>0} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \hat{m}_{h,p}(\mathbf{X}_i)}{1-\mathrm{tr}\left[\mathbf{H}\right]/n}\right)^2.$$

In order to compare the available selectors h_{ROT}, h_{CV} and h_{GCV} and the possible improvements of h_{ROT}, an extensive simulation study similar to the one given in Chapter 4 in the density situation will be carried out. For the design of the simulation study three components must be chosen: the regression function m, the generating process of the directional covariate $\mathbf{X}$ and the noise structure in the response, $\sigma(\mathbf{X})\varepsilon$. To account for different combinations, twelve regression models with directional covariate and scalar response, six directional densities and two kinds of noise are proposed.

Figure 8.1 represents the collection of regression models with the following codification: the radius of the q-sphere represents the response $m(\mathbf{x})$ for a $\mathbf{x}$ direction. As the response of the models is contained in the interval $[-4,4]$, a translation is applied such that a response -4 is mapped to a zero radius and a response 4 is mapped to a radius 8. The graphs then show the response as a distortion from a perfect circle or sphere, for a sequence of parametric regression models with increasing complexity. Figure 8.2 shows the six kinds of directional densities, encoded as D1 to D6 and with increasing complexity. These densities are obtained from the ones in Chapter 4 (M1 to M12, see Subsection 4.5.1 for notation related with the densities): D1 is M1, D2 is the mixture $\frac{2}{5}$M1+$\frac{3}{5}$M4, D3 is $\frac{2}{5}$M1+$\frac{3}{5}$M5, D4 is $\frac{2}{5}$M1+$\frac{3}{5}$M12, D5 is M15 and D6 is M20. The noise considered is $\varepsilon \sim \mathcal{N}(0,1)$, with two different conditional variances given by $\sigma_1^2(\mathbf{x}) = \frac{1}{2}$ (homocedastic) and $\sigma_2^2(\mathbf{x}) = \left(\frac{1}{4} + 3\left(\frac{1}{2}f_{\mathrm{SC}}(\mathbf{x},(\mathbf{0}_q,1),10) + \frac{1}{2}f_{\mathrm{SC}}(\mathbf{x},(1,\mathbf{0}_q),10)\right)\right)^2$ (heteroskedastic). Finally, the combination of these three components is shown in Table 8.1. With these simulation scenarios a large variety of situations for homocedasticity and heteroskedasticity are covered for different degrees of complexity: from simple linear models like S1–S2 to high non-linear ones like S10–S12, and from simple uniform or unimodal design densities (D1–D3) to complex multimodal ones (D5–D6).

The error criterion proposed for measuring the performance of each selector is the Averaged Squared Error (ASE), defined as

$$\mathrm{ASE}(h,p) = \frac{1}{n}\sum_{i=1}^{n}\left(m(\mathbf{X}_i) - \hat{m}_{h,p}(\mathbf{X}_i)\right)^2,$$

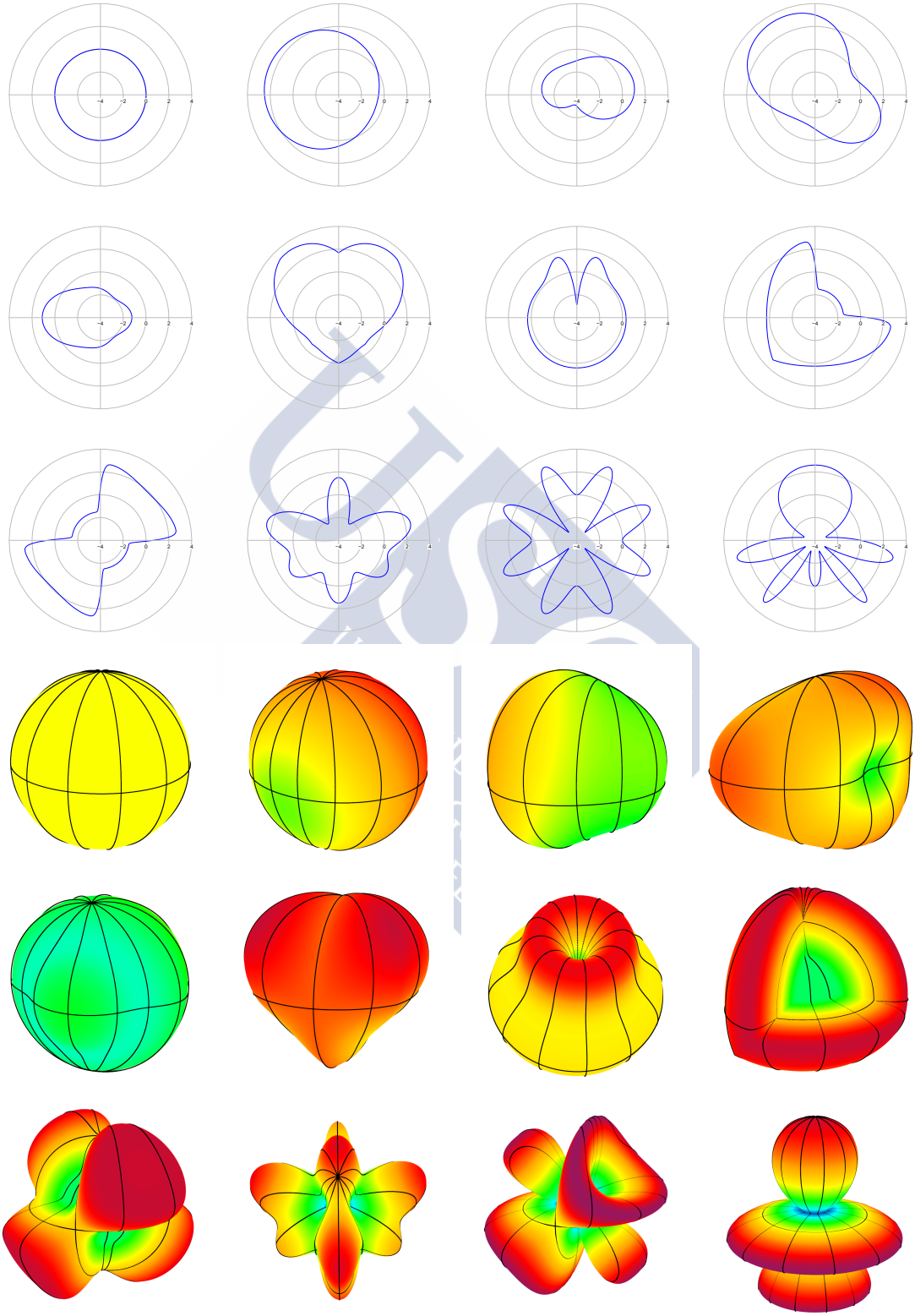


Figure 8.1: Directional regression models proposed for the simulation study. From left to right and up to down, models used in scenarios S1 to S12, with the first three rows fo the circular versions and the last three for the spherical ones.

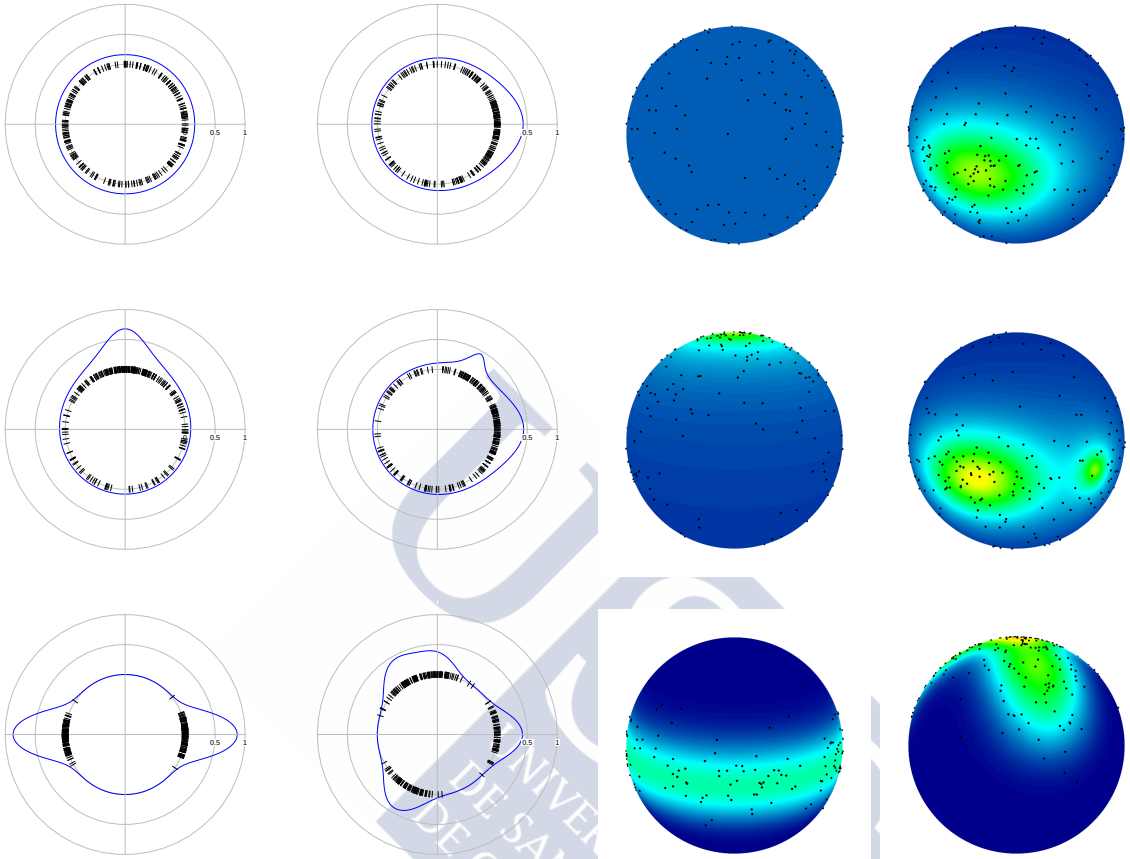


Figure 8.2: Directional densities to be considered in the simulation study. From left to right and up to down, the first two columns represent the directional densities D1 to D6 in the circular case. The corresponsing spherical versions are in the third and fourth columns. For each density, a sample of size $n = 250$ is drawn.

where m is the true regression function. This criterion can be seen as the Monte Carlo version of the conditional MISE with weight $w(\mathbf{x}) = f(\mathbf{x})$. However, unlike the latter, the ASE does not require to compute an integral, avoiding numerical instabilities when the design density is almost zero and the regression function takes large values (this happens for the projected local linear estimator, as well as for the usual Euclidean local linear estimator). For $M = 1000$ replicates, the ASE values will be averaged on the different scenarios for sample sizes $n = 100, 250, 500$, dimensions $q = 1, 2, 3, 4, 5$, estimators $p = 0, 1$ and bandwidth selectors.

The improvements related with the h_{ROT} bandwidth, the comparison of all the selectors in a simulation study and the examination of the empirical results constitute the remaining issues to be done in this future work.

8.2 Kernel density estimation under rotational symmetry

A common feature among the most popular distributions in directional statistics is rotational symmetry. For example, in the circular case most of the parametric distributions proposed in the

statistical literature are symmetric around an angle, although skewed distributions have gained certain attention in the last years (see Pewsey (2006) and Abe and Pewsey (2011), among others). Directional distributions that are rotationally symmetric include the von Mises (Watson, 1983), specific cases of the projected normal (Pukkila and Rao, 1988) and the skew normal directional and directional Cauchy, defined in Chapter 4. Furthermore, axial distributions (that is, with density f satisfying $f(\mathbf{x}) = f(-\mathbf{x})$, $\forall \mathbf{x} \in \Omega_q$) can also be regarded as rotational symmetric with location parameter not uniquely defined. Some examples of axial distributions are the Watson (Mardia and Jupp, 2000) and small circle (Bingham and Mardia, 1978) distributions.

Scenario	Regression function	Parameters	Density	Noise
S1	$m(\mathbf{x}) = m$	$m = 1$	D1	Het.
S2	$m(\mathbf{x}) = m + \boldsymbol{\eta}^T\mathbf{x}$	$m = 1$, $\boldsymbol{\eta} = (-\frac{3}{2}, \mathbf{\frac{1}{2}}_q)$	D2	Het.
S3	$m(\mathbf{x}) = m + \boldsymbol{\eta}^T\mathbf{x} + \boldsymbol{\gamma}^T(x_1^2, \ldots, x_q^2)$	$m = -2$, $\boldsymbol{\eta} = \mathbf{1}_{q+1}$, $\boldsymbol{\gamma} = (2, \mathbf{\frac{1}{2}}_{q-1})$	D3	Het.
S4	$m(\mathbf{x}) = m + a(x_{q+1}^2 - x_q^3) + bx_1x_2$	$m = 2$, $a = 2$, $b = -2$	D4	Het.
S5	$m(\mathbf{x}) = m + ax_1^4 + \frac{b}{\sqrt{\frac{3}{2} - x_1}}$	$m = 1$, $a = 2$, $b = -3$	D5	Het.
S6	$m(\mathbf{x}) = ae^{bx_2}\log(cx_{q+1}\lvert x_1\rvert) + d\max(\lvert x_1\rvert, \lvert x_2\rvert)$	$a = 4$, $b = \frac{1}{2}$, $c = \frac{3}{2}$, $d = -1$	D6	Het.
S7	$m(\mathbf{x}) = m + d_1 f_{\text{vM}}(\mathbf{x}, (\mathbf{0}_q, 1), \kappa_1) - d_2 f_{\text{DC}}(\mathbf{x}, (\mathbf{0}_q, 1), \kappa_2)$	$m = \frac{1}{2}$, $d_1 = 5$, $d_2 = 6$, $\kappa_1 = 10$, $\kappa_2 = 25$	D1	Hom.
S8	$m(\mathbf{x}) = m + af_{\text{SN}}(\max(-\mathbf{x}), b, c, d)$	$m = -\frac{3}{2}$, $a = 4$, $b = 0$, $c = \frac{3}{4}$, $d = 15$	D2	Hom.
S9	$m(\mathbf{x}) = m + af_{\text{SN}}(\prod_{i=1}^{q+1} x_i, b, c, d)$	$m = -\frac{3}{2}$, $a = 4$, $b = 0$, $c = \frac{3}{4}$, $d = 15$	D3	Hom.
S10	$m(\mathbf{x}) = m + a\sin(2\pi x_2) + b\cos(2\pi x_1)$	$m = 0$, $a = 1$, $b = \frac{3}{2}$	D4	Hom.
S11	$m(\mathbf{x}) = m + a\sin(2\pi b x_1 x_{q+1})$	$m = 0$, $a = 3$, $b = \frac{3}{2}$	D5	Hom.
S12	$m(\mathbf{x}) = m + a\sin\left(\frac{2\pi b}{2 + x_{q+1}}\right)$	$m = 0$, $a = 3$, $b = 4$	D6	Hom.

Table 8.1: Simulation scenarios for comparing the bandwidth selectors. f_{SN} is the Skew Normal distribution of Azzalini (1985), while the rest of notations can be seen in Subsection 4.5.1.

In general, a directional random variable $\mathbf{X}$ with density f is said to be rotationally symmetric around a location $\boldsymbol{\theta} \in \Omega_q$ if the corresponding density is of the form

$$f(\mathbf{x}) = g(\mathbf{x}^T\boldsymbol{\theta}), \quad \forall \mathbf{x} \in \Omega_q, \tag{8.3}$$

where $g : [-1, 1] \longrightarrow [0, \infty)$ is a function such that $\omega_{q-1}\int_{-1}^{1} g(t)(1 - t^2)^{\frac{q}{2}-1}\, dt = 1$. This class of distributions was firstly proposed by Saw (1978) requiring g to be monotone increasing, a necessary assumption to avoid identifiability issues but that excludes the class of axial distributions. The rotational symmetry property (8.3), equivalent to $f(\mathbf{x}) = f((\mathbf{x}^T\boldsymbol{\theta})\boldsymbol{\theta})$, $\forall \mathbf{x} \in \Omega_q$, is intimately related with the tangent normal decomposition around $\boldsymbol{\theta}$:

$$\mathbf{x} = (\mathbf{x}^T\boldsymbol{\theta})\boldsymbol{\theta} + (1 - (\mathbf{x}^T\boldsymbol{\theta})^2)^{\frac{1}{2}}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\xi}, \tag{8.4}$$

where $\boldsymbol{\xi} \in \Omega_{q-1}$ and $\mathbf{B}_{\boldsymbol{\theta}} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$ is the semi-orthonormal matrix resulting from the completion of $\boldsymbol{\theta}$ to the orthonormal basis $\{\boldsymbol{\theta}, \mathbf{b}_1, \ldots, \mathbf{b}_q\}$. Using this decomposition, it is possible to define the *rotsymmetrizer* operator around $\boldsymbol{\theta}$, which transforms any directional density into a rotational symmetric one. Then, for a function $f : \Omega_q \longrightarrow \mathbb{R}$, the rotsymmetrizer operator around $\boldsymbol{\theta}$ is

$$S_{\boldsymbol{\theta}} f(\mathbf{x}) = \frac{1}{\omega_{q-1}} \int_{\Omega_{q-1}} f\left((\mathbf{x}^T\boldsymbol{\theta})\boldsymbol{\theta} + (1-(\mathbf{x}^T\boldsymbol{\theta})^2)^{\frac{1}{2}}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\xi}\right) \omega_{q-1}(d\boldsymbol{\xi}). \tag{8.5}$$

For each point $\mathbf{x} \in \Omega_q$, the rotsymmetrizer operator computes the average value of the function f in the $(q-1)$-sphere orthogonal to the location $\boldsymbol{\theta}$ using (8.4), guaranteeing that the final output is a rotational symmetric density around $\boldsymbol{\theta}$. The next result collects some of the properties of this operator.

Proposition 8.1 (Properties of the rotsymmetrizer operator)**.** *Let $f, f_1, f_2 : \Omega_q \longrightarrow \mathbb{R}$ be directional densities and $\boldsymbol{\theta} \in \Omega_q$. The operator (8.5) has the following properties:*

i. Invariance from $\mathbf{B}_{\boldsymbol{\theta}}$: for different semi-orthogonal matrices $\mathbf{B}_{\boldsymbol{\theta},1}$ and $\mathbf{B}_{\boldsymbol{\theta},2}$ orthogonal to $\boldsymbol{\theta}$, if $S_{\boldsymbol{\theta},1}$ and $S_{\boldsymbol{\theta},2}$ denote the rotsymmetrizer operators constructed from $\mathbf{B}_{\boldsymbol{\theta},1}$ and $\mathbf{B}_{\boldsymbol{\theta},2}$, respectively, then $S_{\boldsymbol{\theta},1}f = S_{\boldsymbol{\theta},2}f$.

ii. Linearity: $S_{\boldsymbol{\theta}}(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 S_{\boldsymbol{\theta}} f_1 + \lambda_2 S_{\boldsymbol{\theta}} f_2$, for $\lambda_1, \lambda_2 \in \mathbb{R}$.

iii. Density preservation: $S_{\boldsymbol{\theta}} f$ is a density.

iv. Rotational symmetry: $S_{\boldsymbol{\theta}} f$ is rotationally symmetric around $\boldsymbol{\theta}$. If f is rotational symmetric around $\boldsymbol{\theta}$, then $S_{\boldsymbol{\theta}} f = f$.

v. For the von Mises density,

$$S_{\boldsymbol{\theta}} f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = \frac{C_q(\kappa)\exp\left\{\kappa \mathbf{x}^T\boldsymbol{\theta}\boldsymbol{\theta}^T\boldsymbol{\mu}\right\}}{\omega_{q-1}C_{q-1}\left(\kappa\left[(1-(\mathbf{x}^T\boldsymbol{\theta})^2)(1-(\boldsymbol{\mu}^T\boldsymbol{\theta})^2)\right]^{\frac{1}{2}}\right)}.$$

Recall that if $\boldsymbol{\theta} = \pm\boldsymbol{\mu}$, then $S_{\boldsymbol{\theta}} f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa)$. If $\boldsymbol{\mu}^T\boldsymbol{\theta} = 0$, then $S_{\boldsymbol{\theta}} f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = \frac{C_q(\kappa)}{\omega_{q-1}C_{q-1}\left(\kappa(1-(\mathbf{x}^T\boldsymbol{\theta})^2)^{\frac{1}{2}}\right)}$.

vi. Particular case when $q = 1$: denoting $\boldsymbol{\theta} = (\theta_1, \theta_2)$ and $\boldsymbol{\theta}^\perp = (-\theta_2, \theta_1)$, then

$$S_{\boldsymbol{\theta}} f(\mathbf{x}) = \frac{1}{2}\left\{f\left((\mathbf{x}^T\boldsymbol{\theta})\boldsymbol{\theta} + (1-(\mathbf{x}^T\boldsymbol{\theta})^2)^{\frac{1}{2}}\boldsymbol{\theta}^\perp\right) + f\left((\mathbf{x}^T\boldsymbol{\theta})\boldsymbol{\theta} - (1-(\mathbf{x}^T\boldsymbol{\theta})^2)^{\frac{1}{2}}\boldsymbol{\theta}^\perp\right)\right\}.$$

From the rotsymmetrizer definition and its properties, it is easy to derive a *rotational kernel density estimator* intended to use when rotational symmetry holds. For a random sample $\{\mathbf{X}_i\}_{i=1}^n$ from a directional random variable, by applying the rotsymmetrizer over the usual kernel density estimator (see Chapter 3 or 4) it follows:

$$\hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x}) = S_{\boldsymbol{\theta}}\hat{f}_h(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} L_{h,\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X}_i), \tag{8.6}$$

where

$$L_{h,\boldsymbol{\theta}}\left(\mathbf{x},\mathbf{X}_i\right)=\frac{c_{h,q}(L)}{\omega_{q-1}}\int_{\Omega_{q-1}}L\left(\frac{1-\left((\mathbf{x}^T\boldsymbol{\theta})\boldsymbol{\theta}+(1-(\mathbf{x}^T\boldsymbol{\theta})^2)^{\frac{1}{2}}\mathbf{B}_{\boldsymbol{\theta}}\boldsymbol{\xi}\right)^T\mathbf{X}_i}{h^2}\right)\omega_{q-1}(d\boldsymbol{\xi}).$$

In addition, due to property v in Proposition 8.1, for the von Mises kernel

$$L_{h,\boldsymbol{\theta}}(\mathbf{x},\mathbf{X}_i)=\frac{C_q(1/h^2)\exp\left\{\mathbf{x}^T\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{X}_i\Big/h^2\right\}}{\omega_{q-1}C_{q-1}\Big(\left((1-(\mathbf{x}^T\boldsymbol{\theta})^2)(1-(\mathbf{X}_i^T\boldsymbol{\theta})^2)\right)^{\frac{1}{2}}\Big/h^2\Big)},$$

which provides a closed expression for the rotational kernel density estimator that does not require integration.

The estimator (8.6) can be regarded as a semiparametric estimator of the unknown rotational symmetric density, as it depends on the location parameter $\boldsymbol{\theta}$. Therefore, the two possibilities are that either $\boldsymbol{\theta}$ is known or unknown. For the last case, $\boldsymbol{\theta}$ should be estimated first, for example by the Maximum Likelihood Estimator (MLE) given by the directional sample mean if g is monotone (Duerinckx and Ley, 2012): $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}=\frac{\sum_{i=1}^n\mathbf{X}_i}{||\sum_{i=1}^n\mathbf{X}_i||}$. Future research will involve the development of an estimator that holds for a non monotone g and that can deal with the identifiability issue. Furthermore, bandwidth selection rules for this estimator can be obtained by adaptations of the classical Likelihood Cross Validation (LCV) and Least Squares Cross Validation (LSCV) rules:

$$h_{\mathrm{LSCV}}=\arg\min_{h>0}\mathrm{CV}_2(h),\quad \mathrm{CV}_2(h)=2n^{-1}\sum_{i=1}^n\hat{f}_{h,\boldsymbol{\theta}}^{-i}(\mathbf{X}_i)-\int_{\Omega_q}\hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x})^2\,\omega_q(d\mathbf{x}),$$

$$h_{\mathrm{LCV}}=\arg\max_{h>0}\mathrm{CV}_{\mathrm{KL}}(h),\quad \mathrm{CV}_{\mathrm{KL}}(h)=\sum_{i=1}^n\log\hat{f}_{h,\boldsymbol{\theta}}^{-i}(\mathbf{X}_i).$$

The bias of the estimator (8.6) follows under similar regularity assumptions that the ones used in Chapters 3 and 6 (see also Appendix A), yielding

$$\mathbb{E}\left[\hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x})\right]=S_{\boldsymbol{\theta}}f(\mathbf{x})+\frac{b_q(L)}{q}\mathrm{tr}\left[S_{\boldsymbol{\theta}}\boldsymbol{\mathcal{H}}f(\mathbf{x})\right]h^2+o\left(h^2\right),$$

uniformly in $\mathbf{x}\in\Omega_q$. If f is rotational symmetric around $\boldsymbol{\theta}$, then the bias is the same that for the usual kernel density estimator, since in that case $S_{\boldsymbol{\theta}}f(\mathbf{x})=f(\mathbf{x})$ and

$$S_{\boldsymbol{\theta}}\boldsymbol{\mathcal{H}}f(\mathbf{x})=S_{\boldsymbol{\theta}}\left(g''(\mathbf{x}^T\boldsymbol{\theta})\right)\boldsymbol{\theta}\boldsymbol{\theta}^T=g''(\mathbf{x}^T\boldsymbol{\theta})\boldsymbol{\theta}\boldsymbol{\theta}^T=\boldsymbol{\mathcal{H}}f(\mathbf{x}).$$

The main advantage of the rotational kernel density estimator is the improvement in variance, which is conjectured to be reduced by a factor of $\frac{1}{\omega_{q-1}}$ with respect to the original kernel density estimator. The derivation of the variance expression, as well as the bias and variance when $\boldsymbol{\theta}$ is estimated, will constitute the next steps on the theory development of this estimator. An idea of the performance of the rotational estimator can be seen in Figure 8.3.

Finally, it is worth to remark that rotational symmetry assumption has played an important role in the development of new methodology with directional data, as it can be seen for example

in Ley et al. (2013), Ley et al. (2014) and Ley and Verdebout (2014). Therefore, it is natural to ask whether this assumption holds or not for a directional random variable, to see if these new methodological tools (including the rotational kernel density estimator itself) can be applied. An interesting future work will be to test the null hypothesis of rotational symmetry

$$H_0 : f \in \mathcal{R} = \left\{ g(\cdot^T \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Omega_q, g : [-1,1] \longrightarrow [0,\infty), \int_{-1}^{1} \omega_{q-1} g(t)(1-t^2)^{\frac{q}{2}-1} = 1 \right\},$$

with $\boldsymbol{\theta}$ either known or unknown. Following previous approaches, a possible test statistic from a random sample could be

$$R_n = \int_{\Omega_q} \left(\hat{f}_h(\mathbf{x}) - \hat{f}_{h,\hat{\boldsymbol{\theta}}}(\mathbf{x}) \right)^2 \omega_q(d\mathbf{x}).$$

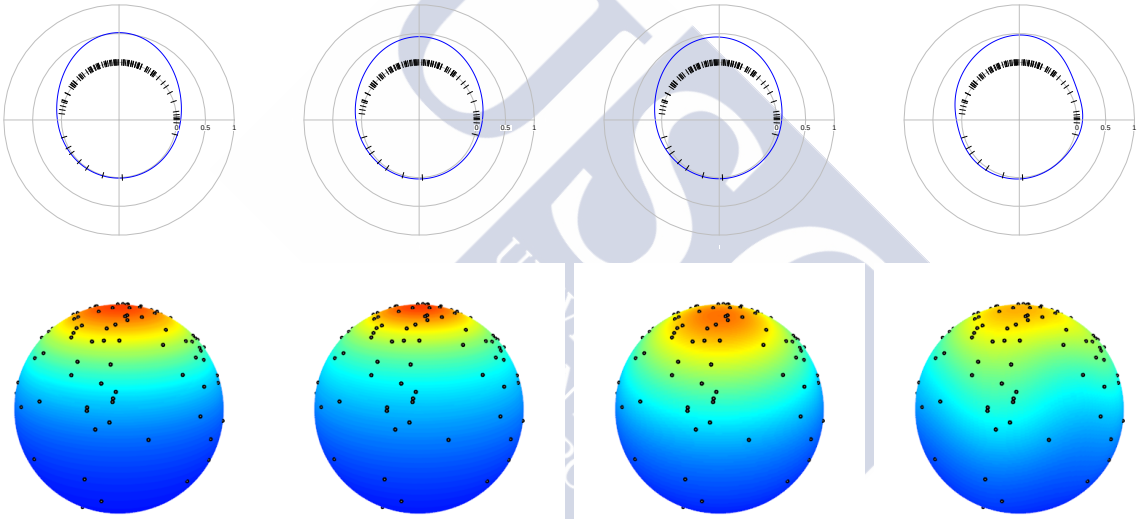


Figure 8.3: From right to left: vM$(\boldsymbol{\theta}, \kappa)$ density, rotational kernel density estimators $\hat{f}_{h_{\mathrm{LCV}},\boldsymbol{\theta}}$ and $\hat{f}_{h_{\mathrm{LCV}},\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}}$ and usual kernel density estimator $\hat{f}_{h_{\mathrm{LCV}}}$. Sample size is $n = 100$, $\boldsymbol{\theta} = (\mathbf{0}_q, 1)$ and $\kappa = 2$.

A decomposition of this statistic for studying its asymptotic distribution could be the following:

$$\begin{aligned}
R_n &= \int_{\Omega_q} \left(\hat{f}_h(\mathbf{x}) - L_h f(\mathbf{x}) + L_{h,\boldsymbol{\theta}} f(\mathbf{x}) - \hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x}) + \hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x}) - \hat{f}_{h,\hat{\boldsymbol{\theta}}}(\mathbf{x}) \right)^2 \omega_q(d\mathbf{x}) \\
&= \int_{\Omega_q} \left(\hat{f}_h(\mathbf{x}) - L_h f(\mathbf{x}) \right)^2 \omega_q(d\mathbf{x}) + \int_{\Omega_q} \left(L_{h,\boldsymbol{\theta}} f(\mathbf{x}) - \hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x})(\mathbf{x}) \right)^2 \omega_q(d\mathbf{x}) \\
&\quad + \int_{\Omega_q} \left(\hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x}) - \hat{f}_{h,\hat{\boldsymbol{\theta}}}(\mathbf{x}) \right)^2 \omega_q(d\mathbf{x}) \\
&\quad - 2 \int_{\Omega_q} \left(\hat{f}_h(\mathbf{x}) - L_h f(\mathbf{x}) \right) \left(L_{h,\boldsymbol{\theta}} f(\mathbf{x}) - \hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x})(\mathbf{x}) \right) \omega_q(d\mathbf{x}) \\
&\quad - 2 \int_{\Omega_q} \left(\hat{f}_h(\mathbf{x}) - L_h f(\mathbf{x}) \right) \left(\hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x}) - \hat{f}_{h,\hat{\boldsymbol{\theta}}}(\mathbf{x}) \right) \omega_q(d\mathbf{x}) \\
&\quad - 2 \int_{\Omega_q} \left(L_{h,\boldsymbol{\theta}} f(\mathbf{x}) - \hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x})(\mathbf{x}) \right) \left(\hat{f}_{h,\boldsymbol{\theta}}(\mathbf{x}) - \hat{f}_{h,\hat{\boldsymbol{\theta}}}(\mathbf{x}) \right) \omega_q(d\mathbf{x})
\end{aligned}$$

$$= R_{n,1} + R_{n,2} + R_{n,3} + R_{n,4} + R_{n,5} + R_{n,6},$$

where the smoothing operators L_h and $L_{h,\boldsymbol{\theta}}$ are included to remove asymptotic bias and are defined as

$$L_h f(\mathbf{x}) = \int_{\Omega_q} L_h(\mathbf{x},\mathbf{y}) f(\mathbf{y})\, \omega_q(d\mathbf{y}), \quad L_{h,\boldsymbol{\theta}} f(\mathbf{x}) = \int_{\Omega_q} L_{h,\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}) f(\mathbf{y})\, \omega_q(d\mathbf{y}),$$

where $L_h(\mathbf{x},\mathbf{y}) = c_{h,q}(L) L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right)$ and $L_{h,\boldsymbol{\theta}}$ was defined in (8.6). For understanding the previous splitting, it is important to note that $L_{h,\boldsymbol{\theta}} f(\mathbf{x}) = L_h(S_{\boldsymbol{\theta}} f(\mathbf{x})) \overset{H_0}{=} L_h f(\mathbf{x})$, $\forall \mathbf{x} \in \Omega_q$.

The asymptotics of $R_{n,1}$ were given by Zhao and Wu (2001), whereas it is expected that $R_{n,2}$ presents a similar limit distribution to $R_{n,1}$ and that $R_{n,4}$ contributes on the asymptotic bias. On the other hand, $R_{n,3}$, $R_{n,5}$ and $R_{n,6}$ are very likely to be negligible in probability if $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$ holds (without requiring that g is monotone). The answers to these unknowns and the design of a bootstrap mechanism to calibrate the test statistic under H_0 will constitute a new research line.

8.3 R package DirStats

All the methods proposed in the thesis have been implemented in R (R Development Core Team, 2014) using calls to FORTRAN to speed up critical computations, tested in simulation studies and applied to real data examples. As the source codes of these methods have not been made public yet, an obvious future step is to collect, organize and document the code in an R package. There exist libraries to analyse circular data in R, such as CircStats (Lund and Agostinelli, 2012), circular (Agostinelli and Lund, 2013), CircNNTSR (Fernández-Durán and Gregorio-Domínguez, 2013), isocir (Barragán et al., 2013) and NPCirc (Oliveira et al., 2014). Since these libraries are only designed to handle circular data, the focus of the DirStats package will be on basic and advanced methods to deal with directional and linear data.

A preliminary skeleton of the DirStats package, organized by topics, is given below:

A. **Class dir and related functions**. A new class named dir, intended for directional data, will be the basis for all the methods employing a directional input. The class will verify if the squared norm of the elements is indeed one, store the dimension, sample size and information about the coordinate system. In addition to the typical functions linked to a class, others like dir.to.rad, rad.to.cir, rad.to.sph, sph.to.aitoff and aitoff.to.sph will allow to make conversions between the most important coordinate systems.

B. **Descriptive statistics**. Basic descriptive measures given in Mardia and Jupp (2000) such as the directional mean (mean.dir), the mean resultant length (R.dir) and the moment of inertia (MI.dir) will be included. The function summary.dir will collect these statistics and maybe more advanced ones, such as the quantiles for rotational symmetric directional data described in Ley et al. (2014).

C. **Graphical functions**. The main graphic function in this topic will be plot.dir, that creates representations for circular, spherical, cylindrical and toroidal data from modifications of codes coming from the plotrix (Lemon, 2006) and rgl (Adler et al., 2014)

packages. Second level graphical functions will be `points.dir` and `lines.dir` to add information in any of the four possible manifolds. The function `jitter.dir` will allow to avoid plotting repeated points in the graphics by perturbing the data in any of the manifolds.

D. **Nonparametric estimation tools**. The estimation procedures are divided in two parts considering the density and the regression functions.

D.1. **Density estimation**. Kernel density estimation on Ω_q (`density.dir` with a flag `rot=TRUE` to account for the estimator of Section 8.2), on $\Omega_q \times \mathbb{R}$ (`density.dir.lin`) and on $\Omega_{q_1} \times \Omega_{q_2}$ (`density.dir.dir`) will be the key functions in this part. Bandwidth selection rules for the directional kernel density estimator (see Chapter 4) will be performed by `bw.dir`, with option `type=c("EMI","AMI","LCV","LSCV","RT")` for choosing a selector from the ones given in Chapter 4 and flag `rot=TRUE` to account for the rotational versions. Cross-validation bandwidths for the other situations will be available throughout the functions `bw.dir.lin` and `bw.dir.dir`, both with option `type=c("LCV","LSCV","RT")`.

D.2. **Regression estimation**. The function `locreg.lin.dir` will implement the projected local regression estimator of a linear variable Y on a directional predictor $\mathbf{X}$ given in Chapter 7, with `p=c(0,1)` specifying the type of local fit. Bandwidth selection will be done by means of function `bw.reg`, with `type=c("GCV","CV","RT","PI")` providing the kind of selector discussed in Section 8.1. It could be possible to implement also the regression estimator with directional response and linear (`locreg.dir.lin`) or directional (`locreg.dir.dir`) predictor, both proposed in Di Marzio et al. (2014).

E. **Nonparametric testing tools**. The testing methods are also dividend into two main subtopics, depending on the kind of curve that generates them.

E.1. **Tests for the density**. The goodness-of-fit tests for parametric models of Boente et al. (2014) (directional) and of Chapter 6 (directional-linear and directional-directional) will be included using the bootstrap resampling by `gof.dir.test`, `gof.dir.lin.test` and `gof.dir.dir.test`, respectively. The independence tests given in Chapter 5, based on permutations for calibration, will be available in `indep.dir.lin.test` and `indep.dir.dir.test`. A possible future inclusion is the test for rotational symmetry presented in Section 8.2 (`rotsym.test`). All the tests will be computed both in a deterministic grid of bandwidths (significance trace) and for data-driven bandwidths. The test for the Johnson and Wehrly copula structure for circular-linear and circular-circular random variables outlined in Section 8.4 could be added as the function `gof.jw`.

E.2. **Tests for the regression**. The goodness-of-fit tests for parametric regression functions (including the test for significance) that were studied in Chapter 7 will be accessible by the `gof.reg.test` function, with the option `p=c(0,1)` for controlling the kind of local estimator used in the fit. The test is performed in a grid of bandwidths and also for data-driven bandwidths.

F. **Parametric models**. Different collections of ready-to-use density and regression models, including simulation and fitting methods are included in this part.

F.1. **Density**. The collection of directional parametric models used in the simulation study of Chapter 4 will be provided by the functions `ddir` (density), `rdir` (simulation) and `fitdir` (fitting), with option `type=1:20` indicating the model. Similarly, the collections of circular-linear and circular-circular parametric densities used in Chapter 6 will be accessible via `dcirlin`, `dcircir`, `rcirlin`, `rcircir`, `fitcirlin` and `fitcircir`.

F.2. **Regression**. The collection of parametric regression models proposed in Section 8.1 will be implemented by the functions `freg` (regression function), `fitreg` (estimation of parameters) and `rreg` (simulation), all of them using option `type=1:12` to select the regression function. Simulation routine `rreg` will depend on functions `noise.dir.lin` and `rdir`.

G. **Datasets**. The datasets described in Chapter 1 will be included in the package or linked to the original source. Object `windSO2` will store the data from Chapter 2, containing the hourly measures of the wind direction and SO_2 concentration during 2004 and 2011. The data from Chapter 4 will be available throughout objects `wind` and `hipparcos`, that will either contain or download the Hipparcos dataset (van Leeuwen, 2007). The `fires` object will contain either the complete Portuguese wildfires dataset considered in Chapter 5 or the short version used in Chapter 6. Finally, the `slashdot` and `slashdot.dtm` objects will store the original dataset and the document term matrix used in Chapter 7. The protein dataset from Chapter 6 is available via `ProteinsAAA` in the `CircNNTSR` package.

H. **Auxiliary functions**. This part comprises, among others: numerical integration rules on Ω_q, $\Omega_q \times \mathbb{R}$ and $\Omega_{q_1} \times \Omega_{q_2}$; deterministic grids on the circle, sphere, cylinder and torus; functions to control multiple 3D devices; specific circular methods not available in other packages and included for comparison, such as the correlation coefficients and tests of Mardia (1976), Johnson and Wehrly (1977) and Fisher and Lee (1981) considered in Chapter 5.

8.4 A goodness-of-fit test for the Johnson and Wehrly copula structure

One of the starting points in this thesis was the Johnson and Wehrly (1978) model for circular-linear densities and the Wehrly and Johnson (1979) model for circular-circular densities, which were introduced in Chapter 2 and references therein, and recently studied by Jones et al. (2014). Essentially, this model assumes that the underlying copula (Nelsen, 2006) has a specific semiparametric structure, indexed by the set of all possible circular densities.

A natural question to ask is whether this semiparametric structure is suitable for modelling a given circular-linear or circular-circular variable. In the particular case where the marginal densities and link function g belong to certain parametric families this question can be answered by the goodness-of-fit tests given in Chapter 6. However, if these functions are not fixed, the interest is on checking if the relation of both variables could be modelled by a copula belonging to the Johnson and Wehrly family.

Denote by (X_1, X_2) any combination of circular and/or linear variables with copula C (see Chapter 2 for further details), by $U_j = F_j(X_j)$ the ranks of X_j, $j = 1, 2$, with F_j the corresponding

cumulative distribution function (cdf) and by $\mathcal{C}$ the class of copula functions. Testing the copula structure of Johnson and Wehrly is equivalent to test

$$H_{0,q} : C \in \mathcal{C}_{\mathcal{JW},q} = \left\{ C \in \mathcal{C} : \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} = 2\pi g\left(2\pi(u_1 + qu_2)\right), g \text{ a circular density} \right\}$$

being $q = \pm 1$ the index for the two possible kinds of correlation between U_1 and U_2: positive ($q = -1$) and negative ($q = 1$). Recall that a circular density $g : \mathbb{R} \longrightarrow [0, \infty)$ should satisfy that: *i)* $\int_0^{2\pi} g(\theta)\, d\theta = 1$ and *ii)* $g(\theta + 2k\pi) = g(\theta)$, $\forall \theta \in [0, 2\pi)$, $\forall k \in \mathbb{Z}$. Note also that independence (when g is the circular uniform density) is included in both $H_{0,-1}$ and $H_{0,1}$.

A possible approach to assess if H_0 holds is to consider the empirical version of the copula C and compare it with an empirical copula under H_0, leading to an empirical process (van der Vaart and Wellner, 1996). If weak convergence of this empirical process is proved, then the asymptotic distribution of continuous functionals such as the Kolmogorov-Smirnov and Cramér-von Mises ones will follow easily by the continuous mapping theorem, providing two suitable goodness-of-fit tests. The empirical copula defined from a sample $\{(X_{1i}, X_{2i})\}_{i=1}^n$ is given by

$$C_n(u_1, u_2) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{F_{1n}(X_{1i}) \leq u_1, F_{2n}(X_{2i}) \leq u_2\}},$$

where $F_{jn}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_{ji} \leq x\}}$ are the empirical cdf of X_j, $j = 1, 2$. It can be seen easily that C_n is not a copula (Nelsen, 2006).

Under $H_{0,q}$, the copula can be analytically computed from its density, yielding

$$\begin{aligned} C_G(u_1, u_2) &= \int_0^{u_1} \int_0^{u_2} 2\pi g\left(2\pi(s + qt)\right)\, dt\, ds \\ &= \frac{1}{q} \int_0^{u_1} G'(2\pi(s + qu_2)) - G'(2\pi s)\, ds \\ &= \frac{1}{2\pi q} \left(G(2\pi(u_1 + qu_2)) - G(2\pi u_1) - G(2\pi(qu_2)) + G(0)\right) \\ &= \frac{1}{2\pi q} \left(G(2\pi(u_1 + qu_2)) - G(2\pi u_1) - G(2\pi q u_2)\right), \end{aligned} \tag{8.7}$$

where $G : [-4\pi, 4\pi] \longrightarrow [0, \infty)$ is the function defined by

$$G(x) = \int_0^x \int_0^s g(t)\, dt\, ds,$$

this is, the integral from 0 to $x \in \mathbb{R}$ of the cdf associated to the circular density g. Based on properties of the circular density g, namely that $\int_0^s g(t + 2\pi k)\, dt = \int_0^s g(t)\, dt$ and $\int_0^{2\pi k} g(t)\, dt = k$ with $k \in \mathbb{Z}$, it happens that the function G satisfies

$$G(x + 2\pi k) = G(x) + G(2\pi k) + kx, \quad \forall x \in \mathbb{R}, \forall k \in \mathbb{Z}. \tag{8.8}$$

This is an important relation that helps in proving that C_G is indeed a copula.

Then, in view of the expression (8.7), an estimator from C_G will arise by plugging-in a suitable estimator G_n for G, which must satisfy relation (8.8). To account for this, recall another

important fact related with g: if the copula of (U_1, U_2) is C_G, then $2\pi(U_1 + qU_2)$ is a circular random variable with density g. This motivates an estimator for G by means of the integration of the empirical cdf on pseudo observations $\{\Theta_i\}_{i=1}^n$, where $\Theta_i = 2\pi\left(F_{1n}(X_{1i}) + qF_{2n}(X_{2i})\right) \mod 2\pi$, $i = 1, \ldots, n$, using the fact that

$$\int_0^x \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{\Theta_i \leq s\}}\, ds = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{\Theta_i \leq x\}}(x - \Theta_i).$$

Then, a possible estimator G_n for G that satisfies (8.8) can be constructed as

$$G_n(x) = \begin{cases} \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{\Theta_i \leq x\}}(x - \Theta_i), & x \in [0, 2\pi), \\ G_n(x - 2\pi) + G_n(2\pi) + x - 2\pi, & x \in [2\pi, 4\pi], \\ G_n(x + 2\pi) + G_n(-2\pi) - (x + 2\pi), & x \in [-2\pi, 0), \\ G_n(x + 4\pi) + G_n(-4\pi) - 2(x + 4\pi), & x \in [-4\pi, -2\pi], \end{cases}$$

where $G_n(-2\pi) = 2\pi - G_n(2\pi)$, $G_n(4\pi) = 2G_n(2\pi) + 2\pi$ and $G_n(-4\pi) = 8\pi - G_n(4\pi)$. The explanation of this estimator is the following: using the modulus 2π all the data $\{\Theta_i\}_{i=1}^n$ belongs to the interval $[0, 2\pi)$, where G is estimated by G_n and later extended to the intervals $[2\pi, 4\pi]$ and $[-4\pi, 0)$ using relation (8.8). The empirical copula C_{G_n} has two main advantages over C_n when $H_{0,q}$ holds: first, it is always a copula for any sample size n; second, it seems more efficient in estimating C_G (see Figure 8.4). Future research will involve the study of the properties of C_{G_n}, such as the bias and the variance.

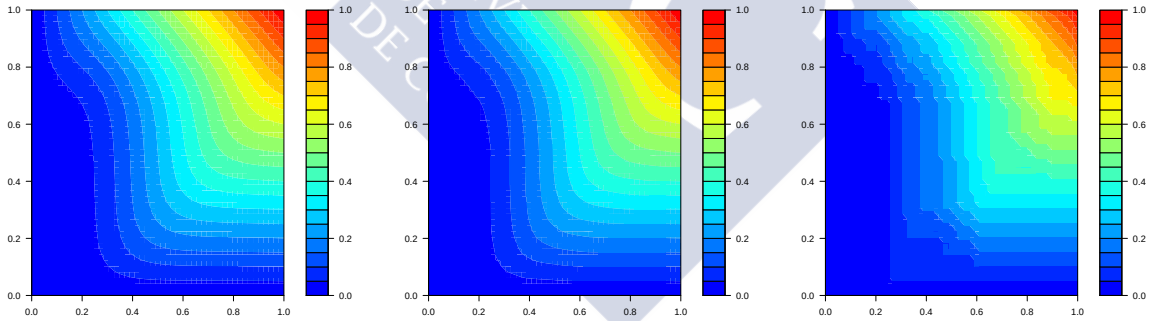


Figure 8.4: From right to left: contourplots for copula C_G, copula C_{G_n} and empirical copula C_n, for a sample size of $n = 50$ with $q = -1$ and $g(\theta) = (2\pi\mathcal{I}_0(\kappa))^{-1} e^{\kappa\cos(\theta-\mu)}$ a circular von Mises density with $\mu = \frac{\pi}{2}$ and $\kappa = 2$.

The empirical copula process for assessing $H_{0,q}$ can be written as

$$\begin{aligned} \mathbb{C}_n &= \sqrt{n}\,(C_n - C_{G_n}) \\ &= \sqrt{n}\,(C_n - C) + \sqrt{n}\,(C - C_G) + \sqrt{n}\,(C_G - C_{G_n}) \\ &= \mathbb{C}_{1n} + \mathbb{C}_{2n} + \mathbb{C}_{3n}. \end{aligned}$$

The weak convergence of $\mathbb{C}_{1n}$ was established by Fermanian et al. (2004), whereas $\mathbb{C}_{2n}$ is a deterministic bias term that vanishes if $H_{0,q}$ holds. The study of the weak convergence of $\mathbb{C}_{3n}$, its relation with $\mathbb{C}_{1n}$ and an effective resampling procedure for the calibration of the test statistic under $H_{0,q}$ is the main part of this future work.

References

Abe, T. and Pewsey, A. (2011). Sine-skewed circular distributions. *Statist. Papers*, 52(3):683–707.

Adler, D., Murdoch, D., and others (2014). *rgl: 3D visualization device system (OpenGL).* R package version 0.93.996.

Agostinelli, C. and Lund, U. (2013). *R package* `circular`*: circular statistics (version 0.4-7).*

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, 12(2):171–178.

Barragán, S., Fernández, M. A., Rueda, C., and Peddada, S. D. (2013). isocir: an R package for constrained inference using isotonic regression for circular data, with an application to cell biology. *J. Stat. Softw.*, 54(4):1–17.

Bingham, C. and Mardia, K. V. (1978). A small circle distribution on the sphere. *Biometrika*, 65(2):379–389.

Boente, G., Rodríguez, D., and González-Manteiga, W. (2014). Goodness-of-fit test for directional data. *Scand. J. Stat.*, 41(1):259–275.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2014). Nonparametric regression for spherical data. *J. Amer. Statist. Assoc.*, 109(506):748–763.

Duerinckx, M. and Ley, C. (2012). Maximum likelihood characterization of rotationally symmetric distributions on the sphere. *Sankhya A*, 74(2):249–262.

Fermanian, J.-D., Radulović, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–860.

Fernández-Durán, J. J. and Gregorio-Domínguez, M. M. (2013). *CircNNTSR: an R package for the statistical analysis of circular data using NonNegative Trigonometric Sums (NNTS) models.* R package version 2.1.

Fisher, N. I. and Lee, A. J. (1981). Nonparametric measures of angular-linear association. *Biometrika*, 68(3):629–636.

Härdle, W. and Marron, J. S. (1995). Fast and simple scatterplot smoothing. *Comput. Statist. Data Anal.*, 20(1):1–17.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43 of *Monographs on Statistics and Applied Probability.* Chapman & Hall, London.

Johnson, R. A. and Wehrly, T. (1977). Measures and models for angular correlation and angular-linear correlation. *J. Roy. Statist. Soc. Ser. B*, 39(2):222–229.

Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *J. Amer. Statist. Assoc.*, 73(363):602–606.

Jones, M. C., Pewsey, A., and Kato, S. (2014). On a class of circulas: copulas for circular distributions. *Ann. Inst. Statist. Math.*, to appear.

Lemon, J. (2006). Plotrix: a package in the red light district of R. *R-News*, 6(4):8–12.

Ley, C., Sabbah, C., and Verdebout, T. (2014). A new concept of quantiles for directional data and the angular Mahalanobis depth. *Electron. J. Stat.*, 8:795–816.

Ley, C., Swan, Y., Thiam, B., and Verdebout, T. (2013). Optimal R-estimation of a spherical location. *Statist. Sinica*, 23(1):305–332.

Ley, C. and Verdebout, T. (2014). Local powers of optimal one-sample and multi-sample tests for the concentration of Fisher-von Mises-Langevin distributions. *Int. Stat. Rev.*, to appear.

Lund, U. and Agostinelli, C. (2012). *CircStats: Circular Statistics, from "Topics in circular statistics" (2001)*. R package version 0.2-4.

Mardia, K. V. (1976). Linear-circular correlation coefficients and rhythmometry. *Biometrika*, 63(2):403–405.

Mardia, K. V. and Jupp, P. E. (2000). *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, second edition.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition.

Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2014). NPCirc: nonparametric circular methods. *J. Stat. Softw.*, to appear.

Pewsey, A. (2006). Modelling asymmetrically distributed circular data using the wrapped skew-normal distribution. *Environ. Ecol. Stat.*, 13(3):257–269.

Pukkila, T. M. and Rao, C. R. (1988). Pattern recognition based on scale invariant discriminant functions. *Inform. Sci.*, 45(3):379–389.

R Development Core Team (2014). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, 90(432):1257–1270.

Saw, J. G. (1978). A family of distributions on the m-sphere and some hypothesis tests. *Biometrika*, 65(1):69–73.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B*, 53(3):683–690.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes: With applications to statistics*. Springer Series in Statistics. Springer, New York.

van Leeuwen, F. (2007). *Hipparcos, the new reduction of the raw data*, volume 350 of *Astrophysics and Space Science Library*. Springer, Dordrecht.

Watson, G. S. (1983). *Statistics on spheres*, volume 6 of *University of Arkansas Lecture Notes in the Mathematical Sciences*. John Wiley & Sons, New York.

Wehrly, T. E. and Johnson, R. A. (1979). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, 67(1):255–256.

Zhao, L. and Wu, C. (2001). Central limit theorem for integrated square error of kernel estimators of spherical density. *Sci. China Ser. A*, 44(4):474–483.

Appendix A

Supplement to Chapter 6

This supplement is organized as follows. Section A.1 contains the detailed proofs of the required technical lemmas used to prove the main results in the paper. The section is divided into four subsections to classify the lemmas used in the CLT of the ISE, the independence test and the goodness-of-fit test, with an extra subsection for general purpose lemmas. Section A.2 presents closed expressions that can be used in the independence test, the extension of the results to the directional-directional situation and some numerical experiments to illustrate the convergence to the asymptotic distribution. Section A.3 describes in detail the simulation study of the goodness-of-fit test to allow its reproducibility: parametric models employed, estimation and simulation methods, the construction of the alternatives, the bandwidth choice and further results omitted in the paper. Finally, Section A.4 shows deeper insights on the real data application.

A.1 Technical lemmas

A.1.1 CLT for the ISE

Lemma A.1 presents a generalization of Theorem 1 in Hall (1984) for degenerate U-statistics that, up to the authors' knowledge, was first stated by Zhao and Wu (2001) under different conditions, but without providing a formal proof. This lemma, written under a general notation, is used to prove asymptotic convergence of the ISE when the variance is large relative to the bias ($n\phi(h,g)h^q g \to 0$) and when the bias is balanced with the variance ($n\phi(h,g)h^q g \to \delta$).

Lemma A.1. *Let $\{X_i\}_{i=1}^n$ be a sequence of independent and identically distributed random variables. Assume that $H_n(x,y)$ is symmetric in x and y,*

$$\mathbb{E}\left[H_n\left(X_1,X_2\right)|X_1\right]=0 \text{ almost surely and } \mathbb{E}\left[H_n^4(X_1,X_2)\right]<\infty,\ \forall n. \tag{A.1}$$

Define $G_n\left(x,y\right)=\mathbb{E}\left[H_n\left(x,X_1\right)H_n\left(y,X_1\right)\right]$ and φ_n, satisfying $\mathbb{E}\left[\varphi_n(X_1)\right]=0$ and $\mathbb{E}\left[\varphi_n^4(X_1)\right]<\infty$. Define also:

$$\begin{aligned}
M_n(X_1)&=\mathbb{E}\left[\varphi_n(X_2)H_n(X_1,X_2)|X_1\right],\\
A_n&=n\mathbb{E}\left[\varphi_n^4(X_1)\right]+n^2\mathbb{E}\left[M_n^2(X_1)\right]+n^3\mathbb{E}\left[H_n^4(X_1,X_2)\right]+n^4\mathbb{E}\left[G_n^2(X_1,X_2)\right],\\
B_n&=n\mathbb{E}\left[\varphi_n^2(X_1)\right]+\frac{1}{2}n^2\mathbb{E}\left[H_n^2(X_1,X_2)\right].
\end{aligned}$$

If $A_nB_n^{-2}\to 0$ as $n\to\infty$ and $U_n=\sum_{i=1}^n\varphi_n(X_i)+\sum_{1\le i<j\le n}H_n\left(X_i,X_j\right)$,

$$B_n^{-\frac{1}{2}}U_n\xrightarrow{d}\mathcal{N}(0,1).$$

Note that when $\varphi_n \equiv 0$, U_n is an U-statistic and Theorem 1 in Hall (1984) is a particular case of Lemma A.1.

Proof of Lemma A.1. To begin with, let consider the sequence of random variables $\{Y_{n_i}\}_{i=1}^n$, defined by

$$Y_{n_i} = \begin{cases} \varphi_n(X_1), & i = 1, \\ \varphi_n(X_i) + \sum_{j=1}^{i-1} H_n(X_i, X_j), & 2 \leq i \leq n. \end{cases}$$

This sequence generates a martingale $S_i = \sum_{j=1}^{i} Y_{n_j}$, $1 \leq i \leq n$ with respect to the sequence of random variables $\{X_i\}_{i=1}^n$, with differences Y_{n_i} and with $S_n = U_n$. To see that $S_i = \sum_{j=1}^{i} Y_{n_j}$, $1 \leq i \leq n$ is indeed a martingale with respect to $\{X_i\}_{i=1}^n$, recall that

$$\begin{aligned} \mathbb{E}\left[S_{i+1}|X_1, \ldots, X_i\right] &= \sum_{j=1}^{i+1} \mathbb{E}\left[\varphi_n(X_j)|X_1, \ldots, X_i\right] + \sum_{j=1}^{i+1} \sum_{k=1}^{j-1} \mathbb{E}\left[H_n(X_j, X_k)|X_1, \ldots, X_i\right] \\ &= \sum_{j=1}^{i} \varphi_n(X_j) + \sum_{j=1}^{i} \sum_{k=1}^{j-1} H_n(X_j, X_k) \\ &= S_i \end{aligned}$$

because of the null expectations of $\mathbb{E}\left[\varphi_n(X)\right]$ and $\mathbb{E}\left[H_n\left(X_1, X_2\right)|X_1\right]$.

The main idea of the proof is to apply the martingale CLT of Brown (1971) (see also Theorem 3.2 of Hall and Heyde (1980)), in the same way as Hall (1984) did for the particular case where $\varphi_n \equiv 0$. Theorem 2 of Brown (1971) ensures that if the conditions

C1. $\lim_{n\to\infty} s_n^{-2} \sum_{i=1}^{n} \mathbb{E}\left[Y_{n_i}^2 \mathbf{1}_{\{|Y_{n_i}| > \varepsilon s_n\}}\right] = 0$, $\forall \varepsilon > 0$,

C2. $s_n^{-2} V_n^2 \xrightarrow{p} 1$,

are satisfied, with $s_n^2 = \mathbb{E}\left[U_n^2\right]$ and $V_n^2 = \sum_{i=1}^{n} \mathbb{E}[Y_{n_i}^2|X_1, \ldots, X_{i-1}]$, then $s_n^{-1} U_n \xrightarrow{d} \mathcal{N}(0,1)$. The aim of this proof is to prove separately both conditions. From now on, expectations will be taken with respect to the random variables $X_1, \ldots, X_n$, except otherwise is stated.

*Proof of **C1**.* The key idea is to give bounds for $\mathbb{E}[Y_{n_i}^4]$ and prove that $s_n^{-4} \sum_{i=1}^{n} \mathbb{E}[Y_{n_i}^4] \to 0$ as $n \to \infty$. In that case, the Lindenberg's condition **C1** follows immediately:

$$\begin{aligned} \lim_{n\to\infty} s_n^{-2} \sum_{i=1}^{n} \mathbb{E}\left[Y_{n_i}^2 \mathbf{1}_{\{|Y_{n_i}| > \varepsilon s_n\}}\right] &\leq \lim_{n\to\infty} s_n^{-2} \sum_{i=1}^{n} \mathbb{E}\left[Y_{n_i}^4 \varepsilon^{-2} s_n^{-2} \times 1\right] \\ &= \varepsilon^{-2} \lim_{n\to\infty} s_n^{-4} \sum_{i=1}^{n} \mathbb{E}\left[Y_{n_i}^4\right] \\ &= 0. \end{aligned}$$

In order to compute $s_n^2 = \mathbb{E}\left[U_n^2\right]$, it is needed

$$\mathbb{E}\left[Y_{n_i}^2\right] = \begin{cases} \mathbb{E}\left[\varphi_n^2(X_1)\right], & i = 1, \\ \mathbb{E}\left[\varphi_n^2(X_i)\right] + (i-1)\mathbb{E}\left[H_n^2(X_1, X_2)\right], & 2 \leq i \leq n, \end{cases}$$

where the second case holds because the independence of the variables, the tower property of the conditional expectation and (A.1) ensure that

$$\mathbb{E}\left[\varphi_n(X_1)H_n(X_1,X_2)\right] = \mathbb{E}\left[H_n(X_1,X_2)H_n(X_1,X_3)\right] = 0.$$

Using these relations and the null expectation of $\varphi_n(X_1)$, it follows that for $j \neq k$,

$$\begin{aligned}\mathbb{E}\left[Y_{n_j}Y_{n_k}\right] =& \mathbb{E}\left[\varphi_n(X_j)\right]\mathbb{E}\left[\varphi_n(X_k)\right] + \sum_{l=1}^{k-1}\mathbb{E}\left[\varphi_n(X_j)H_n(X_k,X_l)\right] \\ &+ \sum_{m=1}^{j-1}\mathbb{E}\left[\varphi_n(X_k)H_n(X_j,X_m)\right] + \sum_{l=1}^{k-1}\sum_{m=1}^{j-1}\mathbb{E}\left[H_n(X_k,X_l)H_n(X_j,X_m)\right] \\ =& 0.\end{aligned}$$

Then:

$$s_n^2 = n\mathbb{E}\left[\varphi_n^2(X_1)\right] + \sum_{j=1}^{n}(j-1)\mathbb{E}\left[H_n^2(X_1,X_2)\right] = \mathcal{O}\left(B_n\right). \tag{A.2}$$

On the other hand,

$$\begin{aligned}\mathbb{E}\left[Y_{n_i}^4\right] =& \mathbb{E}\left[\left(\varphi_n(X_i) + \sum_{j=1}^{i-1}H_n(X_i,X_j)\right)^4\right] \\ =& \mathcal{O}\left(\mathbb{E}\left[\varphi_n^4(X_i)\right]\right) + \mathcal{O}\left(\mathbb{E}\left[\left(\sum_{j=1}^{i-1}H_n(X_i,X_j)\right)^4\right]\right) \\ =& \mathcal{O}\left(\mathbb{E}\left[\varphi_n^4(X_1)\right]\right) + (i-1)\mathcal{O}\left(\mathbb{E}\left[H_n^4(X_1,X_2)\right]\right) \\ &+ 3(i-1)(i-2)\mathcal{O}\left(\mathbb{E}\left[H_n^2(X_1,X_2)H_n^2(X_1,X_3)\right]\right),\end{aligned}$$

where the equalities are true in virtue of Lemma A.12 and because

$$\mathbb{E}\left[H_n\left(X_1,X_2\right)H_n\left(X_1,X_3\right)H_n\left(X_1,X_4\right)H_n\left(X_1,X_5\right)\right] = \mathbb{E}\left[H_n^3\left(X_1,X_2\right)H_n\left(X_1,X_3\right)\right] = 0.$$

Finally,

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{E}\left[Y_{n_i}^4\right] =& n\mathcal{O}\left(\mathbb{E}\left[\varphi_n^4(X_1)\right]\right) + \frac{1}{2}n(n-1)\mathcal{O}\left(\mathbb{E}\left[H_n^4(X_1,X_2)\right]\right) \\ &+ (n^3-n)\mathcal{O}\left(\mathbb{E}\left[G_n^2(X_1,X_2)\right]\right) \\ =& \mathcal{O}\left(A_n\right).\end{aligned} \tag{A.3}$$

Then, joining (A.2) and (A.3),

$$s_n^{-4}\sum_{i=1}^{n}\mathbb{E}\left[Y_{n_i}^4\right] = \mathcal{O}\left(B_n^{-2}A_n\right) \xrightarrow[n\to\infty]{} 0$$

and **C1** is satisfied.

*Proof of **C2**.* Now it is proved the convergence in squared mean of $s_n^{-2}V_n^2$ to 1, which implies that $s_n^{-2}V_n^2 \xrightarrow{p} 1$, by obtaining bounds for $\mathbb{E}\left[V_n^4\right]$.

First of all, let denote $V_n^2 = \sum_{i=1}^n \nu_{n_i}$, where

$$\begin{aligned}
\nu_{n_i} =& \mathbb{E}\left[Y_{n_i}^2 | X_1, \ldots, X_{i-1}\right] \\
=& \mathbb{E}\left[\varphi_n^2(X_i) + 2\varphi_n(X_i)\sum_{j=1}^{i-1} H_n(X_i, X_j) + \sum_{j=1}^{i-1}\sum_{k=1}^{i-1} H_n(X_i, X_j)H_n(X_i, X_k)\middle| X_1, \ldots, X_{i-1}\right] \\
=& \mathbb{E}\left[\varphi_n^2(X_i)\right] + 2\sum_{j=1}^{i-1} M_n(X_j) + \sum_{j=1}^{i-1}\sum_{k=1}^{i-1} \mathbb{E}\left[H_n(X_i, X_j)H_n(X_i, X_k)|X_j, X_k\right] \\
=& \mathbb{E}\left[\varphi_n^2(X_1)\right] + 2\sum_{j=1}^{i-1} M_n(X_j) + \sum_{j=1}^{i-1} G_n(X_j, X_j) + 2\sum_{1\leq j<k\leq i-1} G_n(X_j, X_k).
\end{aligned}$$

Using Lemma A.12, the Jensen inequality and that for $j_1 \leq k_1$, $j_2 \leq k_2$,

$$\mathbb{E}\left[G_n(X_{j_1}, X_{k_1})G_n(X_{j_2}, X_{k_2})\right] = \begin{cases} \mathbb{E}\left[G_n^2(X_1, X_1)\right], & j_1 = k_1 = j_2 = k_2, \\ \mathbb{E}\left[G_n(X_1, X_1)\right]^2, & j_1 = k_1 \neq j_2 = k_2, \\ \mathbb{E}\left[G_n^2(X_1, X_2)\right], & j_1 = j_2 < k_1 = k_2, \\ 0, & \text{otherwise}, \end{cases}$$

it follows:

$$\begin{aligned}
\mathbb{E}\left[\nu_{n_i}^2\right] =& \mathcal{O}\left(\mathbb{E}\left[\varphi_n^2(X_1)\right]^2\right) + \sum_{j=1}^{i-1}\sum_{k=1}^{i-1} \mathcal{O}\left(\mathbb{E}\left[M_n(X_j)M_n(X_k)\right]\right) \\
&+ \sum_{j=1}^{i-1}\sum_{k=1}^{i-1} \mathcal{O}\left(\mathbb{E}\left[G_n(X_j, X_j)G_n(X_k, X_k)\right]\right) + \sum_{\substack{1\leq j_1<k_1\leq i-1 \\ 1\leq j_2<k_2\leq i-1}} \mathcal{O}\left(\mathbb{E}\left[G_n(X_{j_1}, X_{k_1})G_n(X_{j_2}, X_{k_2})\right]\right) \\
=& \mathcal{O}\left(\mathbb{E}\left[\varphi_n^4(X_1)\right]\right) + (i-1)\mathcal{O}\left(\mathbb{E}\left[M_n^2(X_1)\right]\right) + (i-1)(i-2)\mathcal{O}\left(\mathbb{E}\left[M_n(X_1)\right]^2\right) \\
&+ (i-1)\mathcal{O}\left(\mathbb{E}\left[G_n^2(X_1, X_1)\right]\right) + (i-1)(i-2)\mathcal{O}\left(\mathbb{E}\left[G_n(X_1, X_1)\right]^2\right) \\
&+ (i-1)(i-2)\mathcal{O}\left(\mathbb{E}\left[G_n^2(X_1, X_2)\right]\right).
\end{aligned}$$

Applying again the Lemma A.12,

$$\mathbb{E}\left[V_n^4\right] = \mathbb{E}\left[\left(\sum_{i=1}^n \nu_{n_i}\right)^2\right] = \sum_{i=1}^n \mathcal{O}\left(\mathbb{E}\left[\nu_{n_i}^2\right]\right).$$

By the two previous calculus and bearing in mind that $\mathbb{E}\left[G_n(X_1, X_1)\right]^2 = \mathcal{O}\left(\mathbb{E}\left[H_n^4(X_1, X_2)\right]\right)$ (by the Cauchy-Schwartz inequality) and $\mathbb{E}\left[M_n(X_1)\right] = 0$ (by the tower property), it yields:

$$\begin{aligned}
\mathbb{E}\left[V_n^4\right] =& n\mathcal{O}\left(\mathbb{E}\left[\varphi_n^4(X_1)\right]\right) + n(n-1)\mathcal{O}\left(\mathbb{E}\left[M_n^2(X_1)\right]\right) \\
&+ n(n-1)(n-3)\mathcal{O}\left(\mathbb{E}\left[H_n^4(X_1, X_2)\right]\right) \\
&+ n(n-1)(n-3)\mathcal{O}\left(\mathbb{E}\left[G_n^2(X_1, X_2)\right]\right) \\
=& \mathcal{O}\left(A_n\right).
\end{aligned}$$

Then, using the bound for $\mathbb{E}\left[V_n^4\right]$, that $s_n^2 = B_n$ and that $\mathbb{E}\left[V_n^2\right] = s_n^2$, it results

$$\mathbb{E}\left[\left(s_n^{-2}V_n^2 - 1\right)^2\right] = s_n^{-4}\mathbb{E}\left[\left(V_n^2 - s_n^2\right)^2\right] = s_n^{-4}\left(\mathbb{E}\left[V_n^4\right] - s_n^4\right) \leq s_n^{-4}\mathbb{E}\left[V_n^4\right] = \mathcal{O}\left(B_n^{-2}A_n\right).$$

Then $s_n^{-2}V_n^2$ converges to 1 in squared mean, which implies $s_n^{-2}V_n^2 \xrightarrow{p} 1$. □

Lemma A.2. *Under conditions **A1**–**A3**,*

$$n^{\frac{1}{2}}\phi(h,g)^{-\frac{1}{2}}I_{n,1} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof of Lemma A.2. The asymptotic normality of $I_{n,1} = \sum_{i=1}^n I_{n,1}^{(i)}$ will be derived checking the Lindenberg's condition. To that end, it is needed to prove the following relations:

$$\begin{aligned}\mathbb{E}\left[I_{n,1}^{(i)}\right] = 0, \quad \mathbb{E}\left[\left(I_{n,1}^{(i)}\right)^2\right] = n^{-2}\phi(h,g)(1+o(1)),\\ \mathbb{E}\left[\left(I_{n,1}^{(i)}\right)^4\right] = \mathcal{O}\left(n^{-4}(h^8+g^8)\right), \quad s_n^4 = \mathcal{O}\left(n^{-2}(h^8+g^8)\right),\end{aligned}$$

where $s_n^2 = \sum_{i=1}^n \mathbb{E}\left[\left(I_{n,1}^{(i)}\right)^2\right]$ and $\phi(h,g)$ is defined as in Theorem 6.1. If these relations hold, the Lindenberg's condition

$$\lim_{n\to\infty} s_n^{-2}\sum_{i=1}^n \mathbb{E}\left[\left(I_{n,1}^{(i)}\right)^2 \mathbf{1}_{\left\{\left|I_{n,1}^{(i)}\right| > \varepsilon s_n\right\}}\right] = 0, \quad \forall \varepsilon > 0$$

is satisfied:

$$\begin{aligned}s_n^{-2}\sum_{i=1}^n \mathbb{E}\left[\left(I_{n,1}^{(i)}\right)^2 \mathbf{1}_{\left\{\left|I_{n,1}^{(i)}\right| > \varepsilon s_n\right\}}\right] &\leq \sum_{i=1}^n \mathbb{E}\left[\left(I_{n,1}^{(i)}\right)^4 \varepsilon^{-2}s_n^{-4} \times 1\right]\\ &= \varepsilon^{-2}n\mathbb{E}\left[\left(I_{n,1}^{(i)}\right)^4\right]\mathcal{O}\left(n^2(h^8+g^8)^{-1}\right)\\ &= \varepsilon^{-2}\mathcal{O}\left(n^{-1}\right).\end{aligned}$$

Therefore $s_n^{-1}I_{n,1} \xrightarrow{d} \mathcal{N}(0,1)$, which, by Slutsky's theorem, implies that

$$n^{\frac{1}{2}}\phi(h,g)^{-\frac{1}{2}}I_{n,1} \xrightarrow{d} \mathcal{N}(0,1).$$

In order to prove the moment relations for $I_{n,1}^{(i)}$ and bearing in mind the smoothing operator (6.4), let denote

$$\begin{aligned}\widetilde{I}_{n,1}^{(i)} &= 2\frac{c_{h,q}(L)}{ng}\int_{\Omega_q\times\mathbb{R}} LK\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}, \frac{z-Z_i}{g}\right)\left(\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x},z)\right] - f(\mathbf{x},z)\right) dz\,\omega_q(d\mathbf{x}),\\ &= 2n^{-1}LK_{h,g}\left(\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{X}_i,Z_i)\right] - f(\mathbf{X}_i,Z_i)\right)\end{aligned}$$

so that $I_{n,1}^{(i)} = \widetilde{I}_{n,1}^{(i)} - \mathbb{E}[\widetilde{I}_{n,1}^{(i)}]$. Therefore, $\mathbb{E}[I_{n,1}^{(i)}] = 0$ and $\widetilde{I}_{n,1}^{(i)}$ can be decomposed in two addends by virtue of Lemma A.11:

$$\begin{aligned}\widetilde{I}_{n,1}^{(i)} &= 2n^{-1}LK_{h,g}\left(\frac{b_q(L)}{q}\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}}f(\mathbf{X}_i,Z_i)\right]h^2 + \frac{\mu_2(K)}{2}\mathcal{H}_z f(\mathbf{X}_i,Z_i)g^2\right) + o\left(n^{-1}(h^2+g^2)\right)\\ &= \widetilde{I}_{n,1}^{(i,1)} + \widetilde{I}_{n,1}^{(i,2)} + o\left(n^{-1}(h^2+g^2)\right),\end{aligned}$$

where $\widetilde{I}_{n,1}^{(i,j)} = \delta_j LK_{h,g}\varphi_j(f, \mathbf{X}_i, Z_i)$ and

$$\varphi_j(f,\mathbf{x},z) = \begin{cases} \operatorname{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x},z)\right], & j=1, \\ \mathcal{H}_z f(\mathbf{x},z), & j=2, \end{cases} \quad \delta_j = \begin{cases} \frac{2b_q(L)}{q} h^2 n^{-1}, & j=1, \\ \mu_2(K) g^2 n^{-1}, & j=2. \end{cases}$$

Note that as the order $o\left(h^2+g^2\right)$ is uniform in $(\mathbf{x},z) \in \Omega_q \times \mathbb{R}$, then it is possible to extract it from the integrand of $\widetilde{I}_{n,1}^{(i)}$. Applying Lemma A.10 to the functions $\varphi_j(f,\cdot,\cdot)$, that by **A1** are uniformly continuous and bounded, it yields $LK_{h,g}\varphi_j(f,\mathbf{y},t) \to \varphi_j(f,\mathbf{y},t)$ uniformly in $(\mathbf{y},t) \in \Omega_q \times \mathbb{R}$ as $n \to \infty$. So, for any integers k_1 and k_2:

$$\begin{aligned} \lim_{n\to\infty} \delta_1^{-k_1}\delta_2^{-k_2}\mathbb{E}\left[\left(\widetilde{I}_{n,1}^{(i,1)}\right)^{k_1}\left(\widetilde{I}_{n,1}^{(i,2)}\right)^{k_2}\right] & \\ = \lim_{n\to\infty} \int_{\Omega_q\times\mathbb{R}} \left(LK_{h,g}\varphi_1(f,\mathbf{y},t)\right)^{k_1} \left(LK_{h,g}\varphi_2(f,\mathbf{y},t)\right)^{k_2} f(\mathbf{y},t)\, dt\, \omega_q(d\mathbf{y}) & \\ = \int_{\Omega_q\times\mathbb{R}} \varphi_1(f,\mathbf{y},t)^{k_1}\varphi_2(f,\mathbf{y},t)^{k_2} f(\mathbf{y},t)\, dt\, \omega_q(d\mathbf{y}) & \\ = \mathbb{E}\left[\varphi_1(f,\mathbf{X},Z)^{k_1}\varphi_2(f,\mathbf{X},Z)^{k_2}\right]. & \end{aligned}$$

Here the limit can commute with the integral by the Dominated Convergence Theorem (DCT), since the functions $\left(LK_{h,g}\varphi_j(f,\mathbf{y},t)\right)^k$ are bounded by condition **A1** and the construction of the smoothing operator (6.4), being this dominating function integrable:

$$\left(LK_{h,g}\varphi_1(f,\mathbf{y},t)\right)^{k_1}\left(LK_{h,g}\varphi_2(f,\mathbf{y},t)\right)^{k_2} f(\mathbf{y},t) \leq \sup_{(\mathbf{x},z)\in\Omega_q\times\mathbb{R}} \left|\varphi_1(f,\mathbf{x},z)^{k_1}\varphi_2(f,\mathbf{x},z)^{k_2}\right| f(\mathbf{y},t).$$

Recapitulating, the relation obtained is:

$$\mathbb{E}\left[\left(\widetilde{I}_{n,1}^{(i,1)}\right)^{k_1}\left(\widetilde{I}_{n,1}^{(i,2)}\right)^{k_2}\right] \sim 2^{k_1} n^{-(k_1+k_2)} \frac{b_q(L)^{k_1}}{q^{k_1}} \mu_2(K)^{k_2} h^{2k_1} g^{2k_2} \mathbb{E}\left[\operatorname{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}}(f,\mathbf{X},Z)\right]^{k_1} \mathcal{H}_z f(\mathbf{X},Z)^{k_2}\right].$$

Now it is easy to prove:

$$\begin{aligned} \mathbb{E}\left[\left(I_{n,1}^{(i)}\right)^2\right] \sim & \mathbb{E}\left[\left(\widetilde{I}_{n,1}^{(i,1)} + \widetilde{I}_{n,1}^{(i,2)}\right)^2\right] - \left(\mathbb{E}\left[\widetilde{I}_{n,1}^{(i,1)}\right] + \mathbb{E}\left[\widetilde{I}_{n,1}^{(i,2)}\right]\right)^2 \\ = & \mathbb{E}\left[\left(\widetilde{I}_{n,1}^{(i,1)}\right)^2\right] + \mathbb{E}\left[\left(\widetilde{I}_{n,1}^{(i,2)}\right)^2\right] - 2\mathbb{E}\left[\widetilde{I}_{n,1}^{(i,1)}\widetilde{I}_{n,1}^{(i,2)}\right] \\ & - \mathbb{E}\left[\widetilde{I}_{n,1}^{(i,1)}\right]^2 - \mathbb{E}\left[\widetilde{I}_{n,1}^{(i,2)}\right]^2 - 2\mathbb{E}\left[\widetilde{I}_{n,1}^{(i,1)}\right]\mathbb{E}\left[\widetilde{I}_{n,1}^{(i,2)}\right] \\ \sim & n^{-2}\left(\frac{4b_q(L)^2}{q^2}\mathbb{V}\mathrm{ar}\left[\operatorname{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}}(f,\mathbf{X},Z)\right]\right] h^4 + \mu_2(K)^2\mathbb{V}\mathrm{ar}\left[\mathcal{H}_z f(\mathbf{X},Z)\right] g^4\right. \\ & \left. + \frac{4b_q(L)\mu_2(K)}{q}\mathbb{C}\mathrm{ov}\left[\operatorname{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}}(f,\mathbf{X},Z)\right], \mathcal{H}_z f(\mathbf{X},Z)\right] h^2 g^2\right) \\ = & n^{-2}\phi(h,g). \end{aligned}$$

With the previous expression, it follows $\mathbb{E}\left[\left(I_{n,1}^{(i)}\right)^2\right] = \mathcal{O}\left(n^{-2}(h^4+g^4)\right)$ (see the first point of Lemma A.12) and $s_n^2 = n^{-1}\phi(h,g)(1+o(1)) = \mathcal{O}\left(n^{-1}(h^4+g^4)\right)$. Then by the fourth point of Lemma A.12:

$$s_n^4 = \left(s_n^2\right)^2 = \mathcal{O}\left(n^{-2}(h^4+g^4)^2\right) = \mathcal{O}\left(n^{-2}(h^8+g^8)\right),$$

$$\mathbb{E}\left[(I_{n,1}^{(i)})^4\right] = \mathcal{O}\left(\mathbb{E}\left[(\widetilde{I}_{n,1}^{(i)})^4\right] + \mathbb{E}\left[\widetilde{I}_{n,1}^{(i)}\right]^4\right) = \mathcal{O}\left(\mathbb{E}\left[(\widetilde{I}_{n,1}^{(i)})^4\right]\right),$$

where

$$\mathbb{E}\left[(\widetilde{I}_{n,1}^{(i)})^4\right] = \mathcal{O}\left(\mathbb{E}\left[(\widetilde{I}_{n,1}^{(i)})^4\right] + \mathbb{E}\left[(\widetilde{I}_{n,1}^{(i)})^4\right]\right) = \mathcal{O}\left(n^{-4}(h^8+g^8)\right).$$

□

Lemma A.3. *Under conditions **A1**–**A3**,*

$$I_{n,2} = \mathbb{E}\left[I_{n,2}\right] + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{3}{2}}h^{-q}g^{-1}\right) = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^qg} + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{3}{2}}h^{-q}g^{-1}\right).$$

Proof of Lemma A.3. To prove the result the Chebychev inequality will be used. To that end, the expectation and variance of $I_{n,2} = \frac{c_{h,g}(L)^2}{n^2g^2}\sum_{i=1}^n I_{n,2}^{(i)}$ have to be computed. But first recall that, by Lemma A.10 and (6.1), for i and j naturals,

$$\int_{\Omega_q\times\mathbb{R}} LK^j\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}, \frac{z-t}{g}\right)\varphi^i(\mathbf{y},t)\,dt\,\omega_q(d\mathbf{y}) \sim h^qg\lambda_q(L^j)\varphi^i(\mathbf{x},z), \tag{A.4}$$

uniformly in $(\mathbf{x},z)\in\Omega_q\times\mathbb{R}$, with φ a uniformly continuous and bounded function and $\lambda_q(L^j) = \omega_{q-1}2^{\frac{q}{2}-1}\int_0^\infty L^j(r)r^{\frac{q}{2}-1}\,dr$. The following particular cases of this relation are useful to shorten the next computations:

i. $\mathbb{E}[LK(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}, \frac{z-Z}{g})] \sim h^qg\lambda_q(L)f(\mathbf{x},z)$,

ii. $\int_{\Omega_q\times\mathbb{R}} LK^2(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}, \frac{z-t}{g})\,dz\,\omega_q(d\mathbf{x}) \sim h^qg\lambda_q(L^2)R(K)$.

Expectation of $I_{n,2}$. The expectation is divided in two addends, which can be computed by applying the relations *i*–*ii*:

$$\begin{aligned}
\mathbb{E}\left[I_{n,2}^{(i)}\right] &= \mathbb{E}\left[\int_{\Omega_q\times\mathbb{R}} LK_n^2\left((\mathbf{x},z),(\mathbf{X},Z)\right)\,dz\,\omega_q(d\mathbf{x})\right] \\
&= \int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[LK^2\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}, \frac{z-Z}{g}\right)\right]dz\,\omega_q(d\mathbf{x}) \\
&\quad - \int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[LK\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}, \frac{z-Z}{g}\right)\right]^2 dz\,\omega_q(d\mathbf{x}) \\
&= \mathbb{E}\left[\int_{\Omega_q\times\mathbb{R}} LK^2\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}, \frac{z-Z}{g}\right)dz\,\omega_q(d\mathbf{x})\right] - h^{2q}g^2\lambda_q(L)^2R(f)(1+o(1)) \\
&= h^qg\lambda_q(L^2)R(K) + \mathcal{O}\left(h^{2q}g^2\right).
\end{aligned}$$

Therefore, the expectation of $I_{n,2}$ is

$$\mathbb{E}\left[I_{n,2}\right] = \frac{\lambda_q(L)^{-2}}{nh^{2q}g^2}\left(h^qg\lambda_q(L^2)R(K) + \mathcal{O}\left(h^{2q}g^2\right)\right) = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^qg} + \mathcal{O}\left(n^{-1}\right).$$

Variance of $I_{n,2}$. For the variance it suffices to compute its order, which follows considering the third point of Lemma A.12:

$$\begin{aligned}\mathbb{E}\left[\left(I_{n,2}^{(i)}\right)^2\right] &= \int_{\Omega_q\times\mathbb{R}}\left\{\int_{\Omega_q\times\mathbb{R}} LK_n^2\left((\mathbf{x},z),(\mathbf{y},t)\right)\, dz\,\omega_q(d\mathbf{x})\right\}^2 f(\mathbf{y},t)\,dt\,\omega_q(d\mathbf{y})\\ &= \mathcal{O}\left(I_{n,2}^{(i,1)} + I_{n,2}^{(i,2)}\right),\end{aligned}$$

where the involved terms are

$$\begin{aligned}I_{n,2}^{(i,1)} &= \int_{\Omega_q\times\mathbb{R}}\left\{\int_{\Omega_q\times\mathbb{R}} LK^2\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2},\frac{z-t}{g}\right)\, dz\,\omega_q(d\mathbf{x})\right\}^2 f(\mathbf{y},t)\,dt\,\omega_q(d\mathbf{y}),\\ I_{n,2}^{(i,2)} &= \int_{\Omega_q\times\mathbb{R}}\left\{\int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[LK\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2},\frac{z-Z}{g}\right)\right]^2\, dz\,\omega_q(d\mathbf{x})\right\}^2 f(\mathbf{y},t)\,dt\,\omega_q(d\mathbf{y}).\end{aligned}$$

Using relations *i*–*ii* the orders of the addends $I_{n,2}^{(i,k)}$, $k=1,2$, follow easily:

$$\begin{aligned}I_{n,2}^{(i,1)} &\sim \int_{\Omega_q\times\mathbb{R}}\left\{h^q g\lambda_q(L^2)R(K)\right\}^2 f(\mathbf{y},t)\,dt\,\omega_q(d\mathbf{y})\\ &= h^{2q}g^2\lambda_q(L^2)^2R(K)^2,\\ I_{n,2}^{(i,2)} &\sim \int_{\Omega_q\times\mathbb{R}}\left\{\int_{\Omega_q\times\mathbb{R}} \left(h^q g\lambda_q(L)f(\mathbf{x},z)\right)^2\, dz\,\omega_q(d\mathbf{x})\right\}^2 f(\mathbf{y},t)\,dt\,\omega_q(d\mathbf{y})\\ &= h^{4q}g^4\lambda_q(L)^4R(f)^2.\end{aligned}$$

Therefore $I_{n,2}^{(i,1)} = \mathcal{O}\left(h^{2q}g^2\right)$, $I_{n,2}^{(i,2)} = \mathcal{O}\left(h^{4q}g^4\right)$ and $\mathbb{E}\left[\left(I_{n,2}^{(i)}\right)^2\right] = \mathcal{O}(I_{n,2}^{(i,1)}) + \mathcal{O}(I_{n,2}^{(i,2)}) = \mathcal{O}\left(h^{2q}g^2\right)$. The variance of $I_{n,2}$ is

$$\mathbb{V}\mathrm{ar}\left[I_{n,2}\right] \le n^{-4}c_{h,q}(L)^4 g^{-4}\sum_{i=1}^{n}\mathbb{E}\left[\left(I_{n,2}^{(i)}\right)^2\right] = \mathcal{O}\left(n^{-3}h^{-2q}g^{-2}\right),$$

so by Chebychev's inequality

$$\mathbb{P}\left\{\left|I_{n,2}-\mathbb{E}\left[I_{n,2}\right]\right| \ge kn^{-\frac{3}{2}}h^{-q}g^{-1}\right\} \le \frac{1}{k^2},\quad \forall k>0,$$

which, by definition, is

$$I_{n,2} = \mathbb{E}\left[I_{n,2}\right] + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{3}{2}}h^{-q}g^{-1}\right) = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g} + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{3}{2}}h^{-q}g^{-1}\right),$$

because $\mathcal{O}\left(n^{-1}\right) = \mathcal{O}_{\mathbb{P}}(n^{-\frac{3}{2}}h^{-q}g^{-1})$. □

Lemma A.4. *Let be*

$$\begin{aligned}H_n\left((\mathbf{x},z),(\mathbf{y},t)\right) &= \int_{\Omega_q\times\mathbb{R}} LK_n\left((\mathbf{u},v),(\mathbf{x},z)\right)LK_n\left((\mathbf{u},v),(\mathbf{y},t)\right)\, dv\,\omega_q(d\mathbf{u}),\\ G_n\left((\mathbf{x},z),(\mathbf{y},t)\right) &= \mathbb{E}\left[H_n\left((\mathbf{X},Z),(\mathbf{x},z)\right)H_n\left((\mathbf{X},Z),(\mathbf{y},t)\right)\right],\\ M_n(\mathbf{X}_1,Z_1) &= 2\frac{c_{h,q}(L)^2}{n^2g^2}\mathbb{E}\left[I_{n,1}^{(2)}H_n\left((\mathbf{X}_1,Z_1),(\mathbf{X}_2,Z_2)\right)|(\mathbf{X}_1,Z_1)\right].\end{aligned}$$

*Then, under conditions **A1**–**A3**,*

$$\mathbb{E}\left[H_n^2\left((\mathbf{X}_1, Z_1), (\mathbf{X}_2, Z_2)\right)\right] = h^{3q} g^3 \lambda_q(L)^4 \sigma^2 \left(1 + o\left(1\right)\right), \tag{A.5}$$

$$\mathbb{E}\left[H_n^4\left((\mathbf{X}_1, Z_1), (\mathbf{X}_2, Z_2)\right)\right] = \mathcal{O}\left(h^{5q} g^5\right), \tag{A.6}$$

$$\mathbb{E}\left[G_n^2\left((\mathbf{X}_1, Z_1), (\mathbf{X}_2, Z_2)\right)\right] = \mathcal{O}\left(h^{7q} g^7\right), \tag{A.7}$$

$$\mathbb{E}\left[M_n^2(\mathbf{X}_1, Z_1)\right] = \mathcal{O}\left(n^{-6}(h^4 + g^4) h^{-\frac{3q}{2}} g^{-\frac{3}{2}}\right). \tag{A.8}$$

Proof of Lemma A.4. The proof is divided in four sections.

Proof of (A.5). $\mathbb{E}[H_n^2(\mathbf{X}_1, Z_1), (\mathbf{X}_2, Z_2))]$ can be split into three addends:

$$\begin{aligned}
&\mathbb{E}[H_n^2(\mathbf{X}_1, Z_1), (\mathbf{X}_2, Z_2))] \\
&\quad = \mathbb{E}\left[\left(\int_{\Omega_q \times \mathbb{R}} LK_n\left((\mathbf{x}, z), (\mathbf{X}_1, Z_1)\right) LK_n\left((\mathbf{x}, z), (\mathbf{X}_2, Z_2)\right)\, dz\, \omega_q(d\mathbf{x})\right)^2\right] \\
&\quad = \mathbb{E}\Bigg[\int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} LK_n\left((\mathbf{x}, z), (\mathbf{X}_1, Z_1)\right) LK_n\left((\mathbf{x}, z), (\mathbf{X}_2, Z_2)\right) \\
&\qquad \times LK_n\left((\mathbf{y}, t), (\mathbf{X}_1, Z_1)\right) LK_n\left((\mathbf{y}, t), (\mathbf{X}_2, Z_2)\right)\, dz\, \omega_q(d\mathbf{x})\, dt\, \omega_q(d\mathbf{y})\Bigg] \\
&\quad = \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} \mathbb{E}\left[LK_n\left((\mathbf{x}, z), (\mathbf{X}, Z)\right) LK_n\left((\mathbf{y}, t), (\mathbf{X}, Z)\right)\right]^2\, dz\, \omega_q(d\mathbf{x})\, dt\, \omega_q(d\mathbf{y}) \\
&\quad = \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} \left(E_1((\mathbf{x}, z), (\mathbf{y}, t)) - E_2((\mathbf{x}, z), (\mathbf{y}, t))\right)^2\, dz\, \omega_q(d\mathbf{x})\, dt\, \omega_q(d\mathbf{y}) \\
&\quad = A_1 - 2A_2 + A_3,
\end{aligned}$$

where:

$$\begin{aligned}
E_1((\mathbf{x}, z), (\mathbf{y}, t)) &= \mathbb{E}\left[LK\left(\frac{1 - \mathbf{x}^T\mathbf{X}}{h^2}, \frac{z - Z}{g}\right) LK\left(\frac{1 - \mathbf{y}^T\mathbf{X}}{h^2}, \frac{t - Z}{g}\right)\right], \\
E_2((\mathbf{x}, z), (\mathbf{y}, t)) &= \mathbb{E}\left[LK\left(\frac{1 - \mathbf{x}^T\mathbf{X}}{h^2}, \frac{z - Z}{g}\right)\right] \mathbb{E}\left[LK\left(\frac{1 - \mathbf{y}^T\mathbf{X}}{h^2}, \frac{t - Z}{g}\right)\right].
\end{aligned}$$

The dominant term of the three is A_1, which has order $\mathcal{O}\left(h^{3q} g^3\right)$, as it will be seen. The terms A_2 and A_3 have order $\mathcal{O}\left(h^{4q} g^4\right)$, which can be seen applying iteratively the relation (A.4):

$$\begin{aligned}
A_2 &= \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} E_1((\mathbf{x}, z), (\mathbf{y}, t)) E_2((\mathbf{x}, z), (\mathbf{y}, t))\, dt\, \omega_q(d\mathbf{y})\, dz\, \omega_q(d\mathbf{x}) \\
&\sim \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} \left(h^q g \lambda_q(L) LK\left(\frac{1 - \mathbf{x}^T\mathbf{u}}{h^2}, \frac{z - t}{g}\right) f(\mathbf{y}, t)\right) \\
&\qquad \times \left(h^{2q} g^2 \lambda_q(L)^2 f(\mathbf{x}, z) f(\mathbf{y}, t)\right)\, dt\, \omega_q(d\mathbf{y})\, dz\, \omega_q(d\mathbf{x}) \\
&\sim h^{4q} g^4 \lambda_q(L)^4 \int_{\Omega_q \times \mathbb{R}} f(\mathbf{x}, z)^3\, dz\, \omega_q(d\mathbf{x}), \\
A_3 &= \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} L_2^2((\mathbf{x}, z), (\mathbf{y}, t))\, dt\, \omega_q(d\mathbf{y})\, dz\, \omega_q(d\mathbf{x})
\end{aligned}$$

$$\sim \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} h^{4q}g^4\lambda_q(L)^4 f(\mathbf{x},z)^2 f(\mathbf{y},t)^2\,dt\,\omega_q(d\mathbf{y})\,dz\,\omega_q(d\mathbf{x})$$
$$= h^{4q}g^4\lambda_q(L)^4 R(f)^2.$$

Let recall now on the term A_1. In order to clarify the following computations, let denote by $(\mathbf{x},x)$, $(\mathbf{y},y)$ and $(\mathbf{z},z)$ the three variables in $\Omega_q\times\mathbb{R}$ that play the role of $(\mathbf{x},z)$, $(\mathbf{y},t)$ and $(\mathbf{u},v)$, respectively. The addend A_1 in this new notation is:

$$A_1 = \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}}\left[\int_{\Omega_q\times\mathbb{R}} LK\left(\frac{1-\mathbf{x}^T\mathbf{z}}{h^2},\frac{x-z}{g}\right) LK\left(\frac{1-\mathbf{y}^T\mathbf{z}}{h^2},\frac{y-z}{g}\right)\right.$$
$$\left.\times\, f(\mathbf{z},z)\,dz\,\omega_q(d\mathbf{z})\right]^2 dy\,\omega_q(d\mathbf{y})\,dx\,\omega_q(d\mathbf{x}).$$

The calculus of A_1 will be divided in the cases $q\geq 2$ and $q=1$. There are several changes of variables involved, which will be detailed in *i*–*iv*. To begin with, let suppose $q\geq 2$:

$$A_1 \overset{i}{=} \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_{q-1}}\int_{-1}^{1}\int_{\mathbb{R}}\left[\int_{\Omega_{q-2}}\iint_{t^2+\tau^2<1}\int_{\mathbb{R}}\right.$$
$$\times\, LK\left(\frac{1-t}{h^2},\frac{x-z}{g}\right) LK\left(\frac{1-st-\tau(1-s^2)^{\frac{1}{2}}}{h^2},\frac{y-z}{g}\right)$$
$$\times\, f\left(t\mathbf{x}+\tau\mathbf{B}_q\boldsymbol{\xi}+\left(1-t^2-\tau^2\right)^{\frac{1}{2}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta},z\right)(1-t^2-\tau^2)^{\frac{q-3}{2}}$$
$$\left.\times\, dz\,dt\,d\tau\,\omega_{q-2}(d\boldsymbol{\eta})\right]^2 (1-s^2)^{\frac{q}{2}-1}\,dy\,ds\,\omega_{q-1}(d\boldsymbol{\xi})\,dx\,\omega_q(d\mathbf{x})$$
$$\overset{ii}{=} \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_{q-1}}\int_{0}^{2h^{-2}}\int_{\mathbb{R}}\left[\int_{\Omega_{q-2}}\int_{0}^{2h^{-2}}\int_{-1}^{1}\int_{\mathbb{R}} LK\left(\rho,\frac{x-z}{g}\right)\right.$$
$$\times\, LK\left(r+\rho-h^2r\rho-\theta\left[r\rho(2-h^2r)(2-h^2\rho)\right]^{\frac{1}{2}},\frac{y-z}{g}\right)$$
$$\times\, f\left((1-h^2\rho)\mathbf{x}+h\left[\rho(2-h^2\rho)\right]^{\frac{1}{2}}\left[\theta\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}+(1-\theta^2)^{\frac{1}{2}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta}\right],z\right)$$
$$\times\,(1-\theta^2)^{\frac{q-3}{2}}h^{q-3}\left[\rho(2-h^2\rho)\right]^{\frac{q-3}{2}}h^3\left[\rho(2-h^2\rho)\right]^{\frac{1}{2}}\,dz\,d\theta\,d\rho\,\omega_{q-2}(d\boldsymbol{\eta})\Big]^2$$
$$\times\, h^{q-2}r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}h^2\,dy\,dr\,\omega_{q-1}(d\boldsymbol{\xi})\,dx\,\omega_q(d\mathbf{x})$$
$$\overset{iii}{=} h^{3q}g^3\int_{\Omega_q\times\mathbb{R}}\int_{\Omega_{q-1}}\int_{0}^{2h^{-2}}\int_{\mathbb{R}}\left[\int_{\Omega_{q-2}}\int_{0}^{2h^{-2}}\int_{-1}^{1}\int_{\mathbb{R}} LK\,(\rho,u)\right.$$
$$\times\, LK\left(r+\rho-h^2r\rho-\theta\left[r\rho(2-h^2r)(2-h^2\rho)\right]^{\frac{1}{2}},u+v\right)$$
$$\times\, f\left((1-h^2\rho)\mathbf{x}+h\left[\rho(2-h^2\rho)\right]^{\frac{1}{2}}\left[\theta\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}+(1-\theta^2)^{\frac{1}{2}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta}\right],x-ug\right)$$
$$\times\,(1-\theta^2)^{\frac{q-3}{2}}\left[\rho(2-h^2\rho)\right]^{\frac{q}{2}-1}\,du\,d\theta\,d\rho\,\omega_{q-2}(d\boldsymbol{\eta})\Big]^2 r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}$$
$$\times\, dv\,dr\,\omega_{q-1}(d\boldsymbol{\xi})\,dx\,\omega_q(d\mathbf{x})$$

$$
\begin{aligned}
&\overset{iv}{\sim} h^{3q} g^3 \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_{q-1}} \int_0^\infty \int_{\mathbb{R}} \Big[\int_{\Omega_{q-2}} \int_0^\infty \int_{-1}^1 \int_{\mathbb{R}} LK\left(\rho, u\right) \\
&\quad \times LK\left(r + \rho - 2\theta \left(r\rho\right)^{\frac{1}{2}}, u + v\right) f\left(\mathbf{x}, x\right) \left(1 - \theta^2\right)^{\frac{q-3}{2}} \left(2\rho\right)^{\frac{q}{2}-1} \\
&\quad \times \, du \, d\theta \, d\rho \, \omega_{q-2}(d\boldsymbol{\eta}) \Big]^2 \left(2r\right)^{\frac{q}{2}-1} dv \, dr \, \omega_{q-1}(d\boldsymbol{\xi}) \, dx \, \omega_q(d\mathbf{x}) \\
&= h^{3q} g^3 R(f) \omega_{q-1} \omega_{q-2}^2 2^{\frac{3q}{2}-1} \int_0^\infty r^{\frac{q}{2}-1} \int_{\mathbb{R}} \Big[\int_{\mathbb{R}} \int_0^\infty \rho^{\frac{q}{2}-1} LK(\rho, u) \\
&\quad \times \int_{-1}^1 (1 - \theta^2)^{\frac{q-3}{2}} LK\left(r + \rho - 2\theta(r\rho)^{\frac{1}{2}}, u + v\right) d\theta \, d\rho \, du \Big]^2 dv \, dr \\
&= h^{3q} g^3 \lambda_q(L)^4 \sigma^2.
\end{aligned}
$$

The steps for the calculus of the case $q \geq 2$ are the following:

i. Let $\mathbf{x}$ a fixed point in Ω_q, $q \geq 2$. Let be the change of variables:

$$
\mathbf{y} = s\mathbf{x} + (1 - s^2)^{\frac{1}{2}} \mathbf{B}_\mathbf{x} \boldsymbol{\xi}, \quad \omega_q(d\mathbf{y}) = (1 - s^2)^{\frac{q}{2}-1} \, ds \, \omega_{q-1}(d\boldsymbol{\xi}),
$$

where $s \in (-1, 1)$, $\boldsymbol{\xi} \in \Omega_{q-1}$ and $\mathbf{B}_\mathbf{x} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$ is the semi-orthonormal matrix ($\mathbf{B}_\mathbf{x}^T \mathbf{B}_\mathbf{x} = \mathbf{I}_q$ and $\mathbf{B}_\mathbf{x} \mathbf{B}_\mathbf{x}^T = \mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T$) resulting from the completion of $\mathbf{x}$ to the orthonormal basis $\{\mathbf{x}, \mathbf{b}_1, \ldots, \mathbf{b}_q\}$ of $\mathbb{R}^{q+1}$. Here $\mathbf{I}_q$ represents the identity matrix with dimension q. See Lemma 2 of García-Portugués et al. (2013b) for further details. Consider also the other change of variables

$$
\mathbf{z} = t\mathbf{x} + \tau \mathbf{B}_\mathbf{x} \boldsymbol{\xi} + (1 - t^2 - \tau^2)^{\frac{1}{2}} \mathbf{A}_{\boldsymbol{\xi}} \boldsymbol{\eta}, \quad \omega_q(d\mathbf{z}) = (1 - t^2 - \tau^2)^{\frac{q-3}{2}} \, dt \, d\tau \, \omega_{q-2}(d\boldsymbol{\eta}),
$$

where $t, \tau \in (-1, 1)$, $t^2 + \tau^2 < 1$, $\boldsymbol{\eta} \in \Omega_{q-2}$ and $\mathbf{A}_{\boldsymbol{\xi}} = (\mathbf{a}_1, \ldots, \mathbf{a}_q)_{(q+1)\times(q-1)}$ is the semi-orthonormal matrix ($\mathbf{A}_{\boldsymbol{\xi}}^T \mathbf{A}_{\boldsymbol{\xi}} = \mathbf{I}_q$ and $\mathbf{A}_{\boldsymbol{\xi}} \mathbf{A}_{\boldsymbol{\xi}}^T = \mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T - \mathbf{B}_\mathbf{x} \boldsymbol{\xi} \boldsymbol{\xi}^T \mathbf{B}_\mathbf{x}^T$) resulting from the completion of $\{\mathbf{x}, \mathbf{B}_\mathbf{x} \boldsymbol{\xi}\}$ to the orthonormal basis $\{\mathbf{x}, \mathbf{B}_\mathbf{x} \boldsymbol{\xi}, \mathbf{a}_1, \ldots, \mathbf{a}_{q-1}\}$ of $\mathbb{R}^{q+1}$. This change of variables can be obtained by replicating the proof of Lemma 2 in García-Portugués et al. (2013b) with an extra step for the case $q \geq 2$. With these two changes of variables,

$$
\mathbf{y}^T \mathbf{z} = st + \tau(1 - s^2)^{\frac{1}{2}}, \quad \mathbf{x}^T (\mathbf{B}_\mathbf{x} \boldsymbol{\xi}) = \mathbf{x}^T (\mathbf{A}_{\boldsymbol{\xi}} \boldsymbol{\eta}) = (\mathbf{B}_\mathbf{x} \boldsymbol{\xi})^T (\mathbf{A}_{\boldsymbol{\xi}} \boldsymbol{\eta}) = 0.
$$

ii. Consider first the change of variables $r = \frac{1-s}{h^2}$ and then

$$
\begin{cases} \rho = \frac{1-t}{h^2}, \\ \theta = \dfrac{\tau}{h[\rho(2-h^2\rho)]^{\frac{1}{2}}}, \end{cases} \qquad \left| \frac{\partial(t, \tau)}{\partial(\rho, \theta)} \right| = h^3 \left[\rho(2 - h^2\rho)\right]^{\frac{1}{2}}.
$$

With this last change of variables, $\tau = h\theta \left[\rho(2 - h^2\rho)\right]^{\frac{1}{2}}$, $t = 1 - h^2\rho$ and, as a result:

$$
\begin{aligned}
1 - s^2 &= h^2 r(2 - h^2 r), \\
1 - t^2 &= h^2 \rho(2 - h^2 \rho), \\
1 - t^2 - \tau^2 &= (1 - \theta^2) h^2 \rho(2 - h^2 \rho), \\
\frac{1 - st - \tau(1 - s^2)^{\frac{1}{2}}}{h^2} &= r + \rho - h^2 r\rho - \theta \left[r\rho(2 - h^2 r)(2 - h^2 \rho)\right]^{\frac{1}{2}}.
\end{aligned}
$$

iii. Use $u = \frac{x-z}{g}$ and $v = \frac{y-x}{g}$.

iv. By expanding the square, A_1 can be written as

$$\begin{aligned} A_1 = & h^{3q} g^3 \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_{q-1}} \int_0^\infty \int_{\mathbb{R}} \Big[\int_{\Omega_{q-2}} \int_0^\infty \int_{-1}^1 \int_{\mathbb{R}} \int_{\Omega_{q-2}} \int_0^\infty \int_{-1}^1 \int_{\mathbb{R}} \\ & \times \varphi_n(\mathbf{x}, x, r, \rho_1, \theta_1, u_1, v, \boldsymbol{\xi}, \boldsymbol{\eta}_1) \varphi_n(\mathbf{x}, x, r, \rho_2, \theta_2, u_2, v, \boldsymbol{\xi}, \boldsymbol{\eta}_2) \\ & \times \, du_1 \, d\theta_1 \, d\rho_1 \, \omega_{q-2}(d\boldsymbol{\eta}_1) \, du_2 \, d\theta_2 \, d\rho_2 \, \omega_{q-2}(d\boldsymbol{\eta}_2) \Big] \\ & \times \, dv \, dr \, \omega_{q-1}(d\boldsymbol{\xi}) \, dx \, \omega_q(d\mathbf{x}), \end{aligned} \tag{A.9}$$

where

$$\begin{aligned} & \varphi_n(\mathbf{x}, x, r, \rho_i, \theta_i, u_i, v, \boldsymbol{\xi}, \boldsymbol{\eta}_i) \\ & \quad = L(\rho_i) \, L\left(r + \rho_i - h^2 r \rho_i - \theta \left[r\rho_i (2 - h^2 r)(2 - h^2 \rho_i) \right]^{\frac{1}{2}} \right) \\ & \qquad \times K(u_i) \, K(u_i + v) \, f((\mathbf{x}, x) + \boldsymbol{\alpha}_{h,g}) \, (1 - \theta_i^2)^{\frac{q-3}{2}} \\ & \qquad \times \rho_i^{\frac{q}{2}-1} (2 - h^2 \rho_i)^{\frac{q}{2}-1} r^{\frac{q}{4}-\frac{1}{2}} (2 - h^2 r)^{\frac{q}{4}-\frac{1}{2}} \mathbf{1}_{[0,2h^{-2})}(r) \mathbf{1}_{[0,2h^{-2})}(\rho_i), \end{aligned}$$

with $\boldsymbol{\alpha}_{h,g} = \Big(-h^2 \rho_i \mathbf{x} + h \left[\rho_i (2 - h^2 \rho_i) \right]^{\frac{1}{2}} \left[\theta \mathbf{B}_{\mathbf{x}} \boldsymbol{\xi} + (1 - \theta_i^2)^{\frac{1}{2}} \mathbf{A}_{\boldsymbol{\xi}} \boldsymbol{\eta}_i \right], -u_i g \Big)$ and $i = 1, 2$. A first step to apply the DCT is to see that by the Taylor's theorem,

$$f((\mathbf{x}, x) + \boldsymbol{\alpha}_{h,g}) = f(\mathbf{x}, x) + \mathcal{O}\left(\boldsymbol{\alpha}_{h,g}^T \boldsymbol{\nabla} f(\mathbf{x}, x) \right),$$

where the remaining order is $\mathcal{O}((h^2 \rho_i + g^2 u_i^2)^{\frac{1}{2}} \, ||\boldsymbol{\nabla} f(\mathbf{x}, x)|| \,)$ because $||\boldsymbol{\alpha}_{h,g}||^2 = 2h^2 \rho_i + g^2 u_i^2$. Furthermore, the order is uniform for all points $(\mathbf{x}, x)$ because of the boundedness assumption of the second derivative given by **A1** (see the proof of Lemma A.11). Next, as $h, g \to 0$, then the order becomes $o\left((\sqrt{\rho_i} + u_i) \, ||\boldsymbol{\nabla} f(\mathbf{x}, x)|| \right)$.

For bounding the directional kernel L, recall that by completing the square,

$$\begin{aligned} (2 - h^2 r)(2 - h^2 \rho_i) = & \, 4 - 2h^2 (r + \rho_i) + h^4 \left((r + \rho_i)/2 \right)^2 - h^4 \left(\left((r + \rho_i)/2 \right) \right)^2 - r\rho_i \Big) \\ & \leq \left(2 - h^2 \frac{r + \rho_i}{2} \right)^2. \end{aligned}$$

Using this, and the fact that $\theta \in (-1, 1)$, for all $r, \rho_i \in [0, 2h^{-2})$,

$$\begin{aligned} & r + \rho_i - h^2 r \rho_i - \theta \left[r\rho_i (2 - h^2 r)(2 - h^2 \rho_i) \right]^{\frac{1}{2}} \\ & \qquad \geq r + \rho_i - h^2 r \rho_i - (r\rho_i)^{\frac{1}{2}} \left[(2 - h^2 r)(2 - h^2 \rho_i) \right]^{\frac{1}{2}} \\ & \qquad \geq r + \rho_i - h^2 r \rho_i - (r\rho_i)^{\frac{1}{2}} \left(2 - h^2 \frac{r + \rho_i}{2} \right) \\ & \qquad = r + \rho_i - 2(r\rho_i)^{\frac{1}{2}} + h^2 (r\rho_i)^{\frac{1}{2}} \left(\frac{r + \rho_i}{2} - (r\rho_i)^{\frac{1}{2}} \right) \\ & \qquad \geq r + \rho_i - 2(r\rho_i)^{\frac{1}{2}}, \end{aligned}$$

where the last inequality follows because the last addend is positive by the inequality of the geometric and arithmetic means. As L is a decreasing function by **A2**,

$$L\Big(r+\rho_i-h^2r\rho_i-\theta\left[r\rho_i(2-h^2r)(2-h^2\rho_i)\right]^{\frac{1}{2}}\Big)\leq L\left(r+\rho_i-2(r\rho_i)^{\frac{1}{2}}\right).$$

Then for all the variables in the integration domain of A_1,

$$\begin{aligned}\varphi_n(\mathbf{x},x,r,\rho_i,\theta_i,u_i,v,\boldsymbol{\xi},\boldsymbol{\eta}_i)\leq{}& L\left(\rho_i\right)L\left(r+\rho_i-2(r\rho_i)^{\frac{1}{2}}\right)K\left(u_i\right)K\left(u_i+v\right)\\&\times\left(f(\mathbf{x},x)+o\left(\left(\sqrt{\rho_i}+u_i\right)||\boldsymbol{\nabla}f(\mathbf{x},x)||\right)\right)\left(1-\theta_i^2\right)^{\frac{q-3}{2}}\\&\times\rho_i^{\frac{q}{2}-1}2^{\frac{q}{2}-1}r^{\frac{q}{4}-\frac{1}{2}}2^{\frac{q}{4}-\frac{1}{2}}\mathbf{1}_{[0,\infty)}(r)\mathbf{1}_{[0,\infty)}(\rho_i)\\={}&\Psi(\mathbf{x},x,r,\rho_i,\theta_i,u_i,v).\end{aligned}$$

The product of functions φ_n in (A.9) is bounded by the respective product of functions Ψ. The product is also integrable as a consequence of assumptions **A1** (integrability of f and $\boldsymbol{\nabla}f$), **A2** (integrability of kernels) and that the product of integrable functions is integrable. To prove it, recall that by the integral definition of the modified Bessel function of order $\frac{q}{2}-1$ (see equation 10.32.2 of Olver et al. (2010)):

$$\int_{-1}^{1}(1-\theta^2)^{\frac{q-3}{2}}\,d\theta=\frac{\sqrt{\pi}\Gamma\left(\frac{q-1}{2}\right)}{\Gamma\left(\frac{q}{2}\right)}<\infty,\quad\forall q\geq 2.$$

The integral of the linear kernel is proved to be finite using the Cauchy-Schwartz inequality and **A2**:

$$\begin{aligned}\int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}}K(u_1)K(u_1+v)K(u_2)K(u_2+v)\,du_1\,du_2\,dv&\\=\int_{\mathbb{R}}\int_{\mathbb{R}}K(u_1)K(u_2)\left[\int_{\mathbb{R}}K(u_1+v)K(u_2+v)\,dv\right]du_1\,du_2&\\\leq\int_{\mathbb{R}}\int_{\mathbb{R}}K(u_1)K(u_2)\mu_2(K)^{\frac{1}{2}}\mu_2(K)^{\frac{1}{2}}\,du_1\,du_2&\\=\mu_2(K).&\end{aligned}$$

For the directional situation, the following auxiliary result based on **A2** is needed:

$$\begin{aligned}\int_0^\infty L^2\left(\left(\sqrt{r}-\sqrt{\rho_i}\right)^2\right)r^{\frac{q}{2}-1}\,dr&\leq\int_0^\infty L^2\left(s\right)\left(\sqrt{s}+\sqrt{\rho_i}\right)^{q-1}s^{-\frac{1}{2}}\,dr\\&=\int_0^\infty L^2\left(s\right)\sum_{k=0}^{q-1}s^{\frac{k-1}{2}}\rho_i^{\frac{q-1-k}{2}}\,dr\\&=\sum_{k=0}^{q-1}\lambda_{k+1}(L^2)\rho_i^{\frac{q-1-k}{2}}\\&=\mathcal{O}\left(\rho_i^{\frac{q-1}{2}}\right).\end{aligned}$$

Using this and that $\int_0^\infty L(\rho)\rho^{\frac{3q-5}{4}}\,dr\leq\lambda_{\lceil\frac{2q+1}{3}\rceil}(L)<\infty$, it follows:

$$\int_0^\infty\int_0^\infty\int_0^\infty L(\rho_1)L(\rho_2)L\left(r+\rho_1-2(r\rho_1)^{\frac{1}{2}}\right)L\left(r+\rho_2-2(r\rho_2)^{\frac{1}{2}}\right)$$

$$
\begin{aligned}
&\times \rho_1^{\frac{q}{2}-1}\rho_2^{\frac{q}{2}-1} r^{\frac{q}{2}-1}\, dr\, d\rho_1\, d\rho_2 \\
&= \int_0^\infty \int_0^\infty L(\rho_1)L(\rho_2)\rho_1^{\frac{q}{2}-1}\rho_2^{\frac{q}{2}-1} \\
&\quad \times \left[\int_0^\infty L\left(\left(\sqrt{r}-\sqrt{\rho_1}\right)^2\right) L\left(\left(\sqrt{r}-\sqrt{\rho_2}\right)^2\right) r^{\frac{q}{2}-1}\, dr\right] d\rho_1\, d\rho_2 \\
&\leq \int_0^\infty \int_0^\infty L(\rho_1)L(\rho_2)\rho_1^{\frac{q}{2}-1}\rho_2^{\frac{q}{2}-1}\mathcal{O}\left(\rho_1^{\frac{q-1}{4}}\right)\mathcal{O}\left(\rho_2^{\frac{q-1}{4}}\right) d\rho_1\, d\rho_2 \\
&= \mathcal{O}(1)\,.
\end{aligned}
$$

Then, by the DCT,

$$
\begin{aligned}
A_1 \sim h^{3q} g^3 \int_{\Omega_q\times\mathbb{R}} \int_{\Omega_{q-1}} \int_0^\infty \int_{\mathbb{R}} \Bigg[& \int_{\Omega_{q-2}} \int_0^\infty \int_{-1}^1 \int_{\mathbb{R}} \\
& \times LK(\rho,u)\, LK\left(r+\rho-2\theta(r\rho)^{\frac{1}{2}}, u+v\right) \\
& \times f(\mathbf{x},x)\,(1-\theta^2)^{\frac{q-3}{2}}\,(2\rho)^{\frac{q}{2}-1}\, du\, d\theta\, d\rho\, \omega_{q-2}(d\boldsymbol{\eta})\Bigg]^2 \\
& \times (2r)^{\frac{q}{2}-1}\, dv\, dr\, \omega_{q-1}(d\boldsymbol{\xi})\, dx\, \omega_q(d\mathbf{x}),
\end{aligned}
$$

because all the functions involved are continuous almost everywhere.

Turn now to the case $q=1$. As before, the details of the case $q=1$ are explained in *vi–ix*:

$$
\begin{aligned}
A_1 \overset{vi}{=} & \int_{\Omega_1\times\mathbb{R}} \int_{\Omega_0} \int_{-1}^1 \int_{\mathbb{R}} \Bigg[\int_{\Omega_0} \int_{-1}^1 \int_{\mathbb{R}} LK\left(\frac{1-t}{h^2}, \frac{x-z}{g}\right) \\
& \times LK\left(\frac{1-st-(1-t^2)^{\frac{1}{2}}(1-s^2)^{\frac{1}{2}}(\mathbf{B}_\mathbf{x}\boldsymbol{\xi})^T(\mathbf{A}_\mathbf{x}\boldsymbol{\eta})}{h^2}, \frac{y-z}{g}\right) \\
& \times f\left(t\mathbf{x}+(1-t^2)^{\frac{1}{2}}\mathbf{A}_\mathbf{x}\boldsymbol{\eta}, z\right)(1-t^2)^{-\frac{1}{2}}\, dz\, dt\, \omega_0(d\boldsymbol{\eta})\Bigg]^2 \\
& \times (1-s^2)^{-\frac{1}{2}}\, dy\, ds\, \omega_0(d\boldsymbol{\xi})\, dx\, \omega_1(d\mathbf{x}) \\
\overset{vii}{=} & \int_{\Omega_1\times\mathbb{R}} \int_{\Omega_0} \int_0^{2h^{-2}} \int_{\mathbb{R}} \Bigg[\int_{\Omega_0} \int_0^{2h^{-2}} \int_{\mathbb{R}} LK\left(\rho, \frac{x-z}{g}\right) \\
& \times LK\left(r+\rho-h^2r\rho-\left(r\rho(2-h^2r)(2-h^2\rho)\right)^{\frac{1}{2}}(\mathbf{B}_\mathbf{x}\boldsymbol{\xi})^T\mathbf{A}_\mathbf{x}\boldsymbol{\eta}, \frac{y-z}{g}\right) \\
& \times f\left((1-h^2\rho)\mathbf{x}+h\left[\rho(2-h^2\rho)\right]^{\frac{1}{2}}\mathbf{A}_\mathbf{x}\boldsymbol{\eta}, z\right) h^{-1}\rho^{-\frac{1}{2}}(2-h^2\rho)^{-\frac{1}{2}}h^2 \\
& \times\ dz\, d\rho\, \omega_0(d\boldsymbol{\eta})\Bigg]^2 h^{-1}r^{-\frac{1}{2}}(2-h^2r)^{-\frac{1}{2}}h^2\, dy\, dr\, \omega_0(d\boldsymbol{\xi})\, dx\, \omega_1(d\mathbf{x}) \\
\overset{viii}{=} & h^3g^3 \int_{\Omega_1\times\mathbb{R}} \int_{\Omega_0} \int_0^{2h^{-2}} \int_{\mathbb{R}} \Bigg[\int_{\Omega_0} \int_0^{2h^{-2}} \int_{\mathbb{R}} LK(\rho,u) \\
& \times LK\left(r+\rho-h^2r\rho-\left(r\rho(2-h^2r)(2-h^2\rho)\right)^{\frac{1}{2}}(\mathbf{B}_\mathbf{x}\boldsymbol{\xi})^T\mathbf{A}_\mathbf{x}\boldsymbol{\eta}, u+v\right) \\
& \times f\left((1-h^2\rho)\mathbf{x}+h\left[\rho(2-h^2\rho)\right]^{\frac{1}{2}}\mathbf{A}_\mathbf{x}\boldsymbol{\eta}, x-ug\right)
\end{aligned}
$$

$$
\begin{aligned}
&\times \rho^{-\frac{1}{2}}(2-h^2\rho)^{-\frac{1}{2}}\,du\,d\rho\,\omega_0(d\boldsymbol{\eta})\Big]^2 r^{-\frac{1}{2}}(2-h^2 r)^{-\frac{1}{2}}\\
&\times\ dv\,dr\,\omega_0(d\boldsymbol{\xi})\,dx\,\omega_1(d\mathbf{x})\\
&\overset{vi}{=} h^3 g^3 \int_{\Omega_1\times\mathbb{R}}\int_{\Omega_0}\int_0^{2h^{-2}}\int_{\mathbb{R}}\Big[\int_0^{2h^{-2}}\int_{\mathbb{R}} LK\,(\rho,u)\\
&\times\Big[LK\left(r+\rho-h^2 r\rho+\left(r\rho(2-h^2 r)(2-h^2\rho)\right)^{\frac{1}{2}},u+v\right)\\
&\times f\left((1-h^2\rho)\mathbf{x}+h\left[\rho(2-h^2\rho)\right]^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi},x-ug\right)\\
&+LK\left(r+\rho-h^2 r\rho-\left(r\rho(2-h^2 r)(2-h^2\rho)\right)^{\frac{1}{2}},u+v\right)\\
&\times f\left((1-h^2\rho)\mathbf{x}-h\left[\rho(2-h^2\rho)\right]^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi},x-ug\right)\Big]\\
&\times \rho^{-\frac{1}{2}}(2-h^2\rho)^{-\frac{1}{2}}\,du\,d\rho\Big]^2 r^{-\frac{1}{2}}(2-h^2 r)^{-\frac{1}{2}}\,dv\,dr\,\omega_0(d\boldsymbol{\xi})\,dx\,\omega_1(d\mathbf{x})\\
&\overset{ix}{\sim} h^3 g^3 \int_{\Omega_1\times\mathbb{R}}\int_{\Omega_0}\int_0^{\infty}\int_{\mathbb{R}}\Big[\int_0^{\infty}\int_{\mathbb{R}} LK\,(\rho,u)\\
&\times\Big[LK\left(r+\rho+2(r\rho)^{\frac{1}{2}},u+v\right)+LK\left(r+\rho-2(r\rho)^{\frac{1}{2}},u+v\right)\Big]\\
&\times f\,(\mathbf{x},x)\,\rho^{-\frac{1}{2}}2^{-\frac{1}{2}}\,du\,d\rho\Big]^2 r^{-\frac{1}{2}}2^{-\frac{1}{2}}\,dv\,dr\,\omega_0(d\boldsymbol{\xi})\,dx\,\omega_1(d\mathbf{x})\\
&= h^3 g^3 R(f) 2^{-\frac{1}{2}}\int_0^{\infty} r^{-\frac{1}{2}}\int_{\mathbb{R}}\Big[\int_0^{\infty}\int_{\mathbb{R}}\rho^{-\frac{1}{2}}LK\,(\rho,u)\\
&\times\Big[LK\left(r+\rho+2(r\rho)^{\frac{1}{2}},u+v\right)+LK\left(r+\rho-2(r\rho)^{\frac{1}{2}},u+v\right)\Big]\\
&\times\ du\,d\rho\Big]^2\,dv\,dr\\
&= h^{3q} g^3 \lambda_q(L)^4\sigma^2.
\end{aligned}
$$

The steps used for the calculus are the following:

vi. Let $\mathbf{x}$ a fixed point in Ω_q. For $q=1$, let be the changes of variables

$$
\begin{aligned}
\mathbf{y} &= s\mathbf{x}+(1-s^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}, \quad \omega_1(d\mathbf{y}) = (1-s^2)^{\frac{q}{2}-1}\,ds\,\omega_0(d\boldsymbol{\xi}),\\
\mathbf{z} &= t\mathbf{x}+(1-t^2)^{\frac{1}{2}}\mathbf{A}_{\mathbf{x}}\boldsymbol{\eta}, \quad \omega_1(d\mathbf{z}) = (1-t^2)^{\frac{q}{2}-1}\,dt\,\omega_0(d\boldsymbol{\eta}),
\end{aligned}
$$

where $s,t\in(-1,1)$ and $\mathbf{B}_{\mathbf{x}}$ and $\mathbf{A}_{\mathbf{x}}$ are two semi-orthonormal matrices whose q columns are vectors that extend $\mathbf{x}$ to an orthonormal basis of $\mathbb{R}^{q+1}$. Note that as $q=1$ and $\mathbf{x}^T(\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})=\mathbf{x}^T(\mathbf{A}_{\mathbf{x}}\boldsymbol{\eta})=0$, then necessarily $\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}=\mathbf{A}_{\mathbf{x}}\boldsymbol{\eta}$ or $\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}=-\mathbf{A}_{\mathbf{x}}\boldsymbol{\eta}$.

vii. Let be the changes of variables $\rho=\frac{1-t}{h^2}$ and $r=\frac{1-s}{h^2}$. With this change, $t=1-h^2\rho$ and $s=1-h^2 r$. Then $1-s^2=h^2 r(2-h^2 r)$, $1-t^2=h^2\rho(2-h^2\rho)$ and

$$
\begin{aligned}
&\frac{1-st-(1-s^2)^{\frac{1}{2}}(1-t^2)^{\frac{1}{2}}(\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T\mathbf{A}_{\mathbf{x}}\boldsymbol{\eta}}{h^2}\\
&\qquad = r+\rho-h^2 r\rho-\left(r\rho(2-h^2 r)(2-h^2\rho)\right)^{\frac{1}{2}}(\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T\mathbf{A}_{\mathbf{x}}\boldsymbol{\eta}.
\end{aligned}
$$

viii. Use $u = \frac{x-z}{g}$ and $v = \frac{y-x}{g}$.

ix. A_1 can be written as

$$A_1 = h^{3q} g^3 \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_{q-1}} \int_0^\infty \int_{\mathbb{R}} \Big[\int_0^\infty \int_{\mathbb{R}} \int_0^\infty \int_{\mathbb{R}} \varphi_n(\mathbf{x}, x, r, \rho_1, u_1, v, \boldsymbol{\xi}) \times \varphi_n(\mathbf{x}, x, r, \rho_2, u_2, v, \boldsymbol{\xi})\, du_1\, d\rho_1\, du_2\, d\rho_2 \Big]\, dv\, dr\, \omega_{q-1}(d\boldsymbol{\xi})\, dx\, \omega_q(d\mathbf{x}),$$

where

$$\begin{aligned}\varphi_n(\mathbf{x}, x, r, \rho_i, u_i, v, \boldsymbol{\xi}) = L(\rho_i)\, K(u_i) \Big[L\Big(r + \rho_i - h^2 r \rho_i + \Big[r\rho_i (2 - h^2 r)(2 - h^2 \rho_i) \Big]^{\frac{1}{2}} \Big) \\ \times K(u_i + v)\, f\Big((\mathbf{x}, x) + \boldsymbol{\alpha}_{h,g}^{(1)} \Big) + K(u_i + v)\, f\Big((\mathbf{x}, x) + \boldsymbol{\alpha}_{h,g}^{(2)} \Big) \\ \times L\Big(r + \rho_i - h^2 r \rho_i - \Big[r\rho_i (2 - h^2 r)(2 - h^2 \rho_i) \Big]^{\frac{1}{2}} \Big) \Big] \\ \times \rho_i^{-\frac{1}{2}} (2 - h^2 \rho_i)^{-\frac{1}{2}} r^{-\frac{1}{4}} (2 - h^2 r)^{-\frac{1}{4}} \mathbf{1}_{[0, 2h^{-2})}(r) \mathbf{1}_{[0, 2h^{-2})}(\rho_i),\end{aligned}$$

with $\boldsymbol{\alpha}_{h,g}^{(j)} = \Big(-h^2 \rho_i \mathbf{x} + k_j h \left[\rho_i (2 - h^2 \rho_i) \right]^{\frac{1}{2}} \mathbf{B}_{\mathbf{x}} \boldsymbol{\xi}, -u_i g \Big)$ and $k_1 = 1$, $k_2 = -1$. As before, by the Taylor's theorem,

$$f\Big((\mathbf{x}, x) + \boldsymbol{\alpha}_{h,g}^{(k)} \Big) = f(\mathbf{x}, x) + o\left((\sqrt{\rho_i} + u_i)\, ||\boldsymbol{\nabla} f(\mathbf{x}, x)|| \right),$$

where the order is uniform for all points $(\mathbf{x}, x)$. By analogous considerations as for the case $q \geq 2$,

$$\begin{aligned}\varphi_n(\mathbf{x}, x, r, \rho_i, u_i, v, \boldsymbol{\xi}) &\leq 2L(\rho_i)\, L\Big(r + \rho_i - 2(r\rho_i)^{\frac{1}{2}} \Big) K(u_i)\, K(u_i + v) \Big(f(\mathbf{x}, x) \\ &\quad + o\left((\sqrt{\rho_i} + u_i)\, ||\boldsymbol{\nabla} f(\mathbf{x}, x)|| \right) \Big) \rho_i^{-\frac{1}{2}} r^{-\frac{1}{4}} \mathbf{1}_{[0,\infty)}(r) \mathbf{1}_{[0,\infty)}(\rho_i) \\ &= \Psi(\mathbf{x}, x, r, \rho_i, u_i, v),\end{aligned}$$

Then the product of functions φ_n is bounded by the respective product of functions Ψ, which is integrable, and by the DCT the limit commute with the integrals.

Proof of (A.6). $\mathbb{E}\left[H_n^4((\mathbf{X}_1, Z_1), (\mathbf{X}_2, Z_2)) \right]$ can be decomposed in the sum of two terms:

$$\begin{aligned}&\mathbb{E}\big[H_n^4(\mathbf{X}_1, Z_1), (\mathbf{X}_2, Z_2)) \big] \\ &\quad = \mathbb{E}\left[\Big(\int_{\Omega_q \times \mathbb{R}} LK_n((\mathbf{x}, z), (\mathbf{X}_1, Z_1))\, LK_n((\mathbf{x}, z), (\mathbf{X}_2, Z_2))\; dz\, \omega_q(d\mathbf{x}) \Big)^4 \right] \\ &\quad = \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} \left(E_1((\mathbf{x}, z), (\mathbf{y}, t)) - E_2((\mathbf{x}, z), (\mathbf{y}, t)) \right)^4\; dz\, \omega_q(d\mathbf{x})\, dt\, \omega_q(d\mathbf{y}) \\ &\quad = \mathcal{O}(B_1 + B_2).\end{aligned}$$

The calculus of the orders of these terms is analogous to the calculus of A_2 and A_3:

$$B_1 = \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} \left(E_1((\mathbf{x}, z), (\mathbf{y}, t)) \right)^4\; dt\, \omega_q(d\mathbf{y})\, dz\, \omega_q(d\mathbf{x})$$

$$
\begin{aligned}
&\sim \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} \left(h^q g\lambda_q(L)LK\left(\frac{1-\mathbf{x}^T\mathbf{u}}{h^2},\frac{z-t}{g}\right)f(\mathbf{y},t)\right)^4 dt\,\omega_q(d\mathbf{y})\,dz\,\omega_q(d\mathbf{x})\\
&\sim h^{5q}g^5\lambda_q(L)^4\lambda_q(L^4)R(f^2),\\
B_2 &= \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} \left(E_2((\mathbf{x},z),(\mathbf{y},t))\right)^4\, dt\,\omega_q(d\mathbf{y})\,dz\,\omega_q(d\mathbf{x})\\
&\sim \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} h^{8q}g^8\lambda_q(L)^8 f(\mathbf{x},z)^4 f(\mathbf{y},t)^4\,dt\,\omega_q(d\mathbf{y})\,dz\,\omega_q(d\mathbf{x})\\
&= h^{8q}g^8\lambda_q(L)^8 R(f^2)^2.
\end{aligned}
$$

Then $\mathbb{E}\left[H_n^4\left((\mathbf{X}_1,Z_1),(\mathbf{X}_2,Z_2)\right)\right]=\mathcal{O}\left(h^{5q}g^5\right)$.

Proof of (A.7). The notation $(\mathbf{x},x)$, $(\mathbf{y},y)$, $(\mathbf{z},z)$ and $(\mathbf{u},u)$ for variables in $\Omega_q\times\mathbb{R}$ will be employed again:

$$
\begin{aligned}
G_n((\mathbf{x},x),(\mathbf{y},y)) &= \int_{\Omega_q\times\mathbb{R}} H_n\left((\mathbf{z},z),(\mathbf{x},x)\right) H_n\left((\mathbf{z},z),(\mathbf{y},y)\right) f(\mathbf{z},z)\,dz\,\omega_q(d\mathbf{z})\\
&= \int_{\Omega_q\times\mathbb{R}} \left\{\int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{u},u),(\mathbf{x},x))LK_n((\mathbf{u},u),(\mathbf{z},z))\,du\,\omega_q(d\mathbf{u})\right\}\\
&\quad\times\left\{\int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{u},u),(\mathbf{y},y))LK_n((\mathbf{u},u),(\mathbf{z},z))\,du\,\omega_q(d\mathbf{u})\right\}\\
&\quad\times f(\mathbf{z},z)\,dz\,\omega_q(d\mathbf{z}).
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
&\mathbb{E}[G_n^2((\mathbf{X}_1,Z_1),(\mathbf{X}_2,Z_2))]\\
&\quad= \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} \left\{\int_{\Omega_q\times\mathbb{R}} \left[\int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{u},u),(\mathbf{x},x))LK_n((\mathbf{u},u),(\mathbf{z},z))\,du\,\omega_q(d\mathbf{u})\right]\right.\\
&\qquad\left.\times\left[\int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{u},u),(\mathbf{y},y))LK_n((\mathbf{u},u),(\mathbf{z},z))\,du\,\omega_q(d\mathbf{u})\right] f(\mathbf{z},z)\,dz\,\omega_q(d\mathbf{z})\right\}^2\\
&\qquad\times f(\mathbf{y},y)f(\mathbf{x},x)\,dy\,\omega_q(d\mathbf{y})\,dx\,\omega_q(d\mathbf{x}).
\end{aligned}
$$

Then, according to the expression of LK_n, $\mathbb{E}\left[G_n^2\left((\mathbf{X}_1,Z_1),(\mathbf{X}_2,Z_2)\right)\right]$ can be decomposed in 16 summands, which, in view of the symmetric roles of $(\mathbf{x},x)$ and $(\mathbf{y},y)$ can be reduced to 9 different summands. The first of all, C_1, is the dominant and has order $\mathcal{O}\left(h^{7q}g^7\right)$. Again, the orders are computed using (A.4) iteratively:

$$
\begin{aligned}
C_1 = \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} &\left\{\int_{\Omega_q\times\mathbb{R}}\right.\\
&\times\left[\int_{\Omega_q\times\mathbb{R}} LK\left(\frac{1-\mathbf{u}^T\mathbf{x}}{h^2},\frac{u-x}{g}\right)LK\left(\frac{1-\mathbf{u}^T\mathbf{z}}{h^2},\frac{u-z}{g}\right)\,du\,\omega_q(d\mathbf{u})\right]\\
&\times\left[\int_{\Omega_q\times\mathbb{R}} LK\left(\frac{1-\mathbf{u}^T\mathbf{y}}{h^2},\frac{u-y}{g}\right)LK\left(\frac{1-\mathbf{u}^T\mathbf{x}}{h^2},\frac{u-x}{g}\right)\,du\,\omega_q(d\mathbf{u})\right]\\
&\left.\times f(\mathbf{z},z)\,dz\,\omega_q(d\mathbf{z})\right\}^2 f(\mathbf{y},y)f(\mathbf{x},x)\,dy\,\omega_q(d\mathbf{y})\,dx\,\omega_q(d\mathbf{x})
\end{aligned}
$$

$$
\begin{aligned}
&\sim \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}}\left\{\int_{\Omega_q\times\mathbb{R}}\left[\lambda_q(L)h^q gLK\left(\frac{1-\mathbf{x}^T\mathbf{z}}{h^2},\frac{x-z}{g}\right)\right]\right.\\
&\quad\left.\times\left[\lambda_q(L)h^q gLK\left(\frac{1-\mathbf{y}^T\mathbf{z}}{h^2},\frac{y-z}{g}\right)\right]f(\mathbf{z},z)\,dz\,\omega_q(d\mathbf{z})\right\}^2\\
&\quad\times f(\mathbf{y},y)f(\mathbf{x},x)\,dy\,\omega_q(d\mathbf{y})\,dx\,\omega_q(d\mathbf{x})\\
&\sim\lambda_q(L)^4h^{4q}g^4\int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}}\left\{\lambda_q(L)h^q gLK\left(\frac{1-\mathbf{y}^T\mathbf{x}}{h^2},\frac{y-x}{g}\right)f(\mathbf{x},x)\right\}^2\\
&\quad\times f(\mathbf{y},y)f(\mathbf{x},x)\,dy\,\omega_q(d\mathbf{y})\,dx\,\omega_q(d\mathbf{x})\\
&\sim\lambda_q(L)^6h^{6q}g^6\int_{\Omega_q\times\mathbb{R}}\lambda_q(L^2)R(K)h^q gf(\mathbf{x},x)f(\mathbf{x},x)^3\,dx\,\omega_q(d\mathbf{x})\\
&=\lambda_q(L)^6\lambda_q(L^2)R(K)h^{7q}g^7R(f^2).
\end{aligned}
$$

The rest of them have order $\mathcal{O}\left(h^{8q}g^8\right)$, something which can be seen by iteratively applying the Lemma A.12 as before.

Proof of (A.8). It suffices to apply the tower property, the Cauchy-Schwartz inequality, result $\mathbb{E}\left[\left(I_{n,1}^{(2)}\right)^4\right]=\mathcal{O}\left(n^{-4}(h^8+g^8)\right)$ from Lemma A.2 and (A.6):

$$
\begin{aligned}
\mathbb{E}\left[M_n^2(\mathbf{X}_1,Z_1)\right]&=4\frac{c_{h,q}(L)^4}{n^4g^4}\mathbb{E}\left[\mathbb{E}\left[I_{n,1}^{(2)}H_n\left((\mathbf{X}_1,Z_1),(\mathbf{X}_2,Z_2)\right)\middle|(\mathbf{X}_1,Z_1)\right]^2\right]\\
&\leq 4\frac{c_{h,q}(L)^4}{n^4g^4}\mathbb{E}\left[\left(I_{n,1}^{(2)}\right)^2H_n^2\left((\mathbf{X}_1,Z_1),(\mathbf{X}_2,Z_2)\right)\right]\\
&\leq 4\frac{c_{h,q}(L)^4}{n^4g^4}\mathbb{E}\left[\left(I_{n,1}^{(2)}\right)^4\right]^{\frac{1}{2}}\mathbb{E}\left[H_n^4\left((\mathbf{X}_1,Z_1),(\mathbf{X}_2,Z_2)\right)\right]^{\frac{1}{2}}\\
&=\mathcal{O}\left((nh^qg)^{-4}\right)\mathcal{O}\left(n^{-4}(h^8+g^8)\right)^{\frac{1}{2}}\mathcal{O}\left(h^{5q}g^5\right)^{\frac{1}{2}}\\
&=\mathcal{O}\left(n^{-6}(h^4+g^4)h^{-\frac{3q}{2}}g^{-\frac{3}{2}}\right).
\end{aligned}
$$

□

A.1.2 Testing independence with directional data

Lemma A.5. *Under conditions **A1**–**A3**,*

$$
n(h^qg)^{\frac{1}{2}}\left(T_{n,1}-\frac{R(K)\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^qg}\right)\xrightarrow{d}\mathcal{N}\left(0,2\sigma^2\right).
$$

Proof of Lemma A.5. By the decomposition of I_n in the proof of the Theorem 6.1, $T_{n,1}=I_{n,2}+I_{n,3}$ and therefore by (6.9) and (6.12),

$$
T_{n,1}=\mathbb{E}\left[I_{n,2}\right]+\mathcal{O}_{\mathbb{P}}\left(n^{-\frac{3}{2}}h^{-q}g^{-1}\right)+2^{\frac{1}{2}}\sigma n^{-1}(h^qg)^{-\frac{1}{2}}N_n,
$$

where N_n is asymptotically a normal. On the other hand, by (6.9),

$$
\mathbb{E}\left[I_{n,2}\right]=\frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^qg}+\mathcal{O}\left(n^{-1}\right)
$$

and then

$$T_{n,1} = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g} + 2^{\frac{1}{2}}\sigma n^{-1}(h^q g)^{-\frac{1}{2}}N_n + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{3}{2}}h^{-q}g^{-1}\right),$$

because $\left(n^{\frac{3}{2}}h^q g\right)^{-1} = o((nh^{\frac{q}{2}}g^{\frac{1}{2}})^{-1})$. As the last addend is asymptotically negligible compared with the second,

$$n(h^q g)^{\frac{1}{2}}\left(T_{n,1} - \frac{R(K)\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q g}\right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma^2\right).$$

□

Lemma A.6. *Under independence and conditions **A1**–**A3**,*

$$\begin{aligned}\mathbb{E}\left[T_{n,2}\right] &= \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(f_Z)}{nh^q} + \frac{R(K)R(f_{\mathbf{X}})}{ng} + o\left(n^{-1}(h^{-q}+g^{-1})\right),\\ \mathbb{V}\mathrm{ar}\left[T_{n,2}\right] &= \mathcal{O}\left(n^{-2}(h^{-q}+g^{-1})\right).\end{aligned}$$

Proof of Lemma A.6. The term $T_{n,2}$ can be decomposed using the relation

$$\hat{f}_h(\mathbf{x})\hat{f}_g(z) - \mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]\mathbb{E}\left[\hat{f}_g(z)\right] = S_1(\mathbf{x},z) + S_2(\mathbf{x},z) + S_3(\mathbf{x},z),$$

where:

$$\begin{aligned}S_1(\mathbf{x},z) &= \left(\hat{f}_h(\mathbf{x}) - \mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]\right)\left(\hat{f}_g(z) - \mathbb{E}\left[\hat{f}_g(z)\right]\right),\\ S_2(\mathbf{x},z) &= \left(\hat{f}_h(\mathbf{x}) - \mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]\right)\mathbb{E}\left[\hat{f}_g(z)\right],\\ S_3(\mathbf{x},z) &= \left(\hat{f}_g(z) - \mathbb{E}\left[\hat{f}_g(z)\right]\right)\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right].\end{aligned}$$

Hence,

$$\begin{aligned}T_{n,2} =& \int_{\Omega_q\times\mathbb{R}} S_1^2(\mathbf{x},z)\,dz\,\omega_q(d\mathbf{x}) + \int_{\Omega_q\times\mathbb{R}} S_2^2(\mathbf{x},z)\,dz\,\omega_q(d\mathbf{x})\\ &+ \int_{\Omega_q\times\mathbb{R}} S_3^2(\mathbf{x},z)\,dz\,\omega_q(d\mathbf{x}) + 2\int_{\Omega_q\times\mathbb{R}} S_1(\mathbf{x},z)S_2(\mathbf{x},z)\,dz\,\omega_q(d\mathbf{x})\\ &+ 2\int_{\Omega_q\times\mathbb{R}} S_1(\mathbf{x},z)S_3(\mathbf{x},z)\,dz\,\omega_q(d\mathbf{x}) + 2\int_{\Omega_q\times\mathbb{R}} S_2(\mathbf{x},z)S_3(\mathbf{x},z)\,dz\,\omega_q(d\mathbf{x})\\ =& T_{n,2}^{(1)} + T_{n,2}^{(2)} + T_{n,2}^{(3)} + T_{n,2}^{(4)} + T_{n,2}^{(5)} + T_{n,2}^{(6)}.\end{aligned}$$

To compute the expectation of each addend under independence, use the variance and expectation expansions for the directional and linear estimator (see for example García-Portugués et al. (2013b) for both) and relation (6.1). Recall that due to assumption **A1** it is possible to consider Taylor expansions on the marginal densities that have uniform remaining orders.

$$\begin{aligned}\mathbb{E}\left[T_{n,2}^{(1)}\right] &= \int_{\Omega_q}\mathbb{V}\mathrm{ar}\left[\hat{f}_h(\mathbf{x})\right]\omega_q(d\mathbf{x})\int_{\mathbb{R}}\mathbb{V}\mathrm{ar}\left[\hat{f}_g(z)\right]dz\\ &= \mathcal{O}\left((n^2h^q g)^{-1}\right),\\ \mathbb{E}\left[T_{n,2}^{(2)}\right] &= \int_{\Omega_q}\mathbb{V}\mathrm{ar}\left[\hat{f}_h(\mathbf{x})\right]\omega_q(d\mathbf{x})\int_{\mathbb{R}}\mathbb{E}\left[\hat{f}_g(z)\right]^2 dz\end{aligned}$$

$$
\begin{aligned}
&= \left[\frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q} + \mathcal{O}\left(n^{-1}\right)\right][R(f_Z) + o(1)] \\
&= \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(f_Z)}{nh^q} + o\left((nh^q)^{-1}\right), \\
\mathbb{E}\left[T_{n,2}^{(3)}\right] &= \int_{\mathbb{R}} \mathbb{V}\mathrm{ar}\left[\hat{f}_g(z)\right] dz \int_{\Omega_q} \mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]^2 \omega_q(d\mathbf{x}) \\
&= \left[\frac{R(K)}{ng} + \mathcal{O}\left(n^{-1}\right)\right][R(f_{\mathbf{X}}) + o(1)] \\
&= \frac{R(K)R(f_{\mathbf{X}})}{ng} + o\left((ng)^{-1}\right).
\end{aligned}
$$

The expectation of $T_{n,2}^{(4)}$, $T_{n,2}^{(5)}$ and $T_{n,2}^{(6)}$ is zero because of the separability of the directional and linear components. Joining these results,

$$
\mathbb{E}[T_{n,2}] = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(f_Z)}{nh^q} + \frac{R(K)R(f_{\mathbf{X}})}{ng} + o\left(n^{-1}(h^{-q} + g^{-1})\right),
$$

because $(n^2h^qg)^{-1} = o\left(n^{-1}\left(h^{-q} + g^{-1}\right)\right)$.

Computing the variance is not so straightforward as the expectation and some extra results are needed. First of all, recall that by the formula of the variance of the sum, the Cauchy-Schwartz inequality and Lemma A.12,

$$
\mathbb{V}\mathrm{ar}[T_{n,2}] = \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{6} T_{n,2}^{(i)}\right] = \sum_{i=1}^{6} \mathcal{O}\left(\mathbb{V}\mathrm{ar}\left[T_{n,2}^{(i)}\right]\right).
$$

Then the variance of each addend will be computed separately. For that purpose, recall that by the decomposition of the ISE given in Theorem 6.1,

$$
\int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - \mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x}, z)\right]\right)^2 dz\, \omega_q(d\mathbf{x}) = I_{n,2} + I_{n,3},
$$

so by equations (6.9) and (6.11),

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}[I_{n,2} + I_{n,3}] &= \mathcal{O}\left(\mathbb{V}\mathrm{ar}[I_{n,2}] + \mathbb{V}\mathrm{ar}[I_{n,3}]\right) \\
&= \mathcal{O}\left((n^3h^qg)^{-1} + (n^2h^qg)^{-1}\right) \\
&= \mathcal{O}\left((n^2h^qg)^{-1}\right), \\
\mathbb{E}\left[(I_{n,2} + I_{n,3})^2\right] &= \mathbb{V}\mathrm{ar}[I_{n,2} + I_{n,3}] + \mathbb{E}[I_{n,2} + I_{n,3}]^2 \\
&= \mathcal{O}\left((n^2h^qg)^{-1} + (nh^qg)^{-2}\right) \\
&= \mathcal{O}\left((nh^qg)^{-2}\right).
\end{aligned}
$$

The marginal directional and linear versions of these relations will be required:

$$
\mathbb{E}\left[\left(\int_{\Omega_q} \left(\hat{f}_h(\mathbf{x}) - \mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]\right)^2 \omega_q(d\mathbf{x})\right)^2\right] = \mathcal{O}\left((nh^q)^{-2}\right),
$$

$$\mathbb{E}\left[\left(\int_{\mathbb{R}}\left(\hat{f}_g(z)-\mathbb{E}\left[\hat{f}_g(z)\right]\right)^2 dz\right)^2\right]=\mathcal{O}\left((ng)^{-2}\right),$$
$$\mathbb{V}\mathrm{ar}\left[\int_{\Omega_q}\left(\hat{f}_h(\mathbf{x})-\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]\right)^2 \omega_q(d\mathbf{x})\right]=\mathcal{O}\left((n^2h^q)^{-1}\right),$$
$$\mathbb{V}\mathrm{ar}\left[\int_{\mathbb{R}}\left(\hat{f}_g(z)-\mathbb{E}\left[\hat{f}_g(z)\right]\right)^2 dz\right]=\mathcal{O}\left((n^2g)^{-1}\right).$$

Then:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[T_{n,2}^{(1)}\right] &\leq \mathbb{E}\left[\left(T_{n,2}^{(1)}\right)^2\right] \\
&= \mathbb{E}\left[\left(\int_{\Omega_q}\left(\hat{f}_h(\mathbf{x})-\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]\right)^2 \omega_q(d\mathbf{x})\right)^2\right]\mathbb{E}\left[\left(\int_{\mathbb{R}}\left(\hat{f}_g(z)-\mathbb{E}\left[\hat{f}_g(z)\right]\right)^2 dz\right)^2\right] \\
&= \mathcal{O}\left((nh^q)^{-2}\right)\mathcal{O}\left((ng)^{-2}\right) \\
&= \mathcal{O}\left(n^{-4}h^{-2q}g^{-2}\right), \\
\mathbb{V}\mathrm{ar}\left[T_{n,2}^{(2)}\right] &= \mathbb{V}\mathrm{ar}\left[\left(\int_{\Omega_q}\left(\hat{f}_h(\mathbf{x})-\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]\right)^2 \omega_q(d\mathbf{x})\right)\left(\int_{\mathbb{R}}\mathbb{E}\left[\hat{f}_g(z)\right]^2 dz\right)\right] \\
&= \left(\int_{\mathbb{R}}\mathbb{E}\left[\hat{f}_g(z)\right]^2 dz\right)^2 \mathbb{V}\mathrm{ar}\left[\int_{\Omega_q}\left(\hat{f}_h(\mathbf{x})-\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]\right)^2 \omega_q(d\mathbf{x})\right] \\
&= \mathcal{O}(1)\,\mathcal{O}\left((n^2h^q)^{-1}\right) \\
&= \mathcal{O}\left((n^2h^q)^{-1}\right), \\
\mathbb{V}\mathrm{ar}\left[T_{n,2}^{(3)}\right] &= \mathbb{V}\mathrm{ar}\left[\left(\int_{\mathbb{R}}\left(\hat{f}_g(z)-\mathbb{E}\left[\hat{f}_g(z)\right]\right)^2 dz\right)\left(\int_{\Omega_q}\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]^2 \omega_q(d\mathbf{x})\right)\right] \\
&= \left(\int_{\Omega_q}\mathbb{E}\left[\hat{f}_h(\mathbf{x})\right]^2 \omega_q(d\mathbf{x})\right)^2 \mathbb{V}\mathrm{ar}\left[\int_{\mathbb{R}}\left(\hat{f}_g(z)-\mathbb{E}\left[\hat{f}_g(z)\right]\right)^2 dz\right] \\
&= \mathcal{O}(1)\,\mathcal{O}\left((n^2g)^{-1}\right) \\
&= \mathcal{O}\left((n^2g)^{-1}\right).
\end{aligned}$$

The next results follows from applying iteratively Cauchy-Schwartz and the previous orders:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[T_{n,2}^{(4)}\right] &\leq \mathbb{E}\left[\left(T_{n,2}^{(4)}\right)^2\right] \\
&\leq \mathbb{E}\left[\left(\int_{\Omega_q\times\mathbb{R}} S_1^2(\mathbf{x},z)\,dz\,\omega_q(d\mathbf{x})\right)\left(\int_{\Omega_q\times\mathbb{R}} S_2^2(\mathbf{y},t)\,dt\,\omega_q(d\mathbf{y})\right)\right] \\
&\leq \mathbb{E}\left[\left(T_{n,2}^{(1)}\right)^2\right]^{\frac{1}{2}}\mathbb{E}\left[\left(T_{n,2}^{(2)}\right)^2\right]^{\frac{1}{2}} \\
&= \mathcal{O}\left(n^{-3}h^{-2q}\right), \\
\mathbb{V}\mathrm{ar}\left[T_{n,2}^{(5)}\right] &\leq \mathbb{E}\left[\left(T_{n,2}^{(1)}\right)^2\right]^{\frac{1}{2}}\mathbb{E}\left[\left(T_{n,2}^{(3)}\right)^2\right]^{\frac{1}{2}} \\
&= \mathcal{O}\left(n^{-3}g^{-2}\right), \\
\mathbb{V}\mathrm{ar}\left[T_{n,2}^{(6)}\right] &\leq \mathbb{E}\left[\left(T_{n,2}^{(2)}\right)^2\right]^{\frac{1}{2}}\mathbb{E}\left[\left(T_{n,2}^{(3)}\right)^2\right]^{\frac{1}{2}}
\end{aligned}$$

$$= \mathcal{O}\left(n^{-2}\right).$$

Therefore, the order of $\mathbb{V}\mathrm{ar}\left[T_{n,2}\right]$ is $\mathcal{O}\left(n^{-2}(h^{-q}+g^{-1})\right)$ since it dominates $\mathcal{O}\left(n^{-4}h^{-2q}g^{-2}\right)$, $\mathcal{O}\left(n^{-3}(h^{-2q}+g^{-2})\right)$ and $\mathcal{O}\left(n^{-2}\right)$ by assumption **A3**. □

Lemma A.7. *Under independence and conditions **A1**–**A3**,* $\mathbb{E}\left[T_{n,3}\right] = -2\mathbb{E}\left[T_{n,2}\right]$ *and* $\mathbb{V}\mathrm{ar}\left[T_{n,3}\right] = \mathcal{O}\left(n^{-2}(h^{-q}+g^{-1})\right)$.

Proof of Lemma A.7. The term $T_{n,3}$ can be split in a similar fashion to $T_{n,2}$. Let denote

$$S_4(\mathbf{x},z) = \hat{f}_{h,g}(\mathbf{x},z) - \mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x},z)\right].$$

Then:

$$\begin{aligned} T_{n,3} =& -2\int_{\Omega_q\times\mathbb{R}} S_4(\mathbf{x},z)\left(S_1(\mathbf{x},z)+S_2(\mathbf{x},z)+S_3(\mathbf{x},z)\right) dz\,\omega_q(d\mathbf{x}) \\ =& -2\left(T_{n,3}^{(1)}+T_{n,3}^{(2)}+T_{n,3}^{(3)}\right). \end{aligned}$$

The key idea now is to use that, under independence,

$$\begin{aligned} LK_n\left((\mathbf{x},z),(\mathbf{X},Z)\right) =& L_n\left(\mathbf{x},\mathbf{X}\right)K_n\left(z,Z\right) + L_n\left(\mathbf{x},\mathbf{X}\right)\mathbb{E}\left[K\left(\frac{z-Z}{g}\right)\right] \\ &+ K_n\left(z,Z\right)\mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right], \end{aligned} \tag{A.10}$$

where L_n and K_n are the marginal versions of LK_n:

$$L_n\left(\mathbf{x},\mathbf{y}\right) = L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) - \mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right], \quad K_n\left(z,t\right) = K\left(\frac{z-t}{g}\right) - \mathbb{E}\left[K\left(\frac{z-Z}{g}\right)\right].$$

By repeated use of (A.10) in the integrands of $T_{n,3}$ and applying the Fubini theorem, it follows:

$$\begin{aligned} \mathbb{E}\left[T_{n,3}^{(1)}\right] =& \frac{c_{h,q}(L)^2}{n^2g^2}\int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[LK_n\left((\mathbf{x},z),(\mathbf{X},Z)\right)L_n\left(\mathbf{x},\mathbf{X}\right)K_n\left(z,Z\right)\right] dz\,\omega_q(d\mathbf{x}) \\ =& \frac{c_{h,q}(L)^2}{n^2g^2}\int_{\Omega_q\times\mathbb{R}} \mathbb{E}\Bigg[L_n\left(\mathbf{x},\mathbf{X}\right)^2K_n\left(z,Z\right)^2 + L_n\left(\mathbf{x},\mathbf{X}\right)^2K_n\left(z,Z\right)\mathbb{E}\left[K\left(\frac{z-Z}{g}\right)\right] \\ &+ L_n\left(\mathbf{x},\mathbf{X}\right)K_n\left(z,Z\right)^2\mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right]\Bigg] dz\,\omega_q(d\mathbf{x}) \\ =& \frac{c_{h,q}(L)^2}{n^2g^2}\int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[L_n\left(\mathbf{x},\mathbf{X}\right)^2\right]\mathbb{E}\left[K_n\left(z,Z\right)^2\right] dz\,\omega_q(d\mathbf{x}) \\ =& \int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[S_1(\mathbf{x},z)^2\right] dz\,\omega_q(d\mathbf{x}) \\ =& \mathbb{E}\left[T_{n,2}^{(1)}\right], \\ \mathbb{E}\left[T_{n,3}^{(2)}\right] =& \frac{c_{h,q}(L)^2}{n^2g^2}\int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[LK_n\left((\mathbf{x},z),(\mathbf{X},Z)\right)L_n\left(\mathbf{x},\mathbf{X}\right)\mathbb{E}\left[K\left(\frac{z-Z}{g}\right)\right]\right] dz\,\omega_q(d\mathbf{x}) \\ =& \frac{c_{h,q}(L)^2}{n^2g^2}\int_{\Omega_q\times\mathbb{R}} \mathbb{E}\Bigg[L_n\left(\mathbf{x},\mathbf{X}\right)^2K_n\left(z,Z\right)\mathbb{E}\left[K\left(\frac{z-Z}{g}\right)\right] \end{aligned}$$

$$
\begin{aligned}
&+ L_n(\mathbf{x},\mathbf{X})\, K_n(z,Z)\, \mathbb{E}\left[K\left(\frac{z-Z}{g}\right)\right] \mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right] \\
&+ L_n(\mathbf{x},\mathbf{X})^2\, \mathbb{E}\left[K\left(\frac{z-Z}{g}\right)\right]^2 \Bigg]\, dz\, \omega_q(d\mathbf{x}) \\
&= \frac{c_{h,q}(L)^2}{n^2 g^2} \int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[L_n(\mathbf{x},\mathbf{X})^2\right] \mathbb{E}\left[K\left(\frac{z-Z}{g}\right)\right]^2 dz\, \omega_q(d\mathbf{x}) \\
&= \int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[S_2(\mathbf{x},z)^2\right] dz\, \omega_q(d\mathbf{x}) \\
&= \mathbb{E}\left[T_{n,2}^{(2)}\right],
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[T_{n,3}^{(3)}\right] &= \frac{c_{h,q}(L)^2}{n^2 g^2} \int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[LK_n\left((\mathbf{x},z),(\mathbf{X},Z)\right) K_n(z,Z)\, \mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right]\right] dz\, \omega_q(d\mathbf{x}) \\
&= \frac{c_{h,q}(L)^2}{n^2 g^2} \int_{\Omega_q\times\mathbb{R}} \mathbb{E}\Bigg[L_n(\mathbf{x},\mathbf{X})\, K_n(z,Z)\, \mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right] \\
&\quad + L_n(\mathbf{x},\mathbf{X})\, K_n(z,Z)\, \mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right] \mathbb{E}\left[K\left(\frac{z-Z}{g}\right)\right] \\
&\quad + K_n(z,Z)^2\, \mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right]^2 \Bigg] dz\, \omega_q(d\mathbf{x}) \\
&= \frac{c_{h,q}(L)^2}{n^2 g^2} \int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[K_n(z,Z)^2\right] \mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right]^2 dz\, \omega_q(d\mathbf{x}) \\
&= \int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[S_3(\mathbf{x},z)^2\right] dz\, \omega_q(d\mathbf{x}) \\
&= \mathbb{E}\left[T_{n,2}^{(3)}\right].
\end{aligned}
$$

Then $\mathbb{E}[T_{n,3}] = -2\mathbb{E}[T_{n,2}]$.

Computing the variance is much more tedious: the order obtained by bounding the variances by repeated use of the Cauchy-Schwartz inequality is not enough. Instead of, a laborious decomposition of the term $T_{n,3}$ has to be done in order to compute separately the variance of each addend, by following the steps of Rosenblatt and Wahlen (1992). The first step is to split the variance using the Cauchy-Schwartz inequality and Lemma A.12:

$$
\mathbb{V}\mathrm{ar}[T_{n,3}] = \mathcal{O}\left(\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1)}\right] + \mathbb{V}\mathrm{ar}\left[T_{n,3}^{(2)}\right] + \mathbb{V}\mathrm{ar}\left[T_{n,3}^{(3)}\right]\right).
$$

Each of the three terms will be also decomposed into other addends. To simplify their computation the following notation will be employed:

$$
\begin{aligned}
&CLK_n((\mathbf{x}_1,z_1),(\mathbf{X}_1,Z_1);(\mathbf{x}_2,z_2),(\mathbf{X}_2,Z_2)) \\
&\qquad = \mathbb{C}\mathrm{ov}\left[LK\left(\frac{1-\mathbf{x}_1^T\mathbf{X}_1}{h^2}, \frac{z_1-Z_1}{g}\right), LK\left(\frac{1-\mathbf{x}_2^T\mathbf{X}_2}{h^2}, \frac{z_2-Z_2}{g}\right)\right]
\end{aligned}
$$

and also its marginal versions:

$$
CL_n(\mathbf{x}_1,\mathbf{X}_1;\mathbf{x}_2,\mathbf{X}_2) = \mathbb{C}\mathrm{ov}\left[L\left(\frac{1-\mathbf{x}_1^T\mathbf{X}_1}{h^2}\right), L\left(\frac{1-\mathbf{x}_2^T\mathbf{X}_2}{h^2}\right)\right],
$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$CK_n(z_1, Z_1; z_2, Z_2) = \mathbb{C}\text{ov}\left[K\left(\frac{z_1 - Z_1}{g}\right), K\left(\frac{z_2 - Z_2}{g}\right)\right].$$

Term $T_{n,3}^{(2)}$. To begin with, let examine $T_{n,3}^{(2)}$ using the notation of LK_n, L_n and K_n:

$$T_{n,3}^{(2)} = \frac{c_{h,q}(L)^2}{n^2 g^2} \sum_{i=1}^{n}\sum_{j=1}^{n} \int_{\Omega_q \times \mathbb{R}} LK_n((\mathbf{x}, z), (\mathbf{X}_i, Z_i)) L_n(\mathbf{x}, \mathbf{X}_j) \mathbb{E}\left[K\left(\frac{z - Z}{g}\right)\right] dz\, \omega_q(d\mathbf{x}).$$

where the double summation can be split into two summations (a single sum plus the sum of the cross terms). Then,

$$\mathbb{V}\text{ar}\left[T_{n,3}^{(2)}\right] = \frac{c_{h,q}(L)^4}{n^4 g^4} \mathcal{O}\left(n\mathbb{V}\text{ar}\left[T_{n,3}^{(2,1)}\right] + n^2 \mathbb{V}\text{ar}\left[T_{n,3}^{(2,2)}\right]\right),$$

where:

$$T_{n,3}^{(2,1)} = \int_{\Omega_q \times \mathbb{R}} LK_n((\mathbf{x}, z), (\mathbf{X}, Z)) L_n(\mathbf{x}, \mathbf{X}) \mathbb{E}\left[K\left(\frac{z - Z}{g}\right)\right] dz\, \omega_q(d\mathbf{x}),$$
$$T_{n,3}^{(2,2)} = \int_{\Omega_q \times \mathbb{R}} LK_n((\mathbf{x}, z), (\mathbf{X}_1, Z_1)) L_n(\mathbf{x}, \mathbf{X}_2) \mathbb{E}\left[K\left(\frac{z - Z}{g}\right)\right] dz\, \omega_q(d\mathbf{x}).$$

The first term is computed by

$$\begin{aligned}
\mathbb{V}\text{ar}\left[T_{n,3}^{(2,1)}\right] &\leq \mathbb{E}\left[(T_{n,3}^{(2,1)})^2\right] \\
&= \mathcal{O}\left(g^2\right) \mathbb{E}\left[\left(\int_{\Omega_q \times \mathbb{R}} LK_n((\mathbf{x}, z), (\mathbf{X}, Z)) L_n(\mathbf{x}, \mathbf{X})\, dz\, \omega_q(d\mathbf{x})\right)^2\right] \\
&= \mathcal{O}\left(g^2\right) \mathbb{E}\left[\left(\int_{\Omega_q \times \mathbb{R}} \left[LK\left(\frac{1 - \mathbf{x}^T\mathbf{X}}{h^2}, \frac{z - Z}{g}\right) - \mathcal{O}\left(h^q g\right)\right]\right.\right. \\
&\quad \left.\left. \times \left[L\left(\frac{1 - \mathbf{x}^T\mathbf{X}}{h^2}\right) - \mathcal{O}\left(h^q\right)\right] dz\, \omega_q(d\mathbf{x})\right)^2\right] \\
&= \mathcal{O}\left(g^2\right) \left(\int_{\Omega_{q-1}} \int_0^{2h^{-2}} \int_{\mathbb{R}} \left[LK\left(r, t\right) - \mathcal{O}\left(h^q g\right)\right] \left[L\left(r\right) - \mathcal{O}\left(h^q\right)\right]\right. \\
&\quad \left. \times h^q (2 - h^2 r)^{\frac{q}{2}-1} r^{\frac{q}{2}-1} g\, dz\, dr\, \omega_{q-1}(d\boldsymbol{\xi})\right)^2 \\
&= \mathcal{O}\left(h^{2q} g^4\right),
\end{aligned}$$

where the second equality follows from $\mathbb{E}[LK(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}, \frac{z-Z}{g})] = \mathcal{O}\left(h^q g\right)$ and $\mathbb{E}[L(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2})] = \mathcal{O}\left(h^q\right)$, and the third from applying the changes of variables of the proof of Lemma A.4. The second addend is

$$\begin{aligned}
\mathbb{V}\text{ar}\left[T_{n,3}^{(2,2)}\right] \leq \mathbb{E}\Bigg[&\int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} LK_n((\mathbf{x}_1, z_1), (\mathbf{X}_1, Z_1)) L_n(\mathbf{x}_1, \mathbf{X}_2) \\
&\times \mathbb{E}\left[K\left(\frac{z_1 - Z}{g}\right)\right] LK_n((\mathbf{x}_2, z_2), (\mathbf{X}_1, Z_1)) L_n(\mathbf{x}_2, \mathbf{X}_2) \\
&\times \mathbb{E}\left[K\left(\frac{z_2 - Z}{g}\right)\right] dz_1\, \omega_q(d\mathbf{x}_1)\, dz_2\, \omega_q(d\mathbf{x}_2)\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} CLK_n((\mathbf{x}_1,z_1),(\mathbf{X}_1,Z_1);(\mathbf{x}_2,z_2),(\mathbf{X}_1,Z_1)) \\
&\quad\times CL_n(\mathbf{x}_1,\mathbf{X}_2;\mathbf{x}_2,\mathbf{X}_2)\mathbb{E}\left[K\left(\frac{z_1-Z}{g}\right)\right]\mathbb{E}\left[K\left(\frac{z_2-Z}{g}\right)\right] \\
&\quad\times\ dz_1\,\omega_q(d\mathbf{x}_1)\,dz_2\,\omega_q(d\mathbf{x}_2) \\
&\leq \Big(\int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} CLK_n((\mathbf{x}_1,z_1),(\mathbf{X}_1,Z_1);(\mathbf{x}_2,z_2),(\mathbf{X}_1,Z_1)) \\
&\quad\times\ dz_1\,\omega_q(d\mathbf{x}_1)\,dz_2\,\omega_q(d\mathbf{x}_2)\Big)\mathcal{O}\left(h^q g^2\right),
\end{aligned}
$$

because $CL_n(\mathbf{x}_1,\mathbf{X}_2;\mathbf{x}_2,\mathbf{X}_2)=\mathcal{O}\left(h^q\right)$ by Cauchy-Schwartz and the directional version of Lemma A.11, and $\mathbb{E}[K(\frac{z-Z}{g})]=\mathcal{O}\left(g\right)$. Also, the integral of the covariance is

$$
\begin{aligned}
&\int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} CLK_n((\mathbf{x}_1,z_1),(\mathbf{X}_1,Z_1);(\mathbf{x}_2,z_2),(\mathbf{X}_1,Z_1))\,dz_1\,\omega_q(d\mathbf{x}_1)\,dz_2\,\omega_q(d\mathbf{x}_2) \\
&\quad= \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[LK\left(\frac{1-\mathbf{x}_1^T\mathbf{X}_1}{h^2},\frac{z_1-Z_1}{g}\right)LK\left(\frac{1-\mathbf{x}_2^T\mathbf{X}_1}{h^2},\frac{z_2-Z_1}{g}\right)\right] \\
&\qquad\times\ dz_1\,\omega_q(d\mathbf{x}_1)\,dz_2\,\omega_q(d\mathbf{x}_2)-\mathcal{O}\left(h^{2q}g^2\right) \\
&\quad= \int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} LK\left(\frac{1-\mathbf{x}_1^T\mathbf{y}}{h^2},\frac{z_1-t}{g}\right)LK\left(\frac{1-\mathbf{x}_2^T\mathbf{y}}{h^2},\frac{z_2-t}{g}\right) \\
&\qquad\times\ f(\mathbf{y},t)\,dt\,\omega_q(d\mathbf{y})\,dz_1\,\omega_q(d\mathbf{x}_1)\,dz_2\,\omega_q(d\mathbf{x}_2)-\mathcal{O}\left(h^{2q}g^2\right) \\
&\quad=\mathcal{O}\left(h^{2q}g^2\right),
\end{aligned}
$$

as it follows that the order of the first addend is $\mathcal{O}\left(h^{2q}g^2\right)$ by applying *i–ix* in the same way as in the calculus of A_1 in Lemma A.4 (recall that the square in A_1 is not present here and therefore the order is larger). Then $\mathbb{V}\mathrm{ar}[T_{n,3}^{(2,2)}]=\mathcal{O}\left(h^{3q}g^4\right)$ and as a consequence,

$$
\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(2)}\right]=\frac{c_{h,q}(L)^4}{n^4g^4}\mathcal{O}\left(nh^{2q}g^4+n^2h^{3q}g^4\right)=\mathcal{O}\left(n^{-2}h^{-q}\right). \tag{A.11}
$$

Term $T_{n,3}^{(3)}$. This addend follows analogously from $T_{n,3}^{(2)}$, as the only difference is the swapping of the roles of the directional and linear components:

$$
T_{n,3}^{(3)}=\frac{c_{h,q}(L)^2}{n^2g^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x},z),(\mathbf{X}_i,Z_i))K_n(z,Z_j)\mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right]dz\,\omega_q(d\mathbf{x}),
$$

with the same decomposition that gives

$$
\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(3)}\right]=\frac{c_{h,q}(L)^4}{n^4g^4}\mathcal{O}\left(n\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(3,1)}\right]+n^2\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(3,2)}\right]\right),
$$

where:

$$
T_{n,3}^{(3,1)}=\int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x},z),(\mathbf{X},Z))K_n(z,Z)\mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right]dz\,\omega_q(d\mathbf{x}),
$$

$$T_{n,3}^{(3,2)} = \int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x},z),(\mathbf{X}_1,Z_1))K_n(z,Z_2)\mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)\right] dz\,\omega_q(d\mathbf{x}).$$

Then, by similar computations to those of $T_{n,3}^{(3)}$, $\mathbb{V}\mathrm{ar}\big[T_{n,3}^{(3,1)}\big] = \mathcal{O}\left(h^{4q}g^2\right)$, $\mathbb{V}\mathrm{ar}\big[T_{n,3}^{(3,2)}\big] = \mathcal{O}\left(h^{4q}g^3\right)$ and

$$\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(3)}\right] = \frac{c_{h,q}(L)^4}{n^4g^4}\mathcal{O}\left(nh^{4q}g^2 + n^2h^{4q}g^3\right) = \mathcal{O}\left(n^{-2}g^{-1}\right). \tag{A.12}$$

Term $T_{n,3}^{(1)}$. This is the hardest part, as it presents more combinations. As with the previous terms,

$$T_{n,3}^{(1)} = \frac{c_{h,q}(L)^2}{n^3g^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x},z),(\mathbf{X}_i,Z_i))L_n(\mathbf{x},\mathbf{X}_j)K_n(z,Z_k)\,dz\,\omega_q(d\mathbf{x}).$$

and now the triple summation can be split into five summations

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1)}\right] = \frac{c_{h,q}(L)^4}{n^6g^4}\mathcal{O}\Big(&n\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1,1)}\right] + n^2\left(\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1,2a)}\right] + \mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1,2b)}\right] + \mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1,2c)}\right]\right)\\ &+ n^3\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1,3)}\right]\Big),\end{aligned}$$

where:

$$\begin{aligned}T_{n,3}^{(1,1)} &= \int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x},z),(\mathbf{X}_1,Z_1))L_n(\mathbf{x},\mathbf{X}_1)K_n(z,Z_1)\,dz\,\omega_q(d\mathbf{x}),\\ T_{n,3}^{(1,2a)} &= \int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x},z),(\mathbf{X}_1,Z_1))L_n(\mathbf{x},\mathbf{X}_2)K_n(z,Z_2)\,dz\,\omega_q(d\mathbf{x}),\\ T_{n,3}^{(1,2b)} &= \int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x},z),(\mathbf{X}_1,Z_1))L_n(\mathbf{x},\mathbf{X}_1)K_n(z,Z_2)\,dz\,\omega_q(d\mathbf{x}),\\ T_{n,3}^{(1,2c)} &= \int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x},z),(\mathbf{X}_1,Z_1))L_n(\mathbf{x},\mathbf{X}_2)K_n(z,Z_1)\,dz\,\omega_q(d\mathbf{x}),\\ T_{n,3}^{(1,3)} &= \int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x},z),(\mathbf{X}_1,Z_1))L_n(\mathbf{x},\mathbf{X}_2)K_n(z,Z_3)\,dz\,\omega_q(d\mathbf{x}).\end{aligned}$$

The first term is computed by

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1,1)}\right] &\leq \mathbb{E}\left[\left(\int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x},z),(\mathbf{X},Z))L_n(\mathbf{x},\mathbf{X})K_n(z,Z)\,dz\,\omega_q(d\mathbf{x})\right)^2\right]\\ &= \mathbb{E}\Bigg[\Bigg(\int_{\Omega_q\times\mathbb{R}}\left[LK\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2},\frac{z-Z}{g}\right) - \mathcal{O}\left(h^qg\right)\right]\\ &\quad\times\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right) - \mathcal{O}\left(h^q\right)\right]\left[K\left(\frac{z-Z}{g}\right) - \mathcal{O}\left(g\right)\right]\,dz\,\omega_q(d\mathbf{x})\Bigg)^2\Bigg]\\ &= \Bigg(\int_{\Omega_{q-1}}\int_0^{2h^{-2}}\int_{\mathbb{R}}\left[LK\left(r,t\right) - \mathcal{O}\left(h^qg\right)\right]\left[L\left(r\right) - \mathcal{O}\left(h^q\right)\right]\\ &\quad\times\left[K\left(t\right) - \mathcal{O}\left(g\right)\right]h^q(2-h^2r)^{\frac{q}{2}-1}r^{\frac{q}{2}-1}g\,dr\,\omega_{q-1}(d\boldsymbol{\xi})\,dz\Bigg)^2\end{aligned}$$

$$= \mathcal{O}\left(h^{2q} g^2\right),$$

by the same arguments as for $T_{n,3}^{(2,1)}$. The fifth addend is

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1,3)}\right] \leq & \mathbb{E}\Bigg[\int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} LK_n((\mathbf{x}_1, z_1), (\mathbf{X}_1, Z_1)) L_n(\mathbf{x}_1, \mathbf{X}_2) K_n(z_1, Z_3) \\
& \times LK_n((\mathbf{x}_2, z_2), (\mathbf{X}_1, Z_1)) L_n(\mathbf{x}_2, \mathbf{X}_2) K_n(z_2, Z_3)\, dz_1\, \omega_q(d\mathbf{x}_1)\, dz_2\, \omega_q(d\mathbf{x}_2) \Bigg] \\
= & \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} CLK_n((\mathbf{x}_1, z_1), (\mathbf{X}_1, Z_1); (\mathbf{x}_2, z_2), (\mathbf{X}_1, Z_1)) \\
& \times CL_n(\mathbf{x}_1, \mathbf{X}_2; \mathbf{x}_2, \mathbf{X}_2) CK_n(z_1, Z_3; z_2, Z_3)\, dz_1\, \omega_q(d\mathbf{x}_1)\, dz_2\, \omega_q(d\mathbf{x}_2) \\
\leq & \mathcal{O}\left(h^q g\right) \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} CLK_n((\mathbf{x}_1, z_1), (\mathbf{X}_1, Z_1); (\mathbf{x}_2, z_2), (\mathbf{X}_1, Z_1)) \\
& \times\ dz_1\, \omega_q(d\mathbf{x}_1)\, dz_2\, \omega_q(d\mathbf{x}_2) \\
\leq & \mathcal{O}\left(h^{3q} g^3\right),
\end{aligned}$$

again by the same arguments used for $T_{n,3}^{(2,2)}$. It only remains to obtain the variance of $T_{n,3}^{(1,2a)}$, $T_{n,3}^{(1,2b)}$ and $T_{n,3}^{(1,2c)}$. The first one arises from

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1,2a)}\right] \leq & \mathbb{E}\Bigg[\int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} LK_n((\mathbf{x}_1, z_1), (\mathbf{X}_1, Z_1)) L_n(\mathbf{x}_1, \mathbf{X}_2) K_n(z_1, Z_2) \\
& \times LK_n((\mathbf{x}_2, z_2), (\mathbf{X}_1, Z_1)) L_n(\mathbf{x}_2, \mathbf{X}_2) K_n(z_2, Z_2)\, dz_1\, \omega_q(d\mathbf{x}_1)\, dz_2\, \omega_q(d\mathbf{x}_2) \Bigg] \\
= & \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} CLK_n((\mathbf{x}_1, z_1), (\mathbf{X}_1, Z_1); (\mathbf{x}_2, z_2), (\mathbf{X}_1, Z_1)) \\
& \times CL_n(\mathbf{x}_1, \mathbf{X}_2; \mathbf{x}_2, \mathbf{X}_2) CK_n(z_1, Z_2; z_2, Z_2)\, dz_1\, \omega_q(d\mathbf{x}_1)\, dz_2\, \omega_q(d\mathbf{x}_2) \\
= & \mathcal{O}\left(h^{3q} g^3\right),
\end{aligned}$$

in virtue of the assumption of independence and the calculus of $\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1,1)}\right]$. The second one is

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1,2b)}\right] \leq & \mathbb{E}\Bigg[\int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} LK_n((\mathbf{x}_1, z_1), (\mathbf{X}_1, Z_1)) L_n(\mathbf{x}_1, \mathbf{X}_1) K_n(z_1, Z_2) \\
& \times LK_n((\mathbf{x}_2, z_2), (\mathbf{X}_1, Z_1)) L_n(\mathbf{x}_2, \mathbf{X}_1) K_n(z_2, Z_2)\, dz_1\, \omega_q(d\mathbf{x}_1)\, dz_2\, \omega_q(d\mathbf{x}_2) \Bigg] \\
= & \int_{\Omega_q \times \mathbb{R}} \int_{\Omega_q \times \mathbb{R}} \mathbb{E}\big[LK_n((\mathbf{x}_1, z_1), (\mathbf{X}_1, Z_1)) LK_n((\mathbf{x}_2, z_2), (\mathbf{X}_1, Z_1)) \\
& \times L_n(\mathbf{x}_1, \mathbf{X}_1) L_n(\mathbf{x}_2, \mathbf{X}_1) \big] CK_n(z_1, Z_2; z_2, Z_2)\, dz_1\, \omega_q(d\mathbf{x}_1)\, dz_2\, \omega_q(d\mathbf{x}_2) \\
= & \mathcal{O}\left(g\right) \mathbb{E}\left[\left(\int_{\Omega_q \times \mathbb{R}} LK_n((\mathbf{x}, z), (\mathbf{X}_1, Z_1)) L_n(\mathbf{x}, \mathbf{X}_1)\, dz\, \omega_q(d\mathbf{x}) \right)^2 \right] \\
= & \mathcal{O}\left(h^{2q} g^3\right),
\end{aligned}$$

where the order of the expectation is obtained again using the change of variables described in the proof of Lemma A.10,

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x},z),(\mathbf{X}_1,Z_1))L_n(\mathbf{x},\mathbf{X}_1)\,dz\,\omega_q(d\mathbf{x})\right)^2\right]\\
&\quad=\mathbb{E}\left[\left(\int_{\Omega_q\times\mathbb{R}}\left[LK\left(\frac{1-\mathbf{x}^T\mathbf{X}_1}{h^2},\frac{z-Z_1}{g}\right)-\mathcal{O}\left(h^q g\right)\right]\right.\right.\\
&\qquad\left.\left.\times\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}_1}{h^2}\right)-\mathcal{O}\left(h^q\right)\right]\,dz\,\omega_q(d\mathbf{x})\right)^2\right]\\
&\quad=\mathbb{E}\left[\left(\int_{\Omega_{q-1}}\int_0^{2h^{-2}}\int_{\mathbb{R}}\left[LK\left(r,u\right)-\mathcal{O}\left(h^q g\right)\right]\left[L\left(r\right)-\mathcal{O}\left(h^q\right)\right]\right.\right.\\
&\qquad\left.\left.\times h^q(2-h^2r)^{\frac{q}{2}-1}r^{\frac{q}{2}-1}g\,du\,dr\,\omega_{q-1}(d\boldsymbol{\xi})\right)^2\right]\\
&\quad=\mathcal{O}\left(h^{2q}g^2\right).
\end{aligned}$$

The variance of $T_{n,3}^{(1,2c)}$ is obtained analogously:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1,2c)}\right]\leq&\,\mathbb{E}\left[\int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x}_1,z_1),(\mathbf{X}_1,Z_1))L_n(\mathbf{x}_1,\mathbf{X}_2)K_n(z_1,Z_1)\right.\\
&\left.\times LK_n((\mathbf{x}_2,z_2),(\mathbf{X}_1,Z_1))L_n(\mathbf{x}_2,\mathbf{X}_2)K_n(z_2,Z_1)\,dz_1\,\omega_q(d\mathbf{x}_1)\,dz_2\,\omega_q(d\mathbf{x}_2)\right]\\
=&\int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}}\mathbb{E}\big[LK_n((\mathbf{x}_1,z_1),(\mathbf{X}_1,Z_1))LK_n((\mathbf{x}_2,z_2),(\mathbf{X}_1,Z_1))\\
&\times K_n(z_1,Z_1)K_n(z_2,Z_1)\big]CL_n(\mathbf{x}_1,\mathbf{X}_2;\mathbf{x}_2,\mathbf{X}_2)\,dz_1\,\omega_q(d\mathbf{x}_1)\,dz_2\,\omega_q(d\mathbf{x}_2)\\
=&\,\mathcal{O}\left(h^q\right)\mathbb{E}\left[\left(\int_{\Omega_q\times\mathbb{R}} LK_n((\mathbf{x},z),(\mathbf{X}_1,Z_1))K_n(z,Z_1)\,dz\,\omega_q(d\mathbf{x})\right)^2\right]\\
=&\,\mathcal{O}\left(h^{3q}g^2\right).
\end{aligned}$$

Then, putting together the variances of $T_{n,3}^{(1,1)}$, $T_{n,3}^{(1,2a)}$, $T_{n,3}^{(1,2b)}$, $T_{n,3}^{(1,2c)}$ and $T_{n,3}^{(1,3)}$, it follows

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[T_{n,3}^{(1)}\right]&=\frac{c_{h,q}(L)^4}{n^6g^4}\mathcal{O}\left(nh^{2q}g^2+n^2(h^{3q}g^3+h^{2q}g^3+h^{3q}g^2)+n^3h^{3q}g^3\right)\\
&=\frac{c_{h,q}(L)^4}{n^6g^4}\mathcal{O}\left(n^3h^{3q}g^3\right)\\
&=\mathcal{O}\left(n^{-3}h^{-q}g^{-1}\right).
\end{aligned}\tag{A.13}$$

Finally, joining (A.11), (A.12) and (A.13),

$$\mathbb{V}\mathrm{ar}\left[T_{n,3}\right]=\mathcal{O}\left(n^{-3}h^{-q}g^{-1}\right)+\mathcal{O}\left(n^{-2}h^{-q}\right)+\mathcal{O}\left(n^{-2}g^{-1}\right)=\mathcal{O}\left(n^{-2}(h^{-q}+g^{-1})\right),$$

which proves the lemma. □

A.1.3 Goodness-of-fit test for models with directional data

Lemma A.8. *Under $H_0 : f = f_{\boldsymbol{\theta}_0}$, with $\boldsymbol{\theta}_0 \in \Theta$ unknown and conditions **A1**–**A3** and **A5**–**A6**, $n(h^q g)^{\frac{1}{2}} R_{n,1} \xrightarrow{p} 0$ and $n(h^q g)^{\frac{1}{2}} R_{n,4} \xrightarrow{p} 0$.*

Proof of Lemma A.8. Under the null $f = f_{\boldsymbol{\theta}_0}$, for a known $\boldsymbol{\theta}_0 \in \Theta$.

Term $R_{n,4}$. Using a first order Taylor expansion of $f_{\hat{\boldsymbol{\theta}}}$ in $\boldsymbol{\theta}_0$,

$$\begin{aligned}
R_{n,4} &= \int_{\Omega_q \times \mathbb{R}} \left(LK_{h,g} \left(f_{\boldsymbol{\theta}_0}(\mathbf{x}, z) - f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) \right) \right)^2 dz\, \omega_q(d\mathbf{x}) \\
&= \int_{\Omega_q \times \mathbb{R}} \left(LK_{h,g} \left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x}, z)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n} \right) \right)^2 dz\, \omega_q(d\mathbf{x}) \\
&\leq ||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||^2 \int_{\Omega_q \times \mathbb{R}} \left(LK_{h,g} \left(\left|\left| \frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x}, z)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n} \right|\right| \right) \right)^2 dz\, \omega_q(d\mathbf{x}) \\
&= \mathcal{O}_{\mathbb{P}} \left(n^{-1} \right) \mathcal{O}_{\mathbb{P}} (1) \\
&= \mathcal{O}_{\mathbb{P}} \left(n^{-1} \right),
\end{aligned}$$

where $\boldsymbol{\theta}_n \in \Theta$ is a certain parameter depending on the sample. The order holds because, on the one hand, $||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||^2 = \mathcal{O}_{\mathbb{P}} (n^{-1})$ by assumption **A6** and on the other, by **A5** and Lemma A.10,

$$\begin{aligned}
&\int_{\Omega_q \times \mathbb{R}} \left(LK_{h,g} \left(\left|\left| \frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x}, z)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n} \right|\right| \right) \right)^2 dz\, \omega_q(d\mathbf{x}) \\
&\qquad = \left(\int_{\Omega_q \times \mathbb{R}} \left|\left| \frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x}, z)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n} \right|\right|^2 dz\, \omega_q(d\mathbf{x}) \right) (1 + o(1)) \\
&\qquad = \mathcal{O}_{\mathbb{P}} (1) .
\end{aligned}$$

Therefore, $R_{n,4} = \mathcal{O}_{\mathbb{P}} (n^{-1})$ and, by assumption **A3**, $n(h^q g)^{\frac{1}{2}} R_{n,4} \xrightarrow{p} 0$.

Term $R_{n,1}$. It follows also by a Taylor expansion of second order centred at $\boldsymbol{\theta}_0$:

$$\begin{aligned}
R_{n,1} &= 2 \frac{c_{h,q}(L)}{ng} \sum_{i=1}^{n} \int_{\Omega_q \times \mathbb{R}} LK_n \left((\mathbf{x}, z), (\mathbf{X}_i, Z_i) \right) LK_{h,g} \left(f_{\boldsymbol{\theta}_0}(\mathbf{x}, z) - f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) \right) dz\, \omega_q(d\mathbf{x}) \\
&= 2 \frac{c_{h,q}(L)}{ng} \sum_{i=1}^{n} \int_{\Omega_q \times \mathbb{R}} LK_n \left((\mathbf{x}, z), (\mathbf{X}_i, Z_i) \right) LK_{h,g} \Bigg((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \frac{\partial f(\mathbf{x}, z)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&\quad + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \frac{\partial^2 f(\mathbf{x}, z)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \Bigg) dz\, \omega_q(d\mathbf{x}) \\
&\leq 2 \frac{c_{h,q}(L)}{ng} \sum_{i=1}^{n} \int_{\Omega_q \times \mathbb{R}} LK_n \left((\mathbf{x}, z), (\mathbf{X}_i, Z_i) \right) \Bigg[||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|| LK_{h,g} \left(\left|\left| \frac{\partial f(\mathbf{x}, z)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right|\right| \right) \\
&\quad + ||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||^2 LK_{h,g} \left(\left|\left| \frac{\partial^2 f(\mathbf{x}, z)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n} \right|\right|_F \right) \Bigg] dz\, \omega_q(d\mathbf{x}) \\
&= ||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|| R_{n,1}^{(1)} + ||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||^2 R_{n,1}^{(2)},
\end{aligned}$$

where $||A||_F$ stands for the Frobenious norm of the matrix A. By Lemma A.10 and assumption of **A5**,

$$R_{n,1}^{(i)} = \mathcal{O}_{\mathbb{P}}\left(\frac{c_{h,q}(L)}{ng}\sum_{i=1}^{n}\int_{\Omega_q\times\mathbb{R}} LK_n\left((\mathbf{x},z),(\mathbf{X}_i,Z_i)\right)\, dz\,\omega_q(d\mathbf{x})\right),$$

for $i = 1, 2$. As a consequence of this and of assumption **A6**, the first addend of $R_{n,1}$ dominates the second. The proof now is based on proving that $R_{n,1}^{(1)} = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$ using the Chebychev inequality and the fact that the integrand of $R_{n,1}^{(1)}$ is deterministic. Now recall that $\mathbb{E}[R_{n,1}^{(i)}] = 0$ and by the proof of (A.5) in Lemma A.4,

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[R_{n,1}^{(1)}\right] &= \frac{c_{h,q}(L)^2}{ng^2}\mathbb{E}\left[\left(\int_{\Omega_q\times\mathbb{R}} LK_n\left((\mathbf{x},z),(\mathbf{X},Z)\right)\, dz\,\omega_q(d\mathbf{x})\right)^2\right] \\
&= \frac{c_{h,q}(L)^2}{ng^2}\int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} \mathbb{E}\left[LK_n\left((\mathbf{x},z),(\mathbf{X},Z)\right) LK_n\left((\mathbf{y},t),(\mathbf{X},Z)\right)\right] \\
&\quad\times\, dz\,\omega_q(d\mathbf{x})\, dt\,\omega_q(d\mathbf{y}) \\
&= \frac{c_{h,q}(L)^2}{ng^2}\int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} \left(E_1\left((\mathbf{x},z),(\mathbf{y},t)\right) - E_2\left((\mathbf{x},z),(\mathbf{y},t)\right)\right) \\
&\quad\times\, dz\,\omega_q(d\mathbf{x})\, dt\,\omega_q(d\mathbf{y}) \\
&= \frac{c_{h,q}(L)^2}{ng^2}\mathcal{O}\left(h^{2q}g^2\right) \\
&= \mathcal{O}\left(n^{-1}\right),
\end{aligned}$$

so by the Chebychev inequality, $R_{n,1}^{(1)} = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$ and as a consequence of **A5**, $R_{n,1} = \mathcal{O}_{\mathbb{P}}\left(n^{-1}\right)$ and $n(h^q g)^{\frac{1}{2}} R_{n,1} \xrightarrow{p} 0$ follows. □

Lemma A.9. *Under the alternative hypothesis (6.6) and conditions **A1**–**A3**, **A5** and **A7**, $n(h^q g)^{\frac{1}{2}}\widetilde{R}_{n,1} \xrightarrow{p} 0$ and $n(h^q g)^{\frac{1}{2}}\widetilde{R}_{n,4} \xrightarrow{p} R(\Delta)$.*

Proof of Lemma A.9. The convergence in probability is obtained using the decompositions $\widetilde{R}_{n,1} = R_{n,1} + \widetilde{R}_{n,1}^{(1)}$ and $\widetilde{R}_{n,4} = R_{n,4} + \widetilde{R}_{n,4}^{(1)} + \widetilde{R}_{n,4}^{(2)}$.

Terms $R_{n,1}$ and $R_{n,4}$. The proofs of $n(h^q g)^{\frac{1}{2}} R_{n,1} \xrightarrow{p} 0$ and $n(h^q g)^{\frac{1}{2}} R_{n,4} \xrightarrow{p} 0$ are analogous to the ones of Lemma A.8 and follow just replacing assumption **A6** by **A7** and H_0 by H_{1P}.

Term $R_{n,1}^{(1)}$. Recall that $\mathbb{E}[\widetilde{R}_{n,1}^{(1)}] = 0$ and its variance, using the same steps as in the proof of $R_{n,1}^{(1)}$ in Lemma A.8, is

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\widetilde{R}_{n,1}^{(1)}\right] &= 4\frac{c_{h,q}(L)^2}{n^2 h^{\frac{q}{2}} g^{\frac{3}{2}}}\int_{\Omega_q\times\mathbb{R}}\int_{\Omega_q\times\mathbb{R}} \left(E_1\left((\mathbf{x},z),(\mathbf{y},t)\right) - E_2\left((\mathbf{x},z),(\mathbf{y},t)\right)\right) \\
&\quad\times LK_{h,g}\Delta(\mathbf{x},z) LK_{h,g}\Delta(\mathbf{y},t)\, dz\,\omega_q(d\mathbf{x})\, dt\,\omega_q(d\mathbf{y}) \\
&= 4\frac{c_{h,q}(L)^2}{n^2 h^{\frac{q}{2}} g^{\frac{3}{2}}}\mathcal{O}\left(h^{2q}g^2\right) \\
&= \mathcal{O}\left((n^2 h^{\frac{q}{2}} g^{\frac{1}{2}})^{-1}\right).
\end{aligned}$$

Then, $\widetilde{R}_{n,1}^{(1)} = \mathcal{O}_{\mathbb{P}}((nh^{\frac{q}{4}}g^{\frac{1}{4}})^{-1})$ and $n(h^q g)^{\frac{1}{2}}\widetilde{R}_{n,1}^{(1)} \xrightarrow{p} 0$.

Term $R_{n,4}^{(1)}$. Applying Lemma A.10,

$$\widetilde{R}_{n,4}^{(1)} = \frac{1}{n(h^q g)^{\frac{1}{2}}} \int_{\Omega_q \times \mathbb{R}} (LK_{h,g}\Delta(\mathbf{x},z))^2 \, dz\, \omega_q(d\mathbf{x}) = \frac{1}{n(h^q g)^{\frac{1}{2}}} R(\Delta)(1+o(1))$$

and as a consequence $n(h^q g)^{\frac{1}{2}}\widetilde{R}_{n,4}^{(1)} \xrightarrow{p} R(\Delta)$.

Term $R_{n,4}^{(2)}$. Applying the Cauchy-Schwartz inequality:

$$\frac{\sqrt{nh^{\frac{q}{2}}g^{\frac{1}{2}}}}{2}\widetilde{R}_{n,4}^{(2)} \leq (R_{n,4})^{\frac{1}{2}}\Big(nh^{\frac{q}{2}}g^{\frac{1}{2}}\widetilde{R}_{n,4}^{(1)}\Big)^{\frac{1}{2}} = \mathcal{O}_{\mathbb{P}}\Big(n^{-\frac{1}{2}}\Big)\mathcal{O}_{\mathbb{P}}(1) = \mathcal{O}_{\mathbb{P}}\Big(n^{-\frac{1}{2}}\Big),$$

Therefore, $\widetilde{R}_{n,4}^{(2)} = \mathcal{O}_{\mathbb{P}}((nh^{\frac{q}{4}}g^{\frac{1}{4}})^{-1})$ and $n(h^q g)^{\frac{1}{2}}\widetilde{R}_{n,4}^{(2)} = \mathcal{O}_{\mathbb{P}}((h^q g)^{\frac{1}{4}}) \xrightarrow{p} 0$. □

A.1.4 General purpose lemmas

For the proofs of some lemmas, three auxiliary lemmas have been used.

Lemma A.10. *Under conditions **A1**–**A3**, for any function $\varphi : \Omega_q \times \mathbb{R} \to \mathbb{R}$ that is uniformly continuous and bounded, the smoothing operator (6.4) satisfies*

$$\sup_{(\mathbf{x},z)\in\Omega_q\times\mathbb{R}} |LK_{h,g}\varphi(\mathbf{x},z) - \varphi(\mathbf{x},z)| \xrightarrow[n\to\infty]{} 0. \tag{A.14}$$

Thus, $LK_{h,g}\varphi(\mathbf{x},z)$ converges to $\varphi(\mathbf{x},z)$ uniformly in $\Omega_q \times \mathbb{R}$.

Proof of Lemma A.10. Let denote $D_n = |LK_{h,g}\varphi(\mathbf{x},z) - \varphi(\mathbf{x},z)|$. Since $\varphi(\mathbf{x},z)$ can be written as $\frac{c_{h,q}(L)}{g}\int_{\Omega_q\times\mathbb{R}} LK(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}, \frac{z-t}{g})\varphi(\mathbf{y},t)\, dz\, \omega_q(d\mathbf{x})$, then

$$\begin{aligned} D_n &= \left|\frac{c_{h,q}(L)}{g}\int_{\Omega_q\times\mathbb{R}} LK\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}, \frac{z-t}{g}\right)(\varphi(\mathbf{y},t)-\varphi(\mathbf{x},z))\, dt\, \omega_q(d\mathbf{y})\right| \\ &\leq \frac{c_{h,q}(L)}{g}\int_{\Omega_q\times\mathbb{R}} LK\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}, \frac{z-t}{g}\right)|\varphi(\mathbf{y},t)-\varphi(\mathbf{x},z)|\, dt\, \omega_q(d\mathbf{y}) \\ &\leq D_{n,1} + D_{n,2}, \end{aligned}$$

where:

$$\begin{aligned} D_{n,1} &= \frac{c_{h,q}(L)}{g}\int_{A_\delta} LK\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}, \frac{z-t}{g}\right)|\varphi(\mathbf{y},t)-\varphi(\mathbf{x},z)|\, dt\, \omega_q(d\mathbf{y}), \\ D_{n,2} &= \frac{c_{h,q}(L)}{g}\int_{\bar{A}_\delta} LK\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}, \frac{z-t}{g}\right)|\varphi(\mathbf{y},t)-\varphi(\mathbf{x},z)|\, dt\, \omega_q(d\mathbf{y}), \\ A_\delta &= \left\{(\mathbf{y},t)\in\Omega_q\times\mathbb{R} : \max\left(\sqrt{2(1-\mathbf{x}^T\mathbf{y})}, |z-t|\right) < \delta\right\}, \\ A_{1,\delta} &= \left\{(\mathbf{y},t)\in\Omega_q\times\mathbb{R} : 1-\mathbf{x}^T\mathbf{y} < \frac{\delta^2}{2}\right\}, \end{aligned}$$

$$A_{2,\delta} = \{(\mathbf{y}, t) \in \Omega_q \times \mathbb{R} : |z - t| < \delta\}$$

and $\bar{A}_\delta$ denotes the complementary set to A_δ for a $\delta > 0$. Recall that $A_\delta = A_{1,\delta} \cap A_{2,\delta}$ and as a consequence $\bar{A}_\delta = \bar{A}_{1,\delta} \cup \bar{A}_{2,\delta}$.

As stated in assumption **A1**, the uniform continuity of the functions defined in $\Omega_q \times \mathbb{R}$ is understood with respect to the product Euclidean norm, that is

$$||(\mathbf{x}, z)||_2 = \sqrt{||\mathbf{x}||^2_{\Omega_q} + ||z||^2_{\mathbb{R}}}, \text{ where } ||\cdot||_{\Omega_q} = ||\cdot||_2 \text{ and } ||\cdot||_{\mathbb{R}} = |\cdot|\,.$$

Nevertheless, given the equivalence between the product 2-norm and the product ∞-norm, defined as $||(\mathbf{x}, z)||_\infty = \max\left(||\mathbf{x}||_{\Omega_q}, ||z||_{\mathbb{R}}\right)$, and for the sake of simplicity, the second norm will be used in the proof. Then, by the uniform continuity of φ, it holds that for any $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\forall (\mathbf{x}, z), (\mathbf{y}, t) \in \Omega_q \times \mathbb{R},\ ||(\mathbf{x}, z) - (\mathbf{y}, t)||_\infty < \delta \implies |\varphi(\mathbf{x}, z) - \varphi(\mathbf{y}, t)| < \varepsilon.$$

Therefore the first term is dominated by

$$D_{n,1} < \varepsilon \frac{c_{h,q}(L)}{g} \int_{A_\delta} LK\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}, \frac{z - t}{g}\right) dt\, \omega_q(d\mathbf{y}) \leq \varepsilon,$$

for any $\varepsilon > 0$, so as a consequence $D_{n,1} = o(1)$ uniformly in $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$.

For the second term, let consider the change of variables introduced in the proof of Lemma A.4 (see Lemma 2 of García-Portugués et al. (2013b) for a detailed derivation):

$$\begin{cases} \mathbf{y} = u\mathbf{x} + (1 - u^2)^{\frac{1}{2}} \mathbf{B}_\mathbf{x} \boldsymbol{\xi}, \\ \omega_q(d\mathbf{y}) = (1 - u^2)^{\frac{q}{2} - 1}\, du\, \omega_{q-1}(d\boldsymbol{\xi}), \end{cases}$$

where $u \in (-1, 1)$, $\boldsymbol{\xi} \in \Omega_{q-1}$ and $\mathbf{B}_\mathbf{x} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$ is the semi-orthonormal matrix resulting from the completion of $\mathbf{x}$ to the orthonormal basis $\{\mathbf{x}, \mathbf{b}_1, \ldots, \mathbf{b}_q\}$. Applying this change of variables and then using the standard changes of variables $r = \frac{1-u}{h^2}$ (for the first addend) and $s = \frac{z-t}{g}$ (second addend), it follows:

$$\begin{aligned}
D_{n,2} &= \frac{c_{h,q}(L)}{g} \int_{\bar{A}_\delta} LK\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}, \frac{z - t}{g}\right) |\varphi(\mathbf{y}, t) - \varphi(\mathbf{x}, z)|\ dt\, \omega_q(d\mathbf{y}) \\
&\leq \frac{c_{h,q}(L)}{g} \int_{\bar{A}_{1,\delta}} LK\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}, \frac{z - t}{g}\right) |\varphi(\mathbf{y}, t) - \varphi(\mathbf{x}, z)|\ dt\, \omega_q(d\mathbf{y}) \\
&\quad + \frac{c_{h,q}(L)}{g} \int_{\bar{A}_{2,\delta}} LK\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}, \frac{z - t}{g}\right) |\varphi(\mathbf{y}, t) - \varphi(\mathbf{x}, z)|\ dt\, \omega_q(d\mathbf{y}) \\
&\leq 2\frac{c_{h,q}(L)}{g} \sup_{(\mathbf{y},t)\in\Omega_q\times\mathbb{R}} |\varphi(\mathbf{y}, t)| \left\{ \int_{\bar{A}_{1,\delta}} LK\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}, \frac{z - t}{g}\right) dt\, \omega_q(d\mathbf{y}) \right. \\
&\quad \left. + \int_{\bar{A}_{2,\delta}} LK\left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}, \frac{z - t}{g}\right) dt\, \omega_q(d\mathbf{y}) \right\} \\
&\leq 2 \sup_{(\mathbf{y},t)\in\Omega_q\times\mathbb{R}} |\varphi(\mathbf{y}, t)| \left\{ c_{h,q}(L)\omega_{q-1} \int_{-1}^{1-\frac{\delta^2}{2}} L\left(\frac{1 - u}{h^2}\right) (1 - u^2)^{\frac{q}{2} - 1}\, du \right.
\end{aligned}$$

$$
\begin{aligned}
&+ 2\int_{\delta g^{-1}}^{\infty} K(s)\, ds\Bigg\} \\
&\leq 2 \sup_{(\mathbf{y},t)\in\Omega_q\times\mathbb{R}} |\varphi(\mathbf{y},t)| \left\{ c_{h,q}(L)\omega_{q-1}\int_{-1}^{1}(1-u^2)^{\frac{q}{2}-1}\, du \times \sup_{r\geq\delta^2/(2h^2)} L(r) r^{\frac{q}{2}} r^{-\frac{q}{2}} \right. \\
&\left. + 2\int_{\delta g^{-1}}^{\infty} K(s)\, ds \right\} \\
&\leq \mathcal{O}(1)\left\{ \lambda_{h,q}(L)^{-1}\omega_{q-1}2^{-\frac{q}{2}}\delta^{-q}\int_{-1}^{1}(1-u^2)^{\frac{q}{2}-1}\, du \times \sup_{r\geq\delta^2/(2h^2)} L(r) r^{\frac{q}{2}} + o(1) \right\} \\
&= \mathcal{O}(1)\left(\mathcal{O}(1)\, o(1) + o(1)\right) \\
&= o(1),
\end{aligned}
$$

by relation (6.1), the fact $\int_{-1}^{1}(1-u^2)^{\frac{q}{2}-1}\, du < \infty$ for all $q \geq 1$ and because by assumption **A2**, $\lambda_{q+2}(L) < \infty$, which implies that $\lim_{r\to\infty} L(r) r^{\frac{q}{2}} = 0$.

Then, $D_n \to 0$ as $n \to \infty$ and this holds regardless the point $(\mathbf{x}, z)$, since φ is uniformly continuous, so (A.14) is satisfied and $LK_{h,g}\varphi(\mathbf{x}, z)$ converges to $\varphi(\mathbf{x}, z)$ uniformly in $\Omega_q \times \mathbb{R}$. □

Lemma A.11. *Under conditions **A1**–**A3**, the bias and the variance for the directional-linear estimator in a point $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$ is given by*

$$
\begin{aligned}
\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x}, z)\right] &= f(\mathbf{x}, z) + \frac{b_q(L)}{q}\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z)\right] h^2 + \frac{1}{2}\mu_2(K)\mathcal{H}_z f(\mathbf{x}, z) g^2 + o\left(h^2 + g^2\right), \\
\mathbb{V}\mathrm{ar}\left[\hat{f}_{h,g}(\mathbf{x}, z)\right] &= \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g} f(\mathbf{x}, z) + o\left((nh^q g)^{-1}\right),
\end{aligned}
$$

where the remainder orders are uniform.

Proof of Lemma A.11. The asymptotic expressions of the bias and the variance are given in García-Portugués et al. (2013b). Recalling the extension of f in condition **A1**, the partial derivative of f for the direction $\mathbf{x}$ and evaluated at $(\mathbf{x}, z)$, that is $\mathbf{x}^T\boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}, z)$, is null:

$$
\mathbf{x}^T\boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}, z) = \lim_{h\to 0}\frac{f((1+h)\mathbf{x}, z) - f(\mathbf{x}, z)}{h} = \lim_{h\to 0}\frac{f(\mathbf{x}, z) - f(\mathbf{x}, z)}{h} = 0.
$$

Using this fact, it also follows that $\mathbf{x}^T\boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z)\mathbf{x} = 0$, since

$$
\mathbf{x}^T\left(\frac{\partial}{\partial\mathbf{x}}\mathbf{x}^T\boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}, z)\right) = \mathbf{x}^T\left(\boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}, z) + \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z)\mathbf{x}\right) = 0.
$$

Therefore, the operator $\boldsymbol{\Psi}_{\mathbf{x}}(f, \mathbf{x}, z)$ appearing in the bias expansion given in García-Portugués et al. (2013b) can be written in the simplified form

$$
\boldsymbol{\Psi}_{\mathbf{x}}(f, \mathbf{x}, z) = -\mathbf{x}^T\boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}, z) + \frac{1}{q}\left(\nabla^2 f(\mathbf{x}, z) - \mathbf{x}^T\boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z)\mathbf{x}\right) = \frac{1}{q}\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z)\right],
$$

because $\nabla^2 f(\mathbf{x}, z)$ represents the directional Laplacian of f (the trace of $\boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z)$).

The uniformity of the orders, not considered in the above paper, can be obtained by using the extra-smoothness assumption **A1** and the integral form of the remainder in the Taylor's theorem on f:

$$f(\mathbf{y}+\boldsymbol{\alpha}) - f(\mathbf{y}) = \boldsymbol{\alpha}^T \boldsymbol{\nabla} f(\mathbf{y}) + \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{\mathcal{H}} f(\mathbf{y})\boldsymbol{\alpha} + R,$$

with $\mathbf{y} \equiv (\mathbf{x}, z)$, $\boldsymbol{\alpha} \in \Omega_q \times \mathbb{R}$ and where the remainder has the exact form

$$R = \int_0^1 \frac{(1-t)^2}{2} \sum_{i,j,k=1}^{q+1} \frac{\partial^3}{\partial x_i \partial x_j \partial x_k} f(\mathbf{x}+t\boldsymbol{\alpha})\alpha_i\alpha_j\alpha_k \, dt \leq \frac{1}{6} M \sum_{i,j,k=1}^{q+1} \alpha_i\alpha_j\alpha_k = o\left(\boldsymbol{\alpha}^T\boldsymbol{\alpha}\right),$$

where M is the bound of the third derivatives of f and in the last equality it is used the second point of Lemma A.12. Then the remainder does not depend on the point $\mathbf{y} \equiv (\mathbf{x}, z)$ and following the proofs of García-Portugués et al. (2013b) the convergence of the bias and variance is uniform on $\Omega_q \times \mathbb{R}$. □

Lemma A.12. *Let a_n, b_n and c_n sequences of positive real numbers. Then:*

i. If $a_n, b_n \to 0$, then $a_n b_n = o(a_n + b_n)$.

ii. If $a_n, b_n, c_n \to 0$, then $a_n b_n c_n = o(a_n^2 + b_n^2 + c_n^2)$.

iii. $a_n^i b_n^j = \mathcal{O}(a_n^k + b_n^k)$, for any integers $i, j \geq 0$ such that $i + j = k$.

iv. $(a_n + b_n)^k = \mathcal{O}(a_n^k + b_n^k)$, for any integer $k \geq 1$.

Proof of Lemma A.12. The first statement follows immediately from the definition of $o(\cdot)$,

$$a_n b_n = o(a_n + b_n) :\iff \lim_{n\to\infty} \frac{a_n b_n}{a_n + b_n} = \lim_{n\to\infty} \frac{1}{\frac{1}{b_n} + \frac{1}{a_n}} = \frac{1}{\infty} = 0.$$

For the second, suppose that, when $n \to \infty$, $a_n = \max(a_n, b_n, c_n)$ to fix notation. Then

$$\lim_{n\to\infty} \frac{a_n b_n c_n}{a_n^2 + b_n^2 + c_n^2} \leq \lim_{n\to\infty} \frac{a_n^3}{a_n^2 + b_n^2 + c_n^2} = \lim_{n\to\infty} \frac{1}{\frac{1}{a_n} + \frac{b_n^2}{a_n^3} + \frac{c_n^2}{a_n^3}} = \frac{1}{\infty} = 0.$$

Let C be a positive constant. The third statement follows from the definition of $\mathcal{O}(\cdot)$,

$$\lim_{n\to\infty} \frac{a_n^i b_n^j}{a_n^k + b_n^k} = \lim_{n\to\infty} \frac{1}{\left(\frac{a_n}{b_n}\right)^j + \left(\frac{b_n}{a_n}\right)^i} = \begin{cases} \frac{1}{0+\infty}, & a_n = o(b_n), \\ \frac{1}{\infty+0}, & b_n = o(a_n), \\ \frac{1}{C^j + C^{-i}}, & a_n \sim C b_n. \end{cases}$$

Then the limit is bounded and $a_n^i b_n^j = \mathcal{O}(a_n^k + b_n^k)$. The last statement arises as a consequence of this result and the Newton binomial:

$$(a_n + b_n)^k = \sum_{i=0}^{k} \binom{k}{i} a_n^{k-i} b_n^i = \sum_{i=0}^{k} \binom{k}{i} \mathcal{O}\left(a_n^k + b_n^k\right) = \mathcal{O}\left(a_n^k + b_n^k\right).$$

□

A.2 Further results for the independence test

A.2.1 Closed expressions

Consider K and L a normal and a von Mises kernel, respectively. In this case $R(K) = \left(2\pi^{\frac{1}{2}}\right)^{-1}$, $\lambda_q(L) = (2\pi)^{\frac{q}{2}}$ and $\lambda_q(L^2)\lambda_q(L)^{-2} = \left(2\pi^{\frac{1}{2}}\right)^{-q}$. Furthermore, it is possible to compute exactly the form of the contributions of these two kernels to the asymptotic variance, resulting:

$$\gamma_q\lambda_q(L)^{-4}\int_0^\infty r^{\frac{q}{2}-1}\left\{\int_0^\infty \rho^{\frac{q}{2}-1}L(\rho)\varphi_q(r,\rho)\,d\rho\right\}^2 dr = (8\pi)^{-\frac{q}{2}},$$
$$\int_{\mathbb{R}}\left\{\int_{\mathbb{R}} K(u)K(u+v)\,du\right\}^2 dv = (8\pi)^{-\frac{1}{2}}.$$

Corollary A.1. *If $L(r) = e^{-r}$ and K is a normal density, then the asymptotic bias and variance in Theorem 6.2 are*

$$A_n = \frac{1}{2^{q+1}\pi^{\frac{q+1}{2}}nh^q g} - \frac{R(f_Z)}{2^q\pi^{\frac{q}{2}}nh^q} - \frac{R(f_{\mathbf{X}})}{2\pi^{\frac{1}{2}}ng}, \quad \sigma_I^2 = (8\pi)^{-\frac{q+1}{2}}R(f_{\mathbf{X}})R(f_Z).$$

In addition, if $f_{\mathbf{X}} = f_{\mathrm{vM}}(\cdot;\boldsymbol{\mu},\kappa)$ and f_Z is the density of a $\mathcal{N}(m,\sigma^2)$, then $R(f_{\mathbf{X}}) = \left(2\pi^{\frac{q+1}{2}}\right)^{-1}\kappa^{\frac{q-1}{2}}\mathcal{I}_{\frac{q-1}{2}}(2\kappa)\mathcal{I}_{\frac{q-1}{2}}(\kappa)^{-2}$ and $R(f_Z) = \left(2\pi^{\frac{1}{2}}\sigma\right)^{-1}$.

Proof of Corollary A.1. The expressions for $R(K)$, $R(f_Z)$ and $\int_{\mathbb{R}}\left\{\int_{\mathbb{R}} K(u)K(u+v)\,du\right\}^2 dv = (8\pi)^{-\frac{1}{2}}$ follow easily from the convolution properties of normal densities. The expressions for $\lambda_q(L)$ and $\lambda_q(L^2)$ can be derived from the definition of the Gamma function. Similarly,

$$R(f_{\mathbf{X}}) = C_q(\kappa)^2\int_{\Omega_q} e^{2\kappa\mathbf{x}^T\boldsymbol{\mu}}\,\omega_q(d\mathbf{x}) = \frac{C_q(\kappa)^2}{C_q(2\kappa)} = \frac{\kappa^{\frac{q-1}{2}}\mathcal{I}_{\frac{q-1}{2}}(2\kappa)}{2\pi^{\frac{q+1}{2}}\mathcal{I}_{\frac{q-1}{2}}(\kappa)^2},$$

$$\gamma_q^{-1}\lambda_q(L)^4 = \begin{cases} 2^{-\frac{5}{4}}\pi^2, & q=1,\\ 2^{\frac{q}{2}}\pi^{\frac{q}{2}+1}\Gamma\left(\frac{q}{2}\right)\Gamma\left(\frac{q-1}{2}\right)^2, & q>1. \end{cases} \tag{A.1}$$

For $q = 1$ the contribution of the directional kernel to the asymptotic variance can be computed using (A.1) and

$$\int_0^\infty \rho^{-\frac{1}{2}}e^{-2\left(\rho\pm\sqrt{r\rho}\right)}\,d\rho = \sqrt{2\pi}e^{\frac{r}{2}}\left(1-\Phi\left(\mp\sqrt{r}\right)\right),$$

where Φ is the cumulative distribution function of a $\mathcal{N}(0,1)$. Then:

$$\begin{aligned}
&\gamma_1\lambda_1(L)^{-4}\int_0^\infty r^{-\frac{1}{2}}\left\{\int_0^\infty \rho^{-\frac{1}{2}}L(\rho)\varphi_1(r,\rho)\,d\rho\right\}^2 dr\\
&\quad= \gamma_1\lambda_1(L)^{-4}\int_0^\infty r^{-\frac{1}{2}}e^{-2r}\left\{\int_0^\infty \rho^{-\frac{1}{2}}e^{-2\rho-2(r\rho)^{\frac{1}{2}}}\,d\rho + \int_0^\infty \rho^{-\frac{1}{2}}e^{-2\rho+2(r\rho)^{\frac{1}{2}}}\,d\rho\right\}^2 dr\\
&\quad= 2^{-\frac{1}{2}}(2\pi)^{-1}\int_0^\infty r^{-\frac{1}{2}}e^{-r}\,dr\\
&\quad= (8\pi)^{-\frac{1}{2}}.
\end{aligned}$$

For $q > 1$, the integral with respect to θ is computed from the definition of the modified Bessel function and the integral with respect to ρ is

$$\int_0^\infty \rho^{\frac{q}{4}-\frac{1}{2}}e^{-2\rho}\mathcal{I}_{\frac{q}{2}-1}\left(2\sqrt{r\rho}\right)\,d\rho = 2^{-\frac{q}{2}}r^{\frac{q}{4}-\frac{1}{2}}e^{\frac{r}{2}}.$$

Using these two facts, it results:

$$\begin{aligned}
&\gamma_q \lambda_q(L)^{-4} \int_0^\infty r^{\frac{q}{2}-1} \left\{ \int_0^\infty \rho^{\frac{q}{2}-1} L(\rho) \varphi_q(r,\rho)\, d\rho \right\}^2 dr \\
&\quad = 2^{-\frac{q}{2}} \pi^{-\left(\frac{q}{2}+1\right)} \Gamma\left(\frac{q}{2}\right)^{-1} \Gamma\left(\frac{q-1}{2}\right)^{-2} \\
&\qquad \times \int_0^\infty r^{\frac{q}{2}-1} \left\{ \int_0^\infty \rho^{\frac{q}{2}-1} e^{-(r+2\rho)} \left[\pi^{\frac{1}{2}} \Gamma\left(\frac{q-1}{2}\right) (r\rho)^{-\frac{q-2}{4}} \mathcal{I}_{\frac{q}{2}-1}\left(2\sqrt{r\rho}\right) \right] d\rho \right\}^2 dr \\
&\quad = 2^{-\frac{q}{2}} \pi^{-\frac{q}{2}} \Gamma\left(\frac{q}{2}\right)^{-1} \int_0^\infty r^{\frac{q}{2}-1} e^{-2r} \left\{ 2^{-\frac{q}{2}} e^{\frac{r}{2}} \right\}^2 dr \\
&\quad = (8\pi)^{-\frac{q}{2}}.
\end{aligned}$$

□

A.2.2 Extension to the directional-directional case

Under the directional-directional analogue of condition **A4**, that is, $h_{1,n}^{q_1} h_{2,n}^{-q_2} \to c$, with $0 < c < \infty$, the directional-linear independence test can be directly adapted to this setting, considering the following test statistic:

$$T_n = \int_{\Omega_{q_1} \times \Omega_{q_2}} \left(\hat{f}_{(\mathbf{X},\mathbf{Y});h_1,h_2}(\mathbf{x},\mathbf{y}) - \hat{f}_{\mathbf{X};h_1}(\mathbf{x}) \hat{f}_{\mathbf{Y};h_2}(\mathbf{y}) \right)^2 \omega_{q_2}(d\mathbf{y})\, \omega_{q_1}(d\mathbf{x}).$$

Corollary A.2 (Directional-directional independence test)**.** *Under the directional-directional analogues of conditions **A1**–**A4** and under the null hypothesis of independence,*

$$n(h_1^{q_1} h_2^{q_2})^{\frac{1}{2}} (T_n - A_n) \xrightarrow{d} \mathcal{N}(0, 2\sigma_I^2),$$

where

$$\begin{aligned}
A_n = &\frac{\lambda_{q_1}(L_1^2)\lambda_{q_1}(L_1)^{-2}\lambda_{q_2}(L_2^2)\lambda_{q_2}(L_2)^{-2}}{n h_1^{q_1} h_2^{q_2}} \\
&- \frac{\lambda_{q_1}(L_1^2)\lambda_{q_1}(L_1)^{-2} R(f_{\mathbf{Y}})}{n h_1^{q_1}} - \frac{\lambda_{q_2}(L_2^2)\lambda_{q_2}(L_2)^{-2} R(f_{\mathbf{X}})}{n h_2^{q_2}},
\end{aligned}$$

and σ_I^2 is defined as σ^2 in Corollary 6.2 but with $R(f) = R(f_{\mathbf{X}})R(f_{\mathbf{Y}})$. Further, if L_1 and L_2 are the von Mises kernel,

$$A_n = \frac{1}{2^{q_1+q_2} \pi^{\frac{q_1+q_2}{2}} n h_1^{q_1} h_2^{q_2}} - \frac{R(f_{\mathbf{Y}})}{2^{q_1} \pi^{\frac{q_1}{2}} n h_1^{q_1}} - \frac{R(f_{\mathbf{X}})}{2^{q_2} \pi^{\frac{q_2}{2}} n h_2^{q_2}}$$

and $\sigma_I^2 = (8\pi)^{-\frac{q_1+q_2}{2}} R(f_{\mathbf{X}})R(f_{\mathbf{Y}})$. If $f_{\mathbf{X}}$ and $f_{\mathbf{Y}}$ are von Mises densities, $R(f_{\mathbf{X}})$ and $R(f_{\mathbf{Y}})$ are given as in Corollary A.1.

Proof of Corollary A.2. The proof follows from adapting the proofs of Theorem 6.2 and Corollary A.1 to the directional-directional situation. □

A.2.3 Some numerical experiments

The purpose of this subsection is to provide some numerical experiments to illustrate the degree of misfit between the true distribution of the standardized statistic (approximated by Monte Carlo) and its asymptotic distribution, for increasing sample sizes.

For simplicity, independence will be assessed in a circular-linear framework ($q = 1$), with a $\mathrm{vM}((0,1),1)$ for the circular variable and a $\mathcal{N}(0,1)$ for the linear one. Kernel density estimation is done using von Mises and normal kernels, as in Corollary A.1. Sample sizes considered are $n = 5^j \times 10^k$, $j = 0, 1$, $k = 3, 5$ (see supplementary material for $k = 1, 2, 4$). The sequence of bandwidths is taken as $h_n = g_n = 2n^{-\frac{1}{3}}$, as a compromise between fast convergence and numerical problems avoidance. Figure A.1 presents the histogram of 1000 values from $(nh_n^q g_n)^{\frac{1}{2}}(T_n - A_n)$ for different sample sizes, jointly with the p-values of the Kolmogorov-Smirnov test for the distribution $\mathcal{N}(0, 2\sigma_I^2)$ and of the Shapiro-Wilk test for normality. Both tests are significant, until a very large sample size (close to $500,000$ data) is reached.

It should be noted that, in practical problems, the use of the asymptotic distribution does not seem feasible, and a resampling mechanism for the calibration of the test is required. This issue is addressed in García-Portugués et al. (2014), considering a permutation approach. The reader is referred to the aforementioned paper for the details concerning the practical application.

A.3 Extended simulation study

Some technical details concerning the simulation study and further results are provided in this section. First, the simulated models considered will be described. For constructing the test statistic, parametric estimators as well as simulation methods are required. Different Maximum Likelihood Estimators (MLE) and simulation approaches have been considered, playing copulas a remarkable role in both problems (see Nelsen (2006) for a comprehensive review). Some details on the construction of alternative models and bandwidth choice will be also given, jointly with extended results showing the performance of the tests (for circular-linear and circular-circular cases) for different significance levels.

A.3.1 Parametric models

Two collections of Circular-Linear (CL) and Circular-Circular (CC) parametric scenarios have been considered. The corresponding density contours can be seen in Figures 6.1 and 6.2 in the paper. For the circular-linear case, the first five models (CL1–CL5) contain parametric densities with independent components and different kinds of marginals, for which estimation and simulation are easily accomplished. The models are based on von Mises, wrapped Cauchy, wrapped normal, normal, log-normal, gamma and mixtures of these densities. Models CL6–CL7 represent two parametric choices of the model in Mardia and Sutton (1978) for cylindrical variables, which is constructed conditioning a normal density on a von Mises one. Models CL8–CL9 include two parametric densities of the semiparametric circular-linear model given in Theorem 5 of Johnson and Wehrly (1978). This family is indexed by a circular density g that defines the underlying circular-linear copula density, allowing for flexibility both in the specification of the link density and the marginals. CL10 is the model given in Theorem 1 of Johnson and Wehrly (1978), which considers an exponential density conditioned on a von Mises. CL11 is

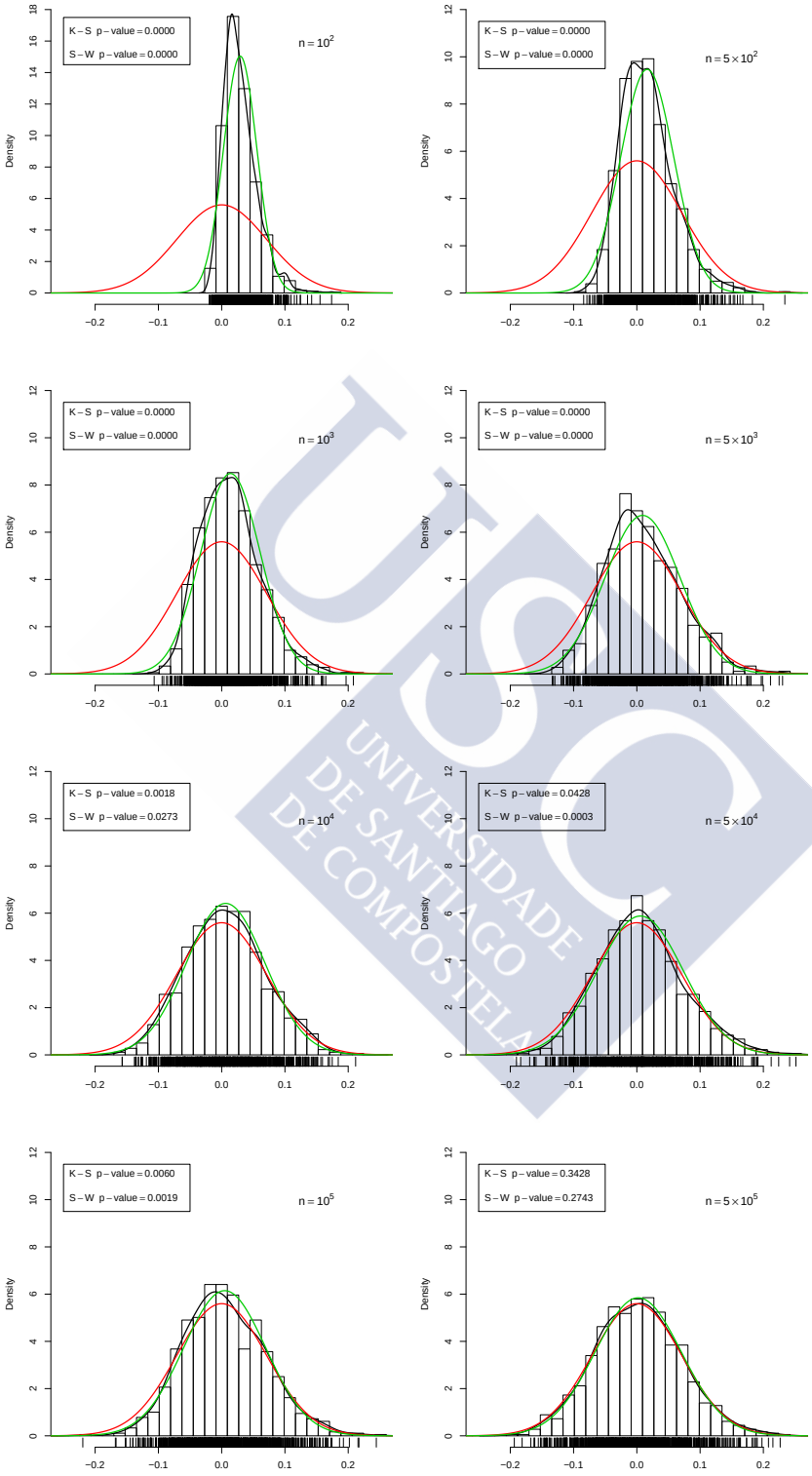


Figure A.1: Comparison of the asymptotic and empirical distributions of $(nh_n^q g_n)^{\frac{1}{2}}(T_n - A_n)$ for sample sizes $n = 5^j \times 10^k$, $j = 0, 1$, $k = 2, 3, 4, 5$. Black curves represent a kernel estimation from 1000 simulations, green curves represent a normal fit to the unknown density and red curves represent the theoretical asymptotic distribution.

constructed considering the QS copula density of García-Portugués et al. (2013a) and cardioid and log-normal marginals. Finally, CL12 is an adaptation of the circular-circular copula density of Kato (2009) to the circular-linear scenario, using an identity matrix in the joint structure and von Mises and log-normal marginals.

The first models (CC1–CC5) of the circular-circular case include also parametric densities with independent components and different kinds of marginals (von Mises, wrapped Cauchy, cardioid and mixtures of them). Models CC6–CC7 represent two parametric choices of the sine model given by Singh et al. (2002). This model introduces elliptical contours for bivariate circular densities and also allows for certain multimodality. Models CC8–CC9 are two densities of the semiparametric models of Wehrly and Johnson (1979), which are based on the previous work of Johnson and Wehrly (1978) and comprise as a particular case the bivariate von Mises model of Shieh and Johnson (2005). Models CC10–CC11 are two parametric choices of the wrapped normal torus density given in Johnson and Wehrly (1977), a natural extension of the circular wrapped normal to the circular-circular setting. Finally, CC12 employs the copula density of Kato (2009) with von Mises marginals.

Density name	Expression
Normal	$f_{\mathcal{N}}(z; m, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(z-m)^2}{2\sigma^2}\right\}$
Log-normal	$f_{\mathcal{LN}}(z; m, \sigma) = \frac{1}{z\sqrt{2\pi}\sigma} \exp\left\{-\frac{(\log z - m)^2}{2\sigma^2}\right\} \mathbf{1}_{(0,\infty)}(z)$
Gamma	$f_{\Gamma}(z; a, p) = \frac{a^p}{\Gamma(p)} z^{p-1} e^{-az} \mathbf{1}_{(0,\infty)}(z)$
Bivariate normal	$f_{\mathcal{N}}(z_1, z_2; m_1, m_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$
	$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(z_1-m_1)^2}{\sigma_1^2} + \frac{(z_2-m_2)^2}{\sigma_2^2} - \frac{2\rho(z_1-m_1)(z_2-m_2)}{\sigma_1\sigma_2}\right)\right\}$
Von Mises	$f_{\mathrm{vM}}(\theta; \mu, \kappa) = \frac{1}{2\pi\mathcal{I}_0(\kappa)} \exp\{\kappa\cos(\theta-\mu)\}$
Cardioid	$f_{\mathrm{Ca}}(\theta; \mu, \rho) = \frac{1}{2\pi}(1 + 2\rho\cos(\theta-\mu))$
Wrapped Cauchy	$f_{\mathrm{WC}}(\theta; m, \sigma) = \frac{1-\rho^2}{2\pi(1+\rho^2-2\rho\cos(\theta-\mu))}$
Wrapped Normal	$f_{\mathrm{WN}}(\theta; \mu, \rho) = \sum_{p=-\infty}^{\infty} f_{\mathcal{N}}(\theta + 2\pi p; m, \sigma)$

Table A.1: Notation for the densities described in Tables A.2 and A.3.

The notation and density expressions used for the construction of the parametric models are collected in Table A.1, whereas Tables A.2 and A.3 show the explicit expressions and parameters for the circular-linear and circular-circular models displayed in Figure 6.1. Most of the circular densities considered in the simulation study are purely circular (and hence not directional) and their circular formulation has been used in order to simplify expressions. The directional notation can be obtained taking into account that $\mathbf{x} = (\cos\theta, \sin\theta)$, $\mathbf{y} = (\cos\psi, \sin\psi)$ and $\boldsymbol{\mu} = (\cos\mu, \sin\mu)$. The distribution function of a circular variable with density f, with $\theta \in [0, 2\pi)$ will be denoted by $F(\theta) = \int_0^\theta f(\varphi)\, d\varphi$.

A.3.2 Estimation

In the scenarios considered, for most of the marginal densities, MLE are available through specific libraries of `R`. For the normal and log-normal densities closed expressions are used and for the gamma density the `fitdistr` function of the `MASS` (Venables and Ripley, 2002) library is

employed. The estimation of the von Mises parameters is done exactly for the mean and numerically for the concentration parameter, whereas for the wrapped Cauchy and wrapped normal densities the numerical routines of the `circular` (Agostinelli and Lund, 2013) package are used. The MLE for the cardioid density are obtained by numerical optimization. Finally, the fitting of mixtures of normals and von Mises was carried out using the Expectation-Maximization algorithms given in packages `nor1mix` (Mächler, 2013) and `movMF` (Hornik and Grün, 2012), respectively.

The fitting of the independent models CL1–CL5 and CC1–CC5 is easily accomplished by marginal fitting of each component. For models CL6–CL7, the closed expressions for the MLE given in Mardia and Sutton (1978) are used. For models CL8–CL9, CL11–CL12, CC8–CC9 and CC12 a two-step Maximum Likelihood (ML) estimation procedure based on the copula density decomposition is used: first, the marginals are fitted by ML and then the copula is estimated by ML using the pseudo-observations computed from the fitted marginals. This procedure is described in more detail in Section 3 of García-Portugués et al. (2013a). In models CL8–CL9 and CC8–CC9 the MLE for the copula are obtained by estimating univariate von Mises or mixtures of von Mises, whereas numerical optimization is required for the copula estimation. For models CC6–CC7 and CC10–CC11, MLE can be also carried out by numerical optimization. Finally, MLE for model CL10 in Johnson and Wehrly (1978) were obtained analytically: given the circular-linear sample $\{(\Theta_i, Z_i)\}_{i=1}^n$,

$$\hat{\lambda} = \frac{\bar{Z}}{(\bar{Z})^2 - (\bar{Z}_c)^2}, \quad \hat{\kappa} = \sqrt{\hat{\lambda}^2 - \hat{\lambda}\bar{Z}^{-1}} \quad \text{and} \quad \sum_{i=1}^n Z_i \sin(\Theta_i - \hat{\mu}) = 0,$$

with $\bar{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$ and $\bar{Z}_c = \frac{1}{n}\sum_{i=1}^n Z_i \cos(\Theta_i - \hat{\mu})$.

A.3.3 Simulation

Simulating from the linear marginals is easily accomplished by the built-in functions in `R`. The simulation of the wrapped Cauchy and wrapped normal is done with the `circular` library, the von Mises is sampled implementing the algorithm described in Wood (1994) and the cardioid by the inversion method, whose equation is solved numerically. Sampling from the independence models is straightforward. Conditioning on the circular variable, it is easy to sample from models CL6–CL7 (sample the circular observation from a von Mises and then the linear from a normal with mean depending on the circular), CL10 (von Mises marginal and exponential with varying rate) and CC6–CC7 (using the properties detailed in Singh et al. (2002) and the inversion method). Simulation in CC10–CC11 is straightforward: sample from a bivariate normal and then wrap around $[0, 2\pi)$ by applying a modulus of 2π. Finally, simulation in two steps using copulas was required for models CL8–CL9, CL12, CC8–CC9 and CC12, where first a pair of uniform random variables (U, V) is sampled from the copula of the density and then the inversion method is applied marginally. See Section 3.1 of García-Portugués et al. (2013a) for more details. The simulation of the pair (U, V) was done by the conditional and inversion methods and, specifically, for the models based on the densities given by Johnson and Wehrly (1978) and Wehrly and Johnson (1979), a transformation method was obtained. It is summarized in the following algorithm.

Algorithm A.1. *Let g be a circular density. A pair (U, V) of uniform variables with joint density $c_g(u, v) = 2\pi g(2\pi(u \pm v))$ is obtained as follows:*

Model	Density	Parameters	Description
CL1	$f_{\text{vM}}(\theta;\mu,\kappa) \times f_{\mathcal{N}}(z;m,\sigma)$	$\mu = \frac{3\pi}{2}$, $\kappa = 2$, $m = 0$, $\sigma = 1$	Independent von Mises and normal
CL2	$f_{\text{WC}}(\theta;\mu,\rho) \times f_{\mathcal{LN}}(z;m,\sigma)$	$\mu = \frac{3\pi}{2}$, $\rho = \sigma = \frac{3}{4}$, $m = \frac{1}{2}$	Independent wrapped Cauchy and log-normal
CL3	$\big(p_1 f_{\text{vM}}(\theta;\mu_1,\kappa_1) + p_2 f_{\text{vM}}(\theta;\mu_2,\kappa_2)\big) \times f_{\Gamma}(z;a,p)$	$\mu_1 = \frac{\pi}{4}$, $\mu_2 = \frac{5\pi}{4}$, $\kappa_1 = \kappa_2 = 2$, $p_1 = p_2 = \frac{1}{2}$, $a = \frac{1}{3}$, $p = 3$	Independent mixture of von Mises and gamma
CL4	$f_{\text{WN}}(\theta;m_1,\sigma_1)$ $\times (p_1 f_{\mathcal{N}}(z;m_2,\sigma_2) + p_2 f_{\mathcal{N}}(z;m_3,\sigma_3))$	$m_1 = \frac{3\pi}{2}$, $\sigma_1 = \sigma_3 = 1$, $m_2 = 0$, $\sigma_2 = \frac{1}{4}$, $m_3 = 2$, $p_1 = p_2 = \frac{1}{2}$	Independent wrapped normal and mixture of normals
CL5	$\big(p_1 f_{\text{vM}}(\theta;\mu_1,\kappa_1) + p_2 f_{\text{vM}}(\theta;\mu_2,\kappa_2)\big)$ $\times\big(p_3 f_{\mathcal{N}}(z;m_1,\sigma_1) + p_4 f_{\mathcal{N}}(z;m_2,\sigma_2)\big)$	$\mu_1 = \frac{5\pi}{4}$, $\mu_2 = \frac{7\pi}{4}$, $\kappa_1 = 10$, $\kappa_2 = 3$, $m_1 = -1$, $m_2 = 2$, $\sigma_1 = 1$, $\sigma_2 = p_1 = p_2 = \frac{1}{2}$, $p_3 = \frac{3}{4}$, $p_4 = \frac{1}{4}$	Independent mixture of von Mises and of normals
CL6	$f_{\text{vM}}(\theta;\mu,\kappa) \times f_{\mathcal{N}}\big(z;m(\theta),\sigma(1-\rho_1-\rho_2)\big)$, with $m(\theta) = m + \sigma\kappa^{\frac{1}{2}}\{\rho_1(\cos(\theta)-\cos(\mu))$ $+\rho_2(\sin(\theta)-\sin(\mu))\}$	$\mu = \frac{3\pi}{2}$, $\kappa = 1$, $m = 0$, $\rho_1 = \rho_2 = \sigma = \frac{1}{2}$	See equation (1.1) of Mardia and Sutton (1978)
CL7		$\mu = \frac{3\pi}{2}$, $\kappa = 5$, $m = 0$, $\rho_1 = \frac{1}{2}$, $\rho_2 = -\frac{3}{4}$, $\sigma = \frac{3}{2}$	
CL8	$f_{\text{vM}}\left(2\pi\left(\frac{\theta}{2\pi} + F_{\mathcal{N}}(z;m,\sigma)\right);\mu_g,\kappa_g\right) \times f_{\mathcal{N}}(z;m,\sigma)$	$m = 0$, $\sigma = 1$, $\mu_g = \frac{5\pi}{4}$, $\kappa_g = \frac{3}{2}$	See Theorem 5 of Johnson and Wehrly (1978) considering a von Mises and a mixture of von Mises as the link functions
CL9	$g\big(2\pi\big(\frac{\theta}{2\pi} - F_{\mathcal{N}}(z;m,\sigma)\big)\big) \times f_{\mathcal{N}}(z;m,\sigma)$, with $g(\theta) = p_{g_1} f_{\text{vM}}\left(\theta;\mu_{g_1},\kappa_{g_1}\right) + p_{g_2} f_{\text{vM}}\left(\theta;\mu_{g_2},\kappa_{g_2}\right)$	$m = 0$, $\sigma = p_{g_1} = p_{g_2} = \frac{1}{2}$, $\mu_{g_1} = \frac{\pi}{4}$, $\kappa_{g_1} = \kappa_{g_2} = 3$, $\mu_{g_2} = \frac{5\pi}{4}$	
CL10	$\frac{(\lambda^2-\kappa^2)^{\frac{1}{2}}}{2\pi} \exp\{-\lambda z + \kappa z\cos(\theta-\mu)\}$	$\mu = \frac{3\pi}{2}$, $\kappa = 2$, $\lambda = 3$	See Theorem 1 of Johnson and Wehrly (1978)
CL11	$\big\{1 + 2\pi\alpha\cos(2\pi F_{\text{Ca}}(\theta;\mu,\rho))(1 - 2F_{\mathcal{N}}(z;m,\sigma))\big\}$ $\times f_{\text{Ca}}(\theta;\mu,\rho) f_{\mathcal{N}}(z;m,\sigma)$	$\mu = \frac{3\pi}{2}$, $\rho = \frac{9}{20}$, $m = 1$, $\sigma = \frac{1}{2}$, $\alpha = \frac{1}{2\pi}$	See equation (7) of García-Portugués et al. (2013a)
CL12	$\left\{4\pi^2\left(1 - 2\rho F_{\text{vM}}(\theta;\mu,\kappa)F_{\mathcal{LN}}(z;m,\sigma) + \rho^2\right)\right\}^{-1}$ $\times(1-\rho^2) f_{\text{vM}}(\theta;\mu,\kappa) f_{\mathcal{LN}}(z;m,\sigma)$	$\mu = \frac{3\pi}{2}$, $\kappa = 1$, $m = \frac{1}{2}$, $\sigma = \rho = \frac{3}{4}$	See Section 4.1 in Kato (2009)

Table A.2: Circular-linear models.

Model	Density	Parameters	Description
CC1	$\frac{1}{2\pi} \times f_{\text{vM}}(\psi; \mu, \kappa)$	$\mu = 0$, $\kappa = 2$	Independent uniform and von Mises
CC2	$f_{\text{vM}}(\theta; \mu_1, \kappa_1) \times f_{\text{vM}}(\psi; \mu_2, \kappa_2)$	$\mu_1 = \frac{3\pi}{2}$, $\kappa_1 = 1$, $\mu_2 = \pi$, $\kappa_2 = 3$	Independent von Mises and von Mises
CC3	$f_{\text{vM}}(\theta; \mu_1, \kappa) \times f_{\text{WC}}(\psi; \mu_2, \rho)$	$\mu_1 = \frac{3\pi}{2}$, $\kappa = 2$, $\mu_2 = \frac{\pi}{4}$, $\rho = \frac{7}{10}$	Independent von Mises and wrapped Cauchy
CC4	$\big(p_1 f_{\text{vM}}(\theta; \mu_1, \kappa_1) + p_2 f_{\text{vM}}(\theta; \mu_2, \kappa_2)\big) \times f_{\text{Ca}}(\psi; \mu_3, \rho)$	$\mu_1 = 0$, $\kappa_1 = \kappa_2 = 10$, $\mu_2 = \frac{3\pi}{2}$, $\mu_3 = 0$, $\rho = \frac{1}{4}$, $p_1 = p_2 = \frac{1}{2}$	Independent mixture von Mises and cardioid
CC5	$\big(p_1 f_{\text{vM}}(\theta; \mu_1, \kappa_1) + p_2 f_{\text{vM}}(\theta; \mu_2, \kappa_2)\big)$ $\times \big(p_3 f_{\text{vM}}(\psi; \mu_3, \kappa_3) + p_4 f_{\text{vM}}(\psi; \mu_4, \kappa_4)\big)$	$\mu_1 = 0$, $\kappa_1 = \kappa_2 = 3$, $\mu_2 = \frac{3\pi}{2}$, $\mu_3 = \frac{\pi}{4}$, $\kappa_3 = \kappa_4 = 5$, $\mu_4 = \frac{7\pi}{4}$, $p_1 = p_2 = p_3 = p_4 = \frac{1}{2}$	Independent mixture of von Mises and of von Mises
CC6	$C \exp\big\{\kappa_1 \cos(\theta - \mu_1) + \kappa_2 \cos(\psi - \mu_2)$ $+\lambda \sin(\theta - \mu_1) \sin(\psi - \mu_2)\big\}$	$\mu_1 = \frac{7\pi}{8}$, $\kappa_1 = \frac{1}{2}$, $\mu_2 = 0$, $\kappa_2 = 1$, $\lambda = -3$	See equation (1.1) of Singh et al. (2002)
CC7		$\mu_1 = \mu_2 = 0$, $\kappa_1 = 5$, $\kappa_2 = 1$, $\lambda = -5$	
CC8	$f_{\text{vM}}\left(2\pi\big(F_{\text{Ca}}(\theta; \mu, \rho) - \frac{\psi}{2\pi}\big); \mu_g, \kappa_g\right) \times f_{\text{Ca}}(\theta; \mu, \rho)$	$\mu = 0$, $\rho = \frac{1}{2}$, $\mu_g = \pi$, $\kappa_g = 7$	See equations (1) and (2) in Wehrly and Johnson (1979) with a von Mises and a mixture of von Mises as links
CC9	$\frac{1}{2\pi}\big(p_{g_1} f_{\text{vM}}(\theta + \psi; \mu_{g_1}, \kappa_{g_1}) +$ $p_{g_2} f_{\text{vM}}(\theta + \psi; \mu_{g_2}, \kappa_{g_2})\big)$	$\mu_{g_1} = \frac{\pi}{4}$, $\kappa_{g_1} = \kappa_{g_2} = 10$, $\mu_{g_2} = \frac{7\pi}{4}$, $p_{g_1} = p_{g_2} = \frac{1}{2}$	
CC10	$\sum_{p_1=-\infty}^{\infty} \sum_{p_1=-\infty}^{\infty}$ $f_{\mathcal{N}}(\theta + 2\pi p_1, \psi + 2\pi p_2; m_1, m_2, \sigma_1, \sigma_2, \rho)$	$m_1 = 0$, $m_2 = \frac{\pi}{6}$, $\sigma_1 = \frac{3}{2}$, $\sigma_2 = \frac{1}{4}$, $\rho = 0$	See Example 7.3 in Johnson and Wehrly (1977)
CC11		$m_1 = m_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = -\frac{9}{10}$	
CC12	$\left\{4\pi^2\left(1 - 2\rho F_{\text{vM}}(\theta; \mu_1, \kappa_1) F_{\text{vM}}(\psi, \mu_2, \kappa_2) + \rho^2\right)\right\}^{-1}$ $\times (1 - \rho^2) f_{\text{vM}}(\theta; \mu_1, \kappa_1) f_{\text{vM}}(\psi; \mu_2, \kappa_2)$	$\mu_1 = \frac{3\pi}{4}$, $\kappa_1 = 5$, $\mu_2 = 0$, $\kappa_2 = 1$, $\rho = \frac{1}{2}$	See Section 4.1 of Kato (2009)

Table A.3: Circular-circular models.

i. Sample Ψ, a random variable with circular density g.

ii. Sample V, a uniform variable in $[0, 1]$.

iii. Set $U = \frac{(\Psi \mp 2\pi V) \mod 2\pi}{2\pi}$.

A.3.4 Alternative models

The alternative hypothesis for the goodness-of-fit test, both in the circular-linear and circular-circular cases, is stated as:

$$H_{k,\delta} : f = (1 - \delta) f^k_{\boldsymbol{\theta}_0} + \delta\Delta, \quad 0 \leq \delta \leq 1.$$

Three mixing densities Δ are considered, two for the circular-linear situation and one for the circular-circular:

$$\begin{aligned}
\Delta_1(\theta, z) &= f_{\mathrm{vM}}(\theta; \mu_1, \kappa) \times f_{\mathcal{N}}(z; m_1, \sigma_1), \\
\Delta_2(\theta, z) &= f_{\mathrm{vM}}(\theta; \mu_1, \kappa) \times f_{\mathcal{LN}}(z; m_2, \sigma_2), \\
\Delta_3(\theta, \psi) &= f_{\mathrm{vM}}(\theta; \mu_2, \kappa) \times f_{\mathrm{vM}}(\psi; \mu_1, \kappa),
\end{aligned}$$

where $\mu_1 = \pi$, $\mu_2 = 0$, $\kappa = 3$, $m_1 = 2$, $\sigma_1 = 1$ and $m_2 = \sigma_2 = \frac{1}{2}$. To account for similar ranges in the linear data obtained under $H_{k,0}$ and under $H_{k,\delta}$, Δ_1 is used in models CL1, CL4–CL11 and CL13, whereas Δ_2 in the other models. In the circular-circular case, the deviation for all models is Δ_3.

A.3.5 Bandwidth choice

The delicate issue of the bandwidth choice for the testing procedure has been approached as follows. In the simulation results presented in Section 6.6, a fixed pair of bandwidths has been chosen based on a Likelihood Cross Validation criterion. Ideally, one would like to run the test in a grid of several bandwidths to check how the test is affected by the bandwidth choice. This has been done for six circular-linear and circular-circular models, as shown in Figure A.2. Specifically, Figure A.2 shows percentages of rejections under the null ($\delta = 0.00$, green) and under the alternative ($\delta = 0.15$, orange), computed from $M = 1000$ Monte Carlo samples for each pair of bandwidths (the same collection of samples for each pair) on a logarithmic spaced 10×10 grid. The sample size considered is $n = 100$ and the number of bootstrap replicates is $B = 1000$.

As it can be seen, the test is correctly calibrated regardless the bandwidths value. In fact, for all the models explored, the rejection rates for each pair of bandwidths in the grid are inside the 95% confidence interval of the proportion $\alpha = 0.05$ (this happens for 95.75% of the bandwidths in the grid). However, the power is notably affected by the choice of the bandwidths, with rather different behaviours depending on the model and on the alternative. Reasonable choices of the bandwidths based on an estimation criterion such as the one obtained by the median of the LCV bandwidths (6.7) lead in general to a competitive power.

A.3.6 Further results

Tables A.4 and A.5 collect the results of the simulation study for each combination of model (CL or CC), deviation (δ), sample size (n) and significance level (α). When the null hypothesis

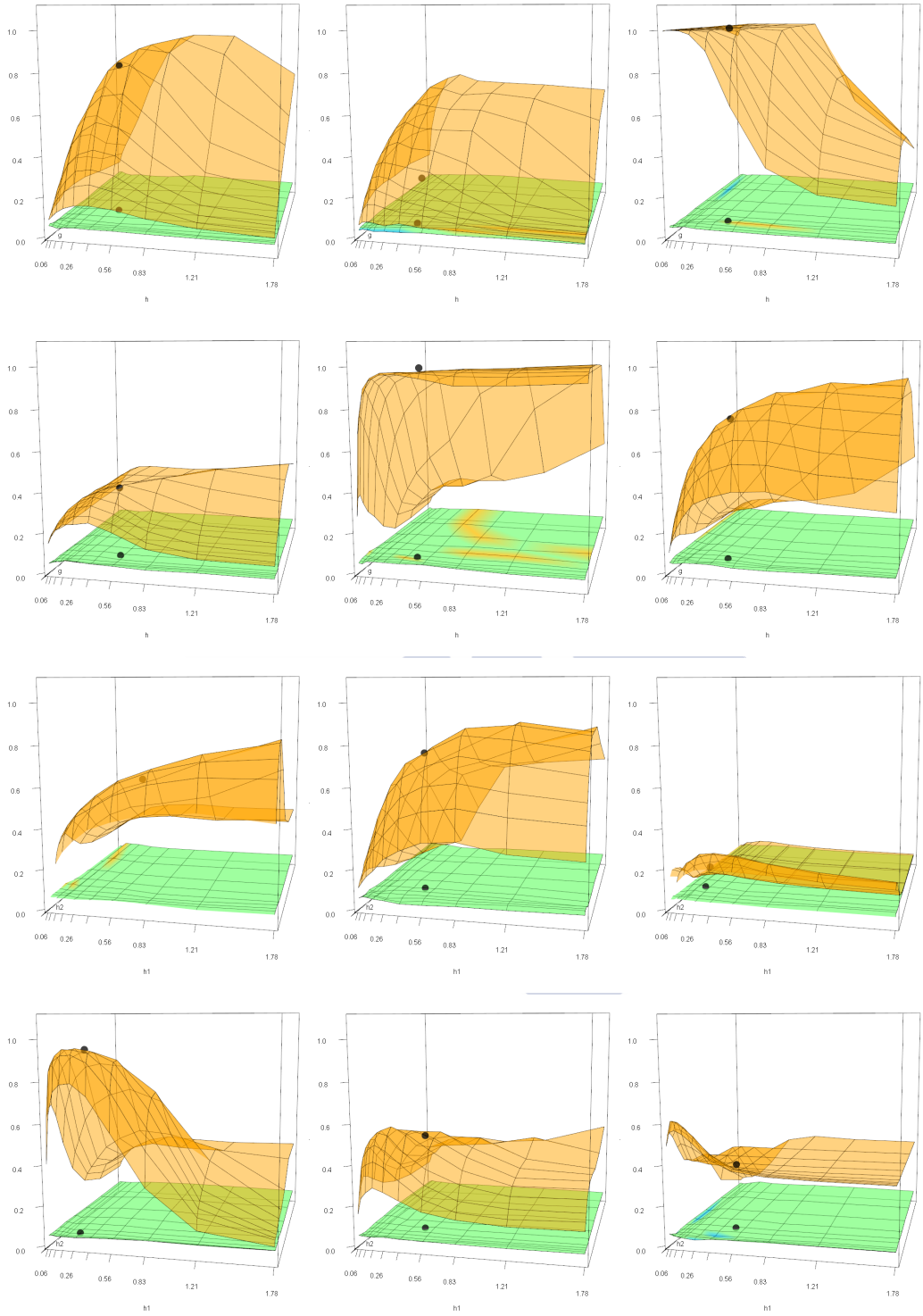


Figure A.2: Empirical size and power of the goodness-of-fit tests for a 10×10 grid of bandwidths. First two rows, from left to right and up to down: models CL1, CL5, CL7, CL8, CL9 and CL11. Last two rows: CC1, CC5, CC7, CC8, CC9 and CC11. Lower surface represents the empirical rejection rate under $H_{0.00}$ and upper surface under $H_{0.15}$. Green colour represent that the empirical rejection is in the 95% confidence interval of $\alpha = 0.05$, blue that is lower and orange that is larger. Black points represent the sized and powers obtained with the median of the LCV bandwidths (for model CC1 under H_0 is outside the grid).

Model	Sample size n and significance level α								
	$n = 100$			$n = 500$			$n = 1000$		
	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01
$H_{1,0.00}$	0.111	0.051	0.010	0.107	0.052	0.013	0.102	0.048	0.013
$H_{2,0.00}$	0.094	0.051	0.013	0.096	0.049	0.010	0.107	0.050	0.009
$H_{3,0.00}$	0.095	0.048	0.014	0.101	0.046	0.014	0.090	0.050	0.009
$H_{4,0.00}$	0.102	0.045	0.009	0.096	0.039	0.011	0.102	0.045	0.008
$H_{5,0.00}$	0.094	0.049	0.009	0.102	0.049	0.009	0.101	0.041	0.009
$H_{6,0.00}$	0.095	0.039	0.010	0.104	0.043	0.010	0.110	0.050	0.015
$H_{7,0.00}$	0.086	0.042	0.013	0.093	0.043	0.008	0.091	0.049	0.016
$H_{8,0.00}$	0.095	0.049	0.011	0.108	0.050	0.003	0.108	0.044	0.006
$H_{9,0.00}$	0.106	0.062	0.016	0.086	0.043	0.010	0.104	0.064	0.015
$H_{10,0.00}$	0.094	0.045	0.007	0.103	0.056	0.018	0.097	0.045	0.005
$H_{11,0.00}$	0.102	0.059	0.009	0.104	0.056	0.010	0.113	0.056	0.013
$H_{12,0.00}$	0.120	0.073	0.020	0.113	0.054	0.013	0.109	0.051	0.010
$H_{1,0.10}$	0.665	0.552	0.355	1.000	0.997	0.981	1.000	1.000	1.000
$H_{2,0.10}$	0.361	0.244	0.107	0.885	0.805	0.579	0.995	0.982	0.898
$H_{3,0.10}$	0.185	0.107	0.032	0.502	0.362	0.166	0.775	0.659	0.421
$H_{4,0.10}$	0.255	0.172	0.060	0.687	0.568	0.322	0.927	0.868	0.697
$H_{5,0.10}$	0.416	0.272	0.087	0.987	0.972	0.894	1.000	1.000	0.999
$H_{6,0.10}$	0.997	0.996	0.988	1.000	1.000	1.000	1.000	1.000	1.000
$H_{7,0.10}$	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
$H_{8,0.10}$	0.325	0.204	0.069	0.940	0.893	0.723	1.000	1.000	0.983
$H_{9,0.10}$	0.947	0.914	0.796	1.000	1.000	1.000	1.000	1.000	1.000
$H_{10,0.10}$	0.340	0.218	0.089	0.829	0.723	0.481	0.962	0.944	0.838
$H_{11,0.10}$	0.618	0.510	0.296	0.996	0.993	0.963	1.000	1.000	1.000
$H_{12,0.10}$	0.230	0.152	0.057	0.788	0.655	0.442	0.991	0.969	0.895
$H_{1,0.15}$	0.883	0.822	0.621	1.000	1.000	1.000	1.000	1.000	1.000
$H_{2,0.15}$	0.650	0.525	0.311	1.000	0.997	0.977	1.000	1.000	1.000
$H_{3,0.15}$	0.281	0.163	0.055	0.776	0.682	0.420	0.970	0.940	0.860
$H_{4,0.15}$	0.399	0.297	0.127	0.910	0.869	0.724	0.998	0.993	0.981
$H_{5,0.15}$	0.663	0.514	0.235	0.999	0.999	0.999	1.000	1.000	1.000
$H_{6,0.15}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$H_{7,0.15}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$H_{8,0.15}$	0.522	0.379	0.168	0.999	0.997	0.976	1.000	1.000	1.000
$H_{9,0.15}$	0.996	0.989	0.962	1.000	1.000	1.000	1.000	1.000	1.000
$H_{10,0.15}$	0.505	0.378	0.154	0.988	0.975	0.893	1.000	1.000	0.996
$H_{11,0.15}$	0.838	0.763	0.567	1.000	1.000	1.000	1.000	1.000	1.000
$H_{12,0.15}$	0.373	0.254	0.114	0.989	0.967	0.872	1.000	1.000	1.000

Table A.4: Empirical size and power of the circular-linear goodness-of-fit test for models CL1–CL12 with different sample sizes, deviations and significance levels.

Model	Sample size n and significance level α								
	$n = 100$			$n = 500$			$n = 1000$		
	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01
$H_{1,0.00}$	0.102	0.061	0.016	0.094	0.047	0.004	0.103	0.048	0.008
$H_{2,0.00}$	0.094	0.054	0.007	0.100	0.043	0.011	0.096	0.056	0.012
$H_{3,0.00}$	0.103	0.061	0.009	0.096	0.042	0.011	0.113	0.058	0.011
$H_{4,0.00}$	0.094	0.049	0.010	0.089	0.048	0.008	0.108	0.052	0.016
$H_{5,0.00}$	0.117	0.059	0.011	0.091	0.050	0.003	0.090	0.051	0.009
$H_{6,0.00}$	0.101	0.069	0.055	0.082	0.045	0.009	0.074	0.034	0.009
$H_{7,0.00}$	0.095	0.048	0.010	0.100	0.059	0.014	0.105	0.044	0.005
$H_{8,0.00}$	0.094	0.043	0.014	0.100	0.054	0.013	0.097	0.050	0.011
$H_{9,0.00}$	0.094	0.043	0.009	0.104	0.057	0.017	0.098	0.042	0.012
$H_{10,0.00}$	0.095	0.047	0.005	0.096	0.041	0.006	0.088	0.042	0.010
$H_{11,0.00}$	0.088	0.041	0.008	0.096	0.047	0.010	0.108	0.053	0.013
$H_{12,0.00}$	0.117	0.062	0.023	0.116	0.058	0.013	0.092	0.048	0.016
$H_{1,0.10}$	0.587	0.456	0.240	0.996	0.995	0.961	1.000	1.000	1.000
$H_{2,0.10}$	0.634	0.506	0.300	0.998	0.994	0.976	1.000	1.000	1.000
$H_{3,0.10}$	0.786	0.706	0.466	1.000	1.000	1.000	1.000	1.000	1.000
$H_{4,0.10}$	0.890	0.837	0.665	1.000	1.000	1.000	1.000	1.000	1.000
$H_{5,0.10}$	0.601	0.431	0.176	1.000	1.000	0.999	1.000	1.000	1.000
$H_{6,0.10}$	0.237	0.123	0.059	0.875	0.759	0.503	0.982	0.958	0.859
$H_{7,0.10}$	0.210	0.112	0.025	0.838	0.724	0.429	0.996	0.989	0.916
$H_{8,0.10}$	0.794	0.693	0.480	1.000	1.000	1.000	1.000	1.000	1.000
$H_{9,0.10}$	0.471	0.325	0.112	1.000	1.000	1.000	1.000	1.000	1.000
$H_{10,0.10}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$H_{11,0.10}$	0.985	0.973	0.910	1.000	1.000	1.000	1.000	1.000	1.000
$H_{12,0.10}$	0.942	0.899	0.788	1.000	1.000	1.000	1.000	1.000	1.000
$H_{1,0.15}$	0.847	0.751	0.521	1.000	1.000	1.000	1.000	1.000	1.000
$H_{2,0.15}$	0.862	0.798	0.627	1.000	1.000	1.000	1.000	1.000	1.000
$H_{3,0.15}$	0.958	0.932	0.830	1.000	1.000	1.000	1.000	1.000	1.000
$H_{4,0.15}$	0.981	0.958	0.885	1.000	1.000	1.000	1.000	1.000	1.000
$H_{5,0.15}$	0.847	0.720	0.445	1.000	1.000	1.000	1.000	1.000	1.000
$H_{6,0.15}$	0.443	0.270	0.097	0.985	0.960	0.858	0.997	0.993	0.982
$H_{7,0.15}$	0.357	0.201	0.043	0.990	0.976	0.879	1.000	1.000	1.000
$H_{8,0.15}$	0.969	0.945	0.842	1.000	1.000	1.000	1.000	1.000	1.000
$H_{9,0.15}$	0.719	0.600	0.345	1.000	1.000	1.000	1.000	1.000	1.000
$H_{10,0.15}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$H_{11,0.15}$	1.000	1.000	0.993	1.000	1.000	1.000	1.000	1.000	1.000
$H_{12,0.15}$	0.999	0.993	0.975	1.000	1.000	1.000	1.000	1.000	1.000

Table A.5: Empirical size and power of the circular-circular goodness-of-fit test for models CC1–CC12 with different sample sizes, deviations and significance levels.

holds, the level of the test is correctly attained for all significance levels, sample sizes and models. Under the alternative, the tests perform satisfactorily, having both of them a quick detection of the alternative when only a 10% and a 15% of the data come from a density not belonging to the null parametric family.

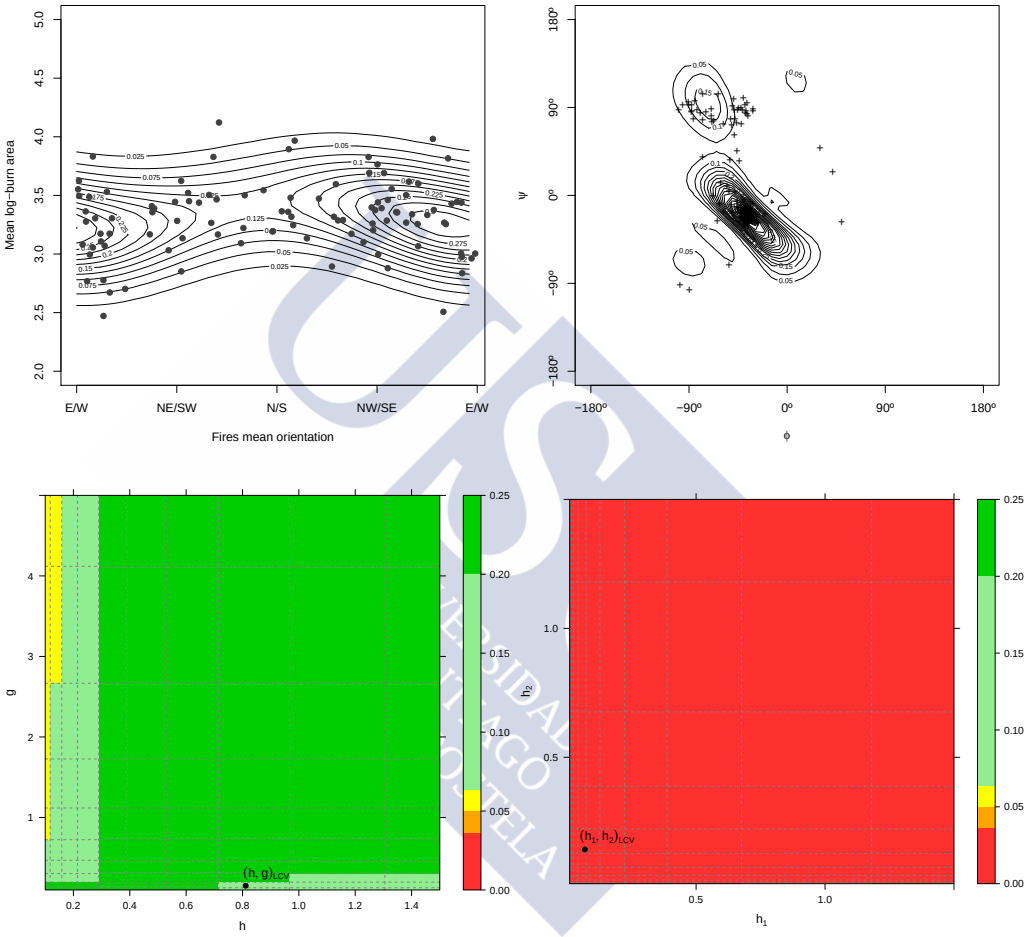


Figure A.3: Upper row, from left to right: parametric fit (model from Mardia and Sutton (1978)) to the circular mean orientation and mean log-burnt area of the fires in each of the 102 watersheds of Portugal; parametric fit (model from Fernández-Durán (2007)) for the dihedral angles of the alanine-alanine-alanine segments. Lower row: p-values of the goodness-of-fit tests for a 10×10 grid, with the LCV bandwidth for the data.

A.4 Extended data application

The analysis of the two real datasets presented in Section 6.7 has been complemented by exploring the effect of different bandwidths in the test. To that aim, Figure A.3 shows the p-values computed from $B = 1000$ bootstrap replicates for a logarithmic spaced 10×10 grid, as well as bandwidths obtained by LCV for each dataset. The graphs shows that there are no evidences against the model of Mardia and Sutton (1978) for modelling the wildfires data and that the model used to describe the proteins dataset is not adequate. This model employs the copula

structure of Wehrly and Johnson (1979) with marginals and link function given by circular densities based on NNTS, specifying Fernández-Durán (2007) that the best fit in terms of BIC arises from considering three components for the NNTS's in the marginals and two for the link function. The fitting of the NNTS densities was performed using the `nntsmanifoldnewtonestimation` function of the package `CircNNTSR` (Fernández-Durán and Gregorio-Domínguez, 2013), which computes the MLE of the NNTS parameters using a Newton algorithm on the hypersphere. The two-step ML procedure described in Section A.3 was employed to fit first the marginals and then the copula. The resulting contour levels of the parametric estimate are quite similar to the ones shown in Figure 5 of Fernández-Durán (2007). The dataset is available as `ProteinsAAA` in the `CircNNTSR` package.

References

Agostinelli, C. and Lund, U. (2013). *R package* `circular`*: circular statistics (version 0.4-7).*

Brown, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.*, 42(1):59–66.

Fernández-Durán, J. J. (2007). Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics*, 63(2):579–585.

Fernández-Durán, J. J. and Gregorio-Domínguez, M. M. (2013). *CircNNTSR: an R package for the statistical analysis of circular data using NonNegative Trigonometric Sums (NNTS) models.* R package version 2.1.

García-Portugués, E., Barros, A. M. G., Crujeiras, R. M., González-Manteiga, W., and Pereira, J. (2014). A test for directional-linear independence, with applications to wildfire orientation and size. *Stoch. Environ. Res. Risk Assess.*, 28(5):1261–1275.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013a). Exploring wind direction and SO_2 concentration by circular-linear density estimation. *Stoch. Environ. Res. Risk Assess.*, 27(5):1055–1067.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013b). Kernel density estimation for directional-linear data. *J. Multivariate Anal.*, 121:152–175.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.*, 14(1):1–16.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application.* Academic Press, New York.

Hornik, K. and Grün, B. (2012). *movMF: mixtures of von Mises-Fisher distributions.* R package version 0.1-0.

Johnson, R. A. and Wehrly, T. (1977). Measures and models for angular correlation and angular-linear correlation. *J. Roy. Statist. Soc. Ser. B*, 39(2):222–229.

Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *J. Amer. Statist. Assoc.*, 73(363):602–606.

Kato, S. (2009). A distribution for a pair of unit vectors generated by Brownian motion. *Bernoulli*, 15(3):898–921.

Mächler, M. (2013). *nor1mix: normal (1-d) mixture models (S3 classes and methods).* R package version 1.1-4.

Mardia, K. V. and Sutton, T. W. (1978). A model for cylindrical variables with applications. *J. Roy. Statist. Soc. Ser. B*, 40(2):229–233.

Nelsen, R. B. (2006). *An introduction to copulas.* Springer Series in Statistics. Springer, New York, second edition.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F., and Clark, C. W., editors (2010). *NIST handbook of mathematical functions.* Cambridge University Press, Cambridge.

Rosenblatt, M. and Wahlen, B. E. (1992). A nonparametric measure of independence under a hypothesis of independent components. *Statist. Probab. Lett.*, 15(3):245–252.

Shieh, G. S. and Johnson, R. A. (2005). Inferences based on a bivariate distribution with von Mises marginals. *Ann. Inst. Statist. Math.*, 57(4):789–802.

Singh, H., Hnizdo, V., and Demchuk, E. (2002). Probabilistic model for two dependent circular variables. *Biometrika*, 89(3):719–723.

Venables, W. N. and Ripley, B. D. (2002). *Modern applied statistics with S.* Statistics and Computing. Springer, New York, four edition.

Wehrly, T. E. and Johnson, R. A. (1979). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, 67(1):255–256.

Wood, A. T. A. (1994). Simulation of the von Mises Fisher distribution. *Commun. Stat. Simulat.*, 23(1):157–164.

Zhao, L. and Wu, C. (2001). Central limit theorem for integrated square error of kernel estimators of spherical density. *Sci. China Ser. A*, 44(4):474–483.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Appendix B

Supplement to Chapter 7

This supplement is organized as follows. Section B.1 contains particular cases of the projected local estimator for the circular and spherical situations and their relations with polar and spherical coordinates. Section B.2 gives the technical lemmas used to prove the main results in the paper. Finally, Section B.3 provides further details about the simulation study and gives extra results omitted in the paper.

B.1 Particular cases of the projected local estimator

Some interesting cases and relations of the projected estimator are the following ones.

B.1.1 Local constant

If $p = 0$, then $W_n^0(\mathbf{x}, \mathbf{X}_i) = \frac{L_h(\mathbf{x},\mathbf{X}_i)}{\sum_{j=1}^n L_h(\mathbf{x},\mathbf{X}_j)}$ and the Nadaraya-Watson estimator for directional predictor and scalar response, firstly proposed by Wang et al. (2000), is obtained:

$$\hat{m}_{h,0}(\mathbf{x}) = \sum_{i=1}^n \frac{L_h(\mathbf{x}, \mathbf{X}_i)}{\sum_{j=1}^n L_h(\mathbf{x}, \mathbf{X}_j)} Y_i = \sum_{i=1}^n \frac{L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right)}{\sum_{j=1}^n \left(\frac{1-\mathbf{x}^T\mathbf{X}_j}{h^2}\right)} Y_i.$$

For $q = 1$, denoting $\mathbf{x} = (\cos\theta, \sin\theta)^T$, for $\theta \in [0, 2\pi)$ the circular sample can be identified with a set of angles $\{\Theta_i\}_{i=1}^n$ and the usual notation for circular statistics applies. Then, the local constant estimator for circular data is given by

$$\hat{m}_{h,0}(\theta) = \sum_{i=1}^n \frac{L\left(\frac{1-\cos(\Theta_i-\theta)}{h^2}\right)}{\sum_{j=1}^n L\left(\frac{1-\cos(\Theta_j-\theta)}{h^2}\right)} Y_i = \sum_{i=1}^n \frac{\exp\left\{-\frac{\cos(\Theta_i-\theta)}{h^2}\right\} Y_i}{\sum_{j=1}^n \exp\left\{-\frac{\cos(\Theta_j-\theta)}{h^2}\right\}},$$

where the second equality holds if L is the von Mises kernel.

For $q = 2$, denoting $\mathbf{x} = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)^T$, for $\theta \in [0, 2\pi)$ and $\phi \in [0, \pi)$, the spherical sample can be identified as the pairs of angles $\{(\Theta_i, \Phi_i)\}_{i=1}^n$. Therefore, the local constant estimator for spherical data is given by

$$\hat{m}_{h,0}(\theta, \phi) = \sum_{i=1}^n \frac{L\left(\frac{1-\sin\phi\sin\Phi_i\cos(\Theta_i-\theta)-\cos\phi\cos\Phi_i}{h^2}\right)}{\sum_{j=1}^n L\left(\frac{1-\sin\phi\sin\Phi_j\cos(\Theta_j-\theta)-\cos\phi\cos\Phi_j}{h^2}\right)} Y_i.$$

B.1.2 Local linear with $q = 1$

Let denote $\mathbf{x} = (\cos\theta, \sin\theta)^T$, for $\theta \in [0, 2\pi)$. The matrix $\mathbf{B}_\mathbf{x}$ is formed by the vector $\mathbf{b}_1 = \pm(-\sin\theta, \cos\theta)^T$, which is the orthonormal vector to $\mathbf{x}$. Then, by the sine subtraction formula $\mathbf{B}_\mathbf{x}^T(\cos\Theta_i - \cos\theta, \sin\Theta_i - \sin\theta)^T = \pm\sin(\Theta_i - \theta)$ and as a consequence (7.4) can be expressed as

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^2} \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 \sin(\Theta_i - \theta)\right)^2 c_{h,1}(L) L\left(\frac{1-\cos(\Theta_i-\theta)}{h^2}\right) \tag{B.1}$$

and the solution (7.5) is given by the design matrix

$$\boldsymbol{\mathcal{X}}_{\mathbf{x},1} = \begin{pmatrix} 1 & \sin(\Theta_1 - \theta) \\ \vdots & \vdots \\ 1 & \sin(\Theta_n - \theta) \end{pmatrix}.$$

The resulting estimate is the local linear estimator proposed by Di Marzio et al. (2009) for circular predictors and for circular kernels which are functions of $\kappa(1 - \cos\theta)$ (the change in notation is $\kappa \equiv 1/h^2$). The equivalence of both estimators can be seen also from examining the equality of their design matrices and weights or from the Taylor expansions that motivate them. By the chain rule, it can be seen that the derivative of the regression function in the circular argument, as considered in Di Marzio et al. (2009), is the same as the projected gradient of m:

$$\frac{d}{d\theta} m(\theta) = \boldsymbol{\nabla} m(\mathbf{x})\big|_{\mathbf{x}=\binom{\cos\theta}{\sin\theta}}^T \frac{\partial \mathbf{x}(\theta)}{\partial\theta} = \boldsymbol{\nabla} m(\mathbf{x})\big|_{\mathbf{x}=\binom{\cos\theta}{\sin\theta}}^T \mathbf{B}_\mathbf{x}.$$

Finally, if θ is close to Θ_i (in modulo 2π), then $\sin(\Theta_i - \theta) \approx \Theta_i - \theta$ and the linear coefficient of the local estimator captures indeed a linear effect of close angles in the response.

The circular case of the local linear estimator in Di Marzio et al. (2014) is different from the circular projected local estimator and the one in Di Marzio et al. (2009). The minimum weighted squares problem, using the tangent-normal decomposition ant translated to this paper's notation, is stated as follows:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^3} \sum_{i=1}^{n} \left(Y_i - \beta_0 - (\beta_1, \beta_2)^T \eta_i \boldsymbol{\xi}_i\right)^2 c_{h,1}(L) L\left(\frac{1-\cos(\Theta_i-\theta)}{h^2}\right), \tag{B.2}$$

where η_i is such that $\cos\eta_i = \mathbf{x}^T\mathbf{X}_i$ and $\boldsymbol{\xi}_i = \frac{\mathbf{X}_i - \mathbf{x}\cos\eta_i}{\sin\eta_i}$ if $\sin\eta_i \neq 0$ and $\boldsymbol{\xi}_i = \pm\mathbf{X}_i$ otherwise. After considering the polar coordinates and doing some trigonometric algebra, it results that $\eta_i\boldsymbol{\xi}_i = (\theta - \Theta_i)(-\sin\theta, \cos\theta)$, so after identifying $\beta_1' = -\beta_1\sin\theta + \beta_2\cos\theta$, the minimization (B.2) is equivalent to

$$\min_{(\beta_0,\beta_1')\in\mathbb{R}^2} \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1'(\Theta_i - \theta)\right)^2 c_{h,1}(L) L\left(\frac{1-\cos(\Theta_i-\theta)}{h^2}\right).$$

Provided that $\sin(\Theta_i - \theta) \approx \Theta_i - \theta$ for close angles, the practical difference between (B.1) and (B.2) relies only in small samples and for large bandwidths.

Finally, it is possible to compute the exact expression for the estimator using the exact inversion formula of the 2×2 matrix $\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_\mathbf{x} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}$ (as it is done in Wand and Jones (1995), among others). This yields

$$\hat{m}_{h,0}(\theta) = \frac{s_2(\theta)t_0(\theta) - s_1(\theta)t_1(\theta)}{s_2(\theta)s_0(\theta) - s_1(\theta)^2},$$

where, for $j = 0, 1, 2$,

$$s_j(\theta) = \sum_{i=1}^{n} L\left(\frac{1-\cos(\Theta_i - \theta)}{h^2}\right) \sin^j(\Theta_i - \theta),$$
$$t_j(\theta) = \sum_{i=1}^{n} L\left(\frac{1-\cos(\Theta_i - \theta)}{h^2}\right) \sin^j(\Theta_i - \theta) Y_i.$$

B.1.3 Local linear with $q = 2$

Let denote $\mathbf{x} = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)^T$, for $\theta \in [0, 2\pi)$ and $\phi \in [0, \pi)$. Now the matrix $\mathbf{B}_\mathbf{x}$ is given by vectors $\mathbf{b}_1 = \pm\left(x_1^2 + x_2^2\right)^{-\frac{1}{2}}(-x_2, x_1, 0)^T$ and $\mathbf{b}_2 = \pm\left(x_1^2 + x_2^2\right)^{-\frac{1}{2}}(-x_1x_3, -x_2x_3, x_1^2 + x_2^2)^T$ if $|x_3| \neq 1$ (if $|x_3| = 1$, then $\mathbf{b}_1 = \pm(1, 0, 0)$ and $\mathbf{b}_2 = \pm(0, 1, 0)$ complete the orthonormal basis). Therefore, after some trigonometric identities,

$$\mathbf{B}_\mathbf{x}^T(\mathbf{X}_i - \mathbf{x}) = \left(\sin\Phi_i \sin(\Theta_i - \theta),\ -\cos\phi\sin\Phi_i\cos(\Theta_i - \theta) + \sin\phi\cos\Phi_i\right).$$

As a consequence, the solution (7.5) is given by the design matrix

$$\boldsymbol{\mathcal{X}}_{\mathbf{x},1} = \begin{pmatrix} 1 & \sin\Phi_1\sin(\Theta_1 - \theta) & -\cos\phi\sin\Phi_1\cos(\Theta_1 - \theta) + \sin\phi\cos\Phi_1 \\ \vdots & \vdots & \vdots \\ 1 & \sin\Phi_n\sin(\Theta_n - \theta) & -\cos\phi\sin\Phi_n\cos(\Theta_n - \theta) + \sin\phi\cos\Phi_n \end{pmatrix}.$$

The second and third columns are almost linear in the angles θ and ϕ, respectively: if θ is close to Θ_i (in modulo 2π) and ϕ_i is close to Φ_i (in modulo π), then $\cos(\Theta_i - \theta) \approx 1$ and hence $\sin\Phi_i\sin(\Theta_i - \theta) \approx \sin\Phi_i(\Theta_i - \theta)$, so

$$-\cos\phi\sin\Phi_i\cos(\Theta_i - \theta) + \sin\phi\cos\Phi_i \approx \sin(\Phi_i - \phi) \approx \Phi_i - \phi.$$

Furthermore, as happens with the circular case, the projected gradient of m used in the projected local estimator comprises naturally the estimator that follows from considering the function m defined throughout spherical coordinates and taking the derivatives on them:

$$\left(\frac{\partial m(\theta,\phi)}{\partial\theta}, \frac{\partial m(\theta,\phi)}{\partial\phi}\right) = \boldsymbol{\nabla} m(\mathbf{x})\Big|_{\mathbf{x}=\begin{pmatrix}\sin\phi\cos\theta\\ \sin\phi\sin\theta\\ \cos\phi\end{pmatrix}}^T \frac{\partial\mathbf{x}(\theta,\phi)}{\partial\theta\partial\phi} = \boldsymbol{\nabla} m(\mathbf{x})\Big|_{\mathbf{x}=\begin{pmatrix}\sin\phi\cos\theta\\ \sin\phi\sin\theta\\ \cos\phi\end{pmatrix}}^T \mathbf{B}_\mathbf{x}.$$

Finally, the exact expression for the estimator can also be obtained using the exact inversion formula of the 3×3 matrix $\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_\mathbf{x}\boldsymbol{\mathcal{X}}_{\mathbf{x},p}$. To that end, recall that

$$\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_\mathbf{x}\boldsymbol{\mathcal{X}}_{\mathbf{x},p} = c_{h,2}(L)\begin{pmatrix} s_{00}(\theta,\phi) & s_{10}(\theta,\phi) & s_{01}(\theta,\phi) \\ s_{10}(\theta,\phi) & s_{20}(\theta,\phi) & s_{11}(\theta,\phi) \\ s_{01}(\theta,\phi) & s_{11}(\theta,\phi) & s_{02}(\theta,\phi) \end{pmatrix},\ \boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T\boldsymbol{\mathcal{W}}_\mathbf{x}\mathbf{Y} = c_{h,2}(L)\begin{pmatrix} t_{00}(\theta,\phi) \\ t_{10}(\theta,\phi) \\ t_{01}(\theta,\phi) \end{pmatrix},$$

where, for $j, k = 0, 1, 2$,

$$s_{jk}(\theta,\phi) = \sum_{i=1}^{n} L\left(\frac{1-\sin\phi\sin\Phi_i\cos(\Theta_i - \theta) - \cos\phi\cos\Phi_i}{h^2}\right)(\sin\Phi_i\sin(\Theta_i - \theta))^j$$
$$\times\left(-\cos\phi\sin\Phi_i\cos(\Theta_i - \theta) + \sin\phi\cos\Phi_i\right)^k,$$

$$t_{jk}(\theta,\phi) = \sum_{i=1}^{n} L\left(\frac{1-\sin\phi\sin\Phi_i\cos(\Theta_i-\theta)-\cos\phi\cos\Phi_i}{h^2}\right)(\sin\Phi_i\sin(\Theta_i-\theta))^j \times(-\cos\phi\sin\Phi_i\cos(\Theta_i-\theta)+\sin\phi\cos\Phi_i)^k Y_i.$$

Therefore, after some matrix algebra it turns out that

$$\hat{m}_{h,0}(\theta,\phi) = \frac{((s_{20}s_{02}-s_{11}^2)t_{00})(\theta,\phi)-((s_{10}s_{02}-s_{01}s_{11})t_{10})(\theta,\phi)+((s_{10}s_{11}-s_{01}s_{20})t_{01})(\theta,\phi)}{((s_{20}s_{02}-s_{11}^2)s_{00})(\theta,\phi)-((s_{10}s_{02}-s_{01}s_{11})s_{10})(\theta,\phi)+((s_{10}s_{11}-s_{01}s_{20})s_{01})(\theta,\phi)}.$$

B.2 Technical lemmas

B.2.1 Projected local estimator properties

Lemma B.1. *Under assumptions **A1**–**A4**, for a random sample $\{(\mathbf{X}_i,Y_i)\}_{i=1}^n$ the following statements hold with uniform orders for any point $\mathbf{x}\in\Omega_q$:*

i. $\hat{f}_h(\mathbf{x}) = f(\mathbf{x}) + o_{\mathbb{P}}(1)$.

ii. $\frac{1}{n}\sum_{i=1}^n L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x}) = \frac{2b_q(L)}{q}\mathbf{B}_{\mathbf{x}}^T\nabla f(\mathbf{x})h^2 + o(h^2\mathbf{1}) + \mathcal{O}_{\mathbb{P}}\left(\frac{h}{\sqrt{nh^q}}\mathbf{1}\right)$.

iii. $\frac{1}{n}\sum_{i=1}^n L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})Y_i = \mathcal{O}(h^2\mathbf{1}) + \mathcal{O}_{\mathbb{P}}\left(\frac{h}{\sqrt{nh^q}}\mathbf{1}\right)$.

iv. $\frac{1}{n}\sum_{i=1}^n L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}} = \frac{2b_q(L)}{q}\mathbf{I}_q f(\mathbf{x})h^2 + o_{\mathbb{P}}(h^2\mathbf{1}\mathbf{1}^T)$.

v. $\frac{1}{n}\sum_{i=1}^n L_h(\mathbf{x},\mathbf{X}_i)(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\mathcal{H}}_m(\mathbf{x})\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x}) = \frac{2b_q(L)}{q}\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_m(\mathbf{x})\right]f(\mathbf{x})h^2 + o_{\mathbb{P}}(h^2)$.

vi. $\frac{1}{n}\sum_{i=1}^n L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\mathcal{H}}_m(\mathbf{x})\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x}) = o_{\mathbb{P}}(h^3\mathbf{1})$.

vii. $\frac{1}{n}\sum_{i=1}^n L_h^2(\mathbf{x},\mathbf{X}_i)\sigma^2(\mathbf{X}_i) = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{h^q}\sigma^2(\mathbf{x})f(\mathbf{x}) + o_{\mathbb{P}}(h^{-q})$.

viii. $\frac{1}{n}\sum_{i=1}^n L_h^2(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})\sigma^2(\mathbf{X}_i) = o_{\mathbb{P}}(h^{-q}\mathbf{1})$.

ix. $\frac{1}{n}\sum_{i=1}^n L_h^2(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\sigma^2(\mathbf{X}_i) = o_{\mathbb{P}}(h^{-q}\mathbf{1})$.

Proof of Lemma B.1. Chebychev's inequality, Lemma B.6 and Taylor expansions will be used for each statement in which the proof is divided.

Proof of i. By Chebychev's inequality, $\hat{f}_h(\mathbf{x}) = \mathbb{E}[\hat{f}_h(\mathbf{x})] + \mathcal{O}_{\mathbb{P}}\left(\sqrt{\mathbb{V}\mathrm{ar}[\hat{f}_h(\mathbf{x})]}\right)$. It follows by Lemma B.6 that $\mathbb{E}[\hat{f}_h(\mathbf{x})] = f(\mathbf{x}) + o(1)$ and that $\mathbb{V}\mathrm{ar}[\hat{f}_h(\mathbf{x})] = \frac{1}{nh^q\lambda_q(L)}(f(\mathbf{x})+o(1))$, with the remaining orders being uniform in $\mathbf{x}\in\Omega_q$. Then, as f is continuous in Ω_q by assumption **A1** it is also bounded, so by **A4** $\mathbb{V}\mathrm{ar}[\hat{f}_h(\mathbf{x})] = o(1)$ uniformly, which results in $\hat{f}_h(\mathbf{x}) = f(\mathbf{x}) + o_{\mathbb{P}}(1)$ uniformly in $\mathbf{x}\in\Omega_q$.

Proof of ii. Applying Lemma B.4 and the change of variables $r = \frac{1-t}{h^2}$,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})\right] = \mathbb{E}\left[L_h(\mathbf{x},\mathbf{X})\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}-\mathbf{x})\right]$$

$$
\begin{aligned}
&= c_{h,q}(L) \int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) \mathbf{B}_{\mathbf{x}}^T(\mathbf{y}-\mathbf{x}) f(\mathbf{y})\, \omega_q(d\mathbf{y}) \\
&= c_{h,q}(L) \int_{-1}^{1} \int_{\Omega_{q-1}} L\left(\frac{1-t}{h^2}\right) \boldsymbol{\xi} f\left(t\mathbf{x} + (1-t^2)^{\frac{1}{2}} \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\right) (1-t^2)^{\frac{q-1}{2}}\, \omega_{q-1}(d\boldsymbol{\xi})\, dt \\
&= c_{h,q}(L) h^2 \int_0^{2h^{-2}} \int_{\Omega_{q-1}} L(r)\, \boldsymbol{\xi} f\left(\mathbf{x} + \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right) \left[rh^2(2-rh^2)\right]^{\frac{q-1}{2}}\, \omega_{q-1}(d\boldsymbol{\xi})\, dr \\
&= c_{h,q}(L) h^{q+1} \int_0^{2h^{-2}} L(r)\, r^{\frac{q-1}{2}} (2-rh^2)^{\frac{q-1}{2}} \int_{\Omega_{q-1}} f\left(\mathbf{x} + \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right) \boldsymbol{\xi}\, \omega_{q-1}(d\boldsymbol{\xi})\, dr, \qquad \text{(B.3)}
\end{aligned}
$$

where $\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}} = -rh^2\mathbf{x} + \left[rh^2(2-rh^2)\right]^{\frac{1}{2}} \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}$. The inner integral in (B.3) is computed by a Taylor expansion

$$
f(\mathbf{x} + \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}) = f(\mathbf{x}) + \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T \boldsymbol{\nabla} f(\mathbf{x}) + \mathcal{O}\left(\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right), \qquad \text{(B.4)}
$$

where the remaining order involves the second derivative of f, which is bounded, thus being the order uniform in $\mathbf{x}$. Using Lemma B.5, the first and second addends are:

$$
\begin{aligned}
\int_{\Omega_{q-1}} f(\mathbf{x})\, \boldsymbol{\xi}\, \omega_{q-1}(d\boldsymbol{\xi}) &= 0, \\
\int_{\Omega_{q-1}} \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T \boldsymbol{\nabla} f(\mathbf{x})\, \boldsymbol{\xi}\, \omega_{q-1}(d\boldsymbol{\xi}) &= \left[rh^2(2-rh^2)\right]^{\frac{1}{2}} \int_{\Omega_{q-1}} (\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T \boldsymbol{\nabla} f(\mathbf{x})\, \boldsymbol{\xi}\, \omega_{q-1}(d\boldsymbol{\xi}) \\
&= \left[rh^2(2-rh^2)\right]^{\frac{1}{2}} \int_{\Omega_{q-1}} \sum_{i,j=1}^{q} \xi_i \mathbf{B}_{\mathbf{x}}^T \boldsymbol{\nabla} f(\mathbf{x})\, \xi_j\, \omega_{q-1}(d\boldsymbol{\xi}) \\
&= \frac{\omega_{q-1}}{q} \left[rh^2(2-rh^2)\right]^{\frac{1}{2}} \mathbf{B}_{\mathbf{x}}^T \boldsymbol{\nabla} f(\mathbf{x}).
\end{aligned}
$$

The third addend is $\mathcal{O}(\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}) = \mathcal{O}\left(h^2\mathbf{1}\right)$, because $\mathbf{B}_{\mathbf{x}}^T\mathbf{x} = \mathbf{0}$ and $(\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi} = \boldsymbol{\xi}^T\mathbf{I}_q\boldsymbol{\xi} = 1$. Therefore, (B.3) becomes

$$
\begin{aligned}
\text{(B.3)} &= c_{h,q}(L) h^{q+2} \frac{\omega_{q-1}}{q} \int_0^{2h^{-2}} L(r)\, r^{\frac{q}{2}} (2-rh^2)^{\frac{q}{2}}\, dr\, \mathbf{B}_{\mathbf{x}}^T \boldsymbol{\nabla} f(\mathbf{x}) \\
&\quad + c_{h,q}(L) h^{q+1} \int_0^{2h^{-2}} L(r)\, r^{\frac{q}{2}} (2-rh^2)^{\frac{q}{2}}\, dr\, \mathcal{O}\left(h^2\mathbf{1}\right) \\
&= \left(b_q(L) + o(1)\right) \frac{2b_q(L)}{q} \mathbf{B}_{\mathbf{x}}^T \boldsymbol{\nabla} f(\mathbf{x}) h^2 + \mathcal{O}\left(h^3\mathbf{1}\right) \\
&= \frac{2b_q(L)}{q} \mathbf{B}_{\mathbf{x}}^T \boldsymbol{\nabla} f(\mathbf{x}) h^2 + o\left(h^2\mathbf{1}\right), \qquad \text{(B.5)}
\end{aligned}
$$

where the second last equality follows from applying the Dominated Convergence Theorem (DCT), (7.2) and the definition of $b_q(L)$. See the proof of Theorem 1 in García-Portugués et al. (2013) for the technical details involved in a similar situation.

As the Chebychev inequality is going to be applied componentwise, the interest is now in the order of the variance vector. To that end, the square of a vector will denote the vector with correspondent squared components. By analogous computations,

$$
\mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{i=1}^{n} L_h(\mathbf{x}, \mathbf{X}_i) \mathbf{B}_{\mathbf{x}}^T (\mathbf{X}_i - \mathbf{x})\right]
$$

$$
\begin{aligned}
&\leq \frac{1}{n}\mathbb{E}\left[L_h(\mathbf{x},\mathbf{X})^2(\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}-\mathbf{x}))^2\right]\\
&= \frac{c_{h,q}(L)^2}{n}\int_{\Omega_q} L^2\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right)(\mathbf{B}_{\mathbf{x}}^T(\mathbf{y}-\mathbf{x}))^2 f(\mathbf{y})\,\omega_q(d\mathbf{y})\\
&= \frac{c_{h,q}(L)^2}{n}\int_{-1}^{1}\int_{\Omega_{q-1}} L^2\left(\frac{1-t}{h^2}\right)\boldsymbol{\xi}^2 f\left(t\mathbf{x}+(1-t^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\right)(1-t^2)^{\frac{q}{2}}\,\omega_{q-1}(d\boldsymbol{\xi})\,dt\\
&= \frac{c_{h,q}(L)^2h^2}{n}\int_0^{2h^{-2}}\int_{\Omega_{q-1}} L^2(r)\,\boldsymbol{\xi}^2 f(\mathbf{x}+\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}})\left[rh^2(2-rh^2)\right]^{\frac{q}{2}}\,\omega_{q-1}(d\boldsymbol{\xi})\,dr\\
&= \frac{c_{h,q}(L)^2h^{q+2}}{n}\int_0^{2h^{-2}} L^2(r)\,r^{\frac{q}{2}}(2-rh^2)^{\frac{q}{2}}\int_{\Omega_{q-1}} f(\mathbf{x}+\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}})\,\boldsymbol{\xi}^2\,\omega_{q-1}(d\boldsymbol{\xi})\,dr\\
&= \frac{c_{h,q}(L)^2h^{q+2}}{n}\int_0^{2h^{-2}} L^2(r)\,r^{\frac{q}{2}}(2-rh^2)^{\frac{q}{2}}\mathcal{O}(\mathbf{1})\,dr\\
&= \mathcal{O}\left(\frac{h^2}{nh^q}\mathbf{1}\right).
\end{aligned}
\tag{B.6}
$$

The result follows from Chebychev's inequality, (B.5) and (B.6).

Proof of iii. The result is proved form the previous proof and the tower property of the conditional expectation. The expectation can be expressed as

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})Y_i\right] &= \mathbb{E}\left[\mathbb{E}\left[L_h(\mathbf{x},\mathbf{X})\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}-\mathbf{x})Y\middle|\mathbf{X}\right]\right]\\
&= \mathbb{E}\left[L_h(\mathbf{x},\mathbf{X})\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}-\mathbf{x})m(\mathbf{X})\right]\\
&= c_{h,q}(L)\int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right)\mathbf{B}_{\mathbf{x}}^T(\mathbf{y}-\mathbf{x})m(\mathbf{y})f(\mathbf{y})\,\omega_q(d\mathbf{y}).
\end{aligned}
$$

Then, replicating the proof of *ii*, it is easily seen that the order is $\mathcal{O}\left(h^2\mathbf{1}\right)$. The order of the variance is obtained in the same way:

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{i=1}^n L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})Y_i\right] &\leq \frac{1}{n}\mathbb{E}\left[L_h^2(\mathbf{x},\mathbf{X})(\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}-\mathbf{x}))^2Y^2\right]\\
&= \frac{1}{n}\mathbb{E}\left[L_h(\mathbf{x},\mathbf{X})(\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}-\mathbf{x}))^2(\sigma^2(\mathbf{X})+m(\mathbf{X})^2)\right]\\
&= \mathcal{O}\left(\frac{h^2}{nh^q}\mathbf{1}\right).
\end{aligned}
$$

As a consequence, $\frac{1}{n}\sum_{i=1}^n L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})Y_i = \mathcal{O}\left(h^2\mathbf{1}\right)+\mathcal{O}_{\mathbb{P}}\left(\frac{h}{\sqrt{nh^q}}\mathbf{1}\right)$.

Proof of iv. The steps of the proof of *ii* are replicated:

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\right]\\
&\quad = c_{h,q}(L)\int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right)\mathbf{B}_{\mathbf{x}}^T(\mathbf{y}-\mathbf{x})(\mathbf{y}-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}f(\mathbf{y})\,\omega_q(d\mathbf{y})
\end{aligned}
$$

$$
\begin{aligned}
&= c_{h,q}(L) \int_{-1}^{1} \int_{\Omega_{q-1}} L\left(\frac{1-t}{h^2}\right) \boldsymbol{\xi}\boldsymbol{\xi}^T f\left(t\mathbf{x} + (1-t^2)^{\frac{1}{2}} \mathbf{B}_{\mathbf{x}} \boldsymbol{\xi}\right) (1-t^2)^{\frac{q}{2}} \, \omega_{q-1}(d\boldsymbol{\xi}) \, dt \\
&= c_{h,q}(L) h^{q+2} \int_{0}^{2h^{-2}} L(r) \, r^{\frac{q}{2}} (2 - rh^2)^{\frac{q}{2}} \int_{\Omega_{q-1}} f\left(\mathbf{x} + \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right) \boldsymbol{\xi}\boldsymbol{\xi}^T \, \omega_{q-1}(d\boldsymbol{\xi}) \, dr. \qquad \text{(B.7)}
\end{aligned}
$$

The second integral of (B.7) is obtained by expansion (B.4) and Lemma B.5:

$$
\begin{aligned}
\int_{\Omega_{q-1}} f(\mathbf{x}) \, \boldsymbol{\xi}\boldsymbol{\xi}^T \mathbf{B}_{\mathbf{x}} \, \omega_{q-1}(d\boldsymbol{\xi}) &= \frac{\omega_{q-1}}{q} \mathbf{I}_q f(\mathbf{x}), \\
\int_{\Omega_{q-1}} \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}^T \boldsymbol{\nabla} f(\mathbf{x}) \, \boldsymbol{\xi}\boldsymbol{\xi}^T \, \omega_{q-1}(d\boldsymbol{\xi}) &= \int_{\Omega_{q-1}} -rh^2 \mathbf{x}^T \boldsymbol{\xi}\boldsymbol{\xi}^T \, \omega_{q-1}(d\boldsymbol{\xi}) = \mathcal{O}\left(h^2 \mathbf{1}\mathbf{1}^T\right).
\end{aligned}
$$

As the third addend given by expansion (B.4) has order $\mathcal{O}\left(h^2 \mathbf{1}\mathbf{1}^T\right)$, it results that:

$$
\begin{aligned}
\text{(B.7)} &= c_{h,q}(L) h^{q+2} \int_{0}^{2h^{-2}} L(r) \, r^{\frac{q}{2}} (2 - rh^2)^{\frac{q}{2}} \left\{ \frac{\omega_{q-1}}{q} \mathbf{I}_q f(\mathbf{x}) + \mathcal{O}\left(h^2 \mathbf{1}\mathbf{1}^T\right) \right\} dr \\
&= (b_q(L) + o(1)) \left\{ \frac{2}{q} \mathbf{I}_q f(\mathbf{x}) + \mathcal{O}\left(h^2 \mathbf{1}\mathbf{1}^T\right) \right\} h^2 \\
&= \frac{2 b_q(L)}{q} \mathbf{I}_q f(\mathbf{x}) \, h^2 + o\left(h^2 \mathbf{1}\mathbf{1}^T\right), \qquad \text{(B.8)}
\end{aligned}
$$

using the same arguments as in *ii*. The order of the variance is

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}\left[\frac{1}{n} \sum_{i=1}^{n} L_h(\mathbf{x}, \mathbf{X}_i) \mathbf{B}_{\mathbf{x}}^T (\mathbf{X}_i - \mathbf{x}) (\mathbf{X}_i - \mathbf{x})^T \mathbf{B}_{\mathbf{x}} \right] \\
&\quad \leq \frac{1}{n} \mathbb{E}\left[L_h(\mathbf{x}, \mathbf{X})^2 \left(\mathbf{B}_{\mathbf{x}}^T (\mathbf{X} - \mathbf{x}) (\mathbf{X} - \mathbf{x})^T \mathbf{B}_{\mathbf{x}} \right)^2 \right] \\
&\quad = \frac{c_{h,q}(L)^2}{n} \int_{\Omega_q} L^2\left(\frac{1 - \mathbf{x}^T \mathbf{y}}{h^2}\right) \left(\mathbf{B}_{\mathbf{x}}^T (\mathbf{y} - \mathbf{x}) (\mathbf{y} - \mathbf{x})^T \mathbf{B}_{\mathbf{x}} \right)^2 f(\mathbf{y}) \, \omega_q(d\mathbf{y}) \\
&\quad = \frac{c_{h,q}(L)^2}{n} \int_{-1}^{1} \int_{\Omega_{q-1}} L^2\left(\frac{1-t}{h^2}\right) \left(\boldsymbol{\xi}\boldsymbol{\xi}^T \right)^2 f\left(t\mathbf{x} + (1-t^2)^{\frac{1}{2}} \mathbf{B}_{\mathbf{x}} \boldsymbol{\xi}\right) (1-t^2)^{\frac{q}{2}+1} \, \omega_{q-1}(d\boldsymbol{\xi}) \, dt \\
&\quad = \frac{c_{h,q}(L)^2 h^{q+4}}{n} \int_{0}^{2h^{-2}} L^2(r) \, r^{\frac{q}{2}+1} (2 - rh^2)^{\frac{q}{2}+1} \int_{\Omega_{q-1}} f\left(\mathbf{x} + \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right) \left(\boldsymbol{\xi}\boldsymbol{\xi}^T \right)^2 \omega_{q-1}(d\boldsymbol{\xi}) \, dt \\
&\quad = \mathcal{O}\left(\frac{h^4}{n h^q} \mathbf{1}\mathbf{1}^T \right). \qquad \text{(B.9)}
\end{aligned}
$$

The desired result now holds by (B.8) and (B.9), as $\mathcal{O}_{\mathbb{P}}\left(\frac{h^2}{\sqrt{nh^q}} \mathbf{1}\mathbf{1}^T\right) = o_{\mathbb{P}}(h^2 \mathbf{1}\mathbf{1}^T)$ by **A4**.

Proof of v. This is one of the most important results since it determines the dominant term of the bias of the local projected estimator. The expectation is

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} L_h(\mathbf{x}, \mathbf{X}_i) (\mathbf{X}_i - \mathbf{x})^T \mathbf{B}_{\mathbf{x}} \mathbf{B}_{\mathbf{x}}^T \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \mathbf{B}_{\mathbf{x}} \mathbf{B}_{\mathbf{x}}^T (\mathbf{X}_i - \mathbf{x}) \right] \\
&\quad = c_{h,q}(L) \int_{\Omega_q} L\left(\frac{1 - \mathbf{x}^T \mathbf{y}}{h^2}\right) (\mathbf{y} - \mathbf{x})^T \mathbf{B}_{\mathbf{x}} \mathbf{B}_{\mathbf{x}}^T \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \mathbf{B}_{\mathbf{x}} \mathbf{B}_{\mathbf{x}}^T (\mathbf{y} - \mathbf{x}) f(\mathbf{y}) \, \omega_q(d\mathbf{y})
\end{aligned}
$$

$$
\begin{aligned}
&= c_{h,q}(L) \int_{-1}^{1} \int_{\Omega_{q-1}} L\left(\frac{1-t}{h^2}\right) (\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi} f\left(t\mathbf{x} + (1-t^2)^{\frac{1}{2}} \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\right) (1-t^2)^{\frac{q}{2}} \\
&\quad \times \omega_{q-1}(d\boldsymbol{\xi})\, dt \\
&= c_{h,q}(L) h^{q+2} \int_0^{2h^{-2}} L(r)\, r^{\frac{q}{2}} (2-rh^2)^{\frac{q}{2}} \int_{\Omega_{q-1}} f(\mathbf{x} + \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}) (\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi} \\
&\quad \times \omega_{q-1}(d\boldsymbol{\xi})\, dr. \qquad \text{(B.10)}
\end{aligned}
$$

The first addend of the Taylor expansion (B.4) is computed using Lemma B.5 and the following relation of the trace operator:

$$
\mathbf{x}^T \mathbf{A} \mathbf{x} = \operatorname{tr}\left[\mathbf{x}^T \mathbf{A} \mathbf{x}\right] = \operatorname{tr}\left[\mathbf{x}\mathbf{x}^T \mathbf{A}\right], \quad \text{for } \mathbf{x} \text{ a vector and } \mathbf{A} \text{ a matrix.}
$$

Recall also that by definition of $\mathbf{B}_{\mathbf{x}} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$, $\sum_{i=1}^{q} \mathbf{b}_i \mathbf{b}_i^T = \mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T$ and $\mathbf{x}^T \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \mathbf{x} = 0$ by **A1**. Then:

$$
\begin{aligned}
\int_{\Omega_{q-1}} f(\mathbf{x}) (\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\, \omega_{q-1}(d\boldsymbol{\xi}) &= f(\mathbf{x}) \int_{\Omega_{q-1}} \sum_{i,j=1}^{q} \xi_i \xi_j \mathbf{b}_i^T \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \mathbf{b}_j\, \omega_{q-1}(d\boldsymbol{\xi}) \\
&= f(\mathbf{x}) \frac{\omega_{q-1}}{q} \sum_{i=1}^{q} \mathbf{b}_i^T \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \mathbf{b}_i \\
&= f(\mathbf{x}) \frac{\omega_{q-1}}{q} \operatorname{tr}\left[\sum_{i=1}^{q} \mathbf{b}_i \mathbf{b}_i^T \boldsymbol{\mathcal{H}}_m(\mathbf{x})\right] \\
&= f(\mathbf{x}) \frac{\omega_{q-1}}{q} \operatorname{tr}\left[\boldsymbol{\mathcal{H}}_m(\mathbf{x})\right].
\end{aligned}
$$

The second and third addends have the same orders as in *iv*, so

$$
\begin{aligned}
\text{(B.10)} &= c_{h,q}(L) h^{q+2} \int_0^{2h^{-2}} L(r)\, r^{\frac{q}{2}} (2-rh^2)^{\frac{q}{2}} \left\{ \frac{\omega_{q-1}}{q} \operatorname{tr}\left[\boldsymbol{\mathcal{H}}_m(\mathbf{x})\right] f(\mathbf{x}) + \mathcal{O}\left(h^2\right) \right\} dr \\
&= \frac{2b_q(L)}{q} \operatorname{tr}\left[\boldsymbol{\mathcal{H}}_m(\mathbf{x})\right] f(\mathbf{x}) h^2 + o\left(h^2\right).
\end{aligned}
$$

Using the square notation for vectors, the order of the variance is

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{i=1}^{n} L_h(\mathbf{x}, \mathbf{X}_i)(\mathbf{X}_i - \mathbf{x})^T \mathbf{B}_{\mathbf{x}} \mathbf{B}_{\mathbf{x}}^T \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \mathbf{B}_{\mathbf{x}} \mathbf{B}_{\mathbf{x}}^T (\mathbf{X}_i - \mathbf{x})\right] \\
&\quad \leq \frac{c_{h,q}(L)^2}{n} \int_{\Omega_q} L^2\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) \left((\mathbf{y}-\mathbf{x})^T \mathbf{B}_{\mathbf{x}} \mathbf{B}_{\mathbf{x}}^T \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \mathbf{B}_{\mathbf{x}} \mathbf{B}_{\mathbf{x}}^T (\mathbf{y}-\mathbf{x})\right)^2 f(\mathbf{y})\, \omega_q(d\mathbf{y}) \\
&\quad = \frac{c_{h,q}(L)^2 h^{q+4}}{n} \int_0^{2h^{-2}} L^2(r)\, r^{\frac{q}{2}+1} (2-rh^2)^{\frac{q}{2}+1} \\
&\qquad \times \int_{\Omega_{q-1}} f(\mathbf{x} + \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}) \left((\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\right)^2 \omega_{q-1}(d\boldsymbol{\xi})\, dr \\
&\quad = \frac{c_{h,q}(L)^2 h^{q+4}}{n} \int_0^{2h^{-2}} L^2(r)\, r^{\frac{q}{2}+1} (2-rh^2)^{\frac{q}{2}+1} \mathcal{O}(1)\, dr \\
&\quad = \mathcal{O}\left(\frac{h^4}{nh^q}\right)
\end{aligned}
$$

and the square root of this order can be merged into $o_{\mathbb{P}}\left(h^2\right)$.

Proof of vi. Similarly to the previous proofs, the order of the bias is

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\mathcal{H}}_m(\mathbf{x})\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})\right] \\
&\qquad = c_{h,q}(L)h^{q+3}\int_0^{2h^{-2}} L\left(r\right) r^{\frac{q+1}{2}}\left(2-rh^2\right)^{\frac{q+1}{2}} \\
&\qquad\quad \times \int_{\Omega_{q-1}} f\left(\mathbf{x}+\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right)\boldsymbol{\xi}\boldsymbol{\xi}^T\boldsymbol{\mathcal{H}}_m(\mathbf{x})\boldsymbol{\xi}\,\omega_{q-1}(d\boldsymbol{\xi})\,dr \\
&\qquad = \mathcal{O}\left(h^4\mathbf{1}\right)
\end{aligned}$$

because by Lemma B.5 the first element in the Taylor expansion of the inner integral is exactly zero. The variance is

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{i=1}^{n} L_h(\mathbf{x},\mathbf{X}_i)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^T\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T\boldsymbol{\mathcal{H}}_m(\mathbf{x})\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})\right] \\
&\qquad \leq \frac{c_{h,q}(L)^2h^{q+6}}{n}\int_0^{2h^{-2}} L^2\left(r\right) r^{\frac{q}{2}+2}\left(2-rh^2\right)^{\frac{q}{2}+2} \\
&\qquad\quad \times \int_{\Omega_{q-1}} f\left(\mathbf{x}+\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right)\left(\boldsymbol{\xi}\boldsymbol{\xi}^T\boldsymbol{\mathcal{H}}_m(\mathbf{x})\boldsymbol{\xi}\right)^2\omega_{q-1}(d\boldsymbol{\xi})\,dr \\
&\qquad = \mathcal{O}\left(\frac{h^6}{nh^q}\mathbf{1}\right).
\end{aligned}$$

Since $\mathcal{O}\left(h^4\mathbf{1}\right)+\mathcal{O}_{\mathbb{P}}\left(\frac{h^3}{\sqrt{nh^q}}\mathbf{1}\right)=o_{\mathbb{P}}\left(h^3\mathbf{1}\right)$ the result is proved.

Proof of vii. Because of Lemma B.6 and (7.2):

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} L_h^2\left(\mathbf{x},\mathbf{X}_i\right)\sigma^2(\mathbf{X}_i)\right] &= c_{h,q}(L)^2\int_{\Omega_q} L^2\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right)\sigma^2(\mathbf{y})f(\mathbf{y})\,\omega_q(d\mathbf{y}) \\
&= \frac{c_{h,q}(L)^2}{c_{h,q}(L^2)}\left[\sigma^2(\mathbf{x})f(\mathbf{x})+o\left(1\right)\right] \\
&= \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{h^q}\sigma^2(\mathbf{x})f(\mathbf{x})+o\left(h^{-q}\right), \\
\mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{i=1}^{n} L_h^2\left(\mathbf{x},\mathbf{X}_i\right)\sigma^2(\mathbf{X}_i)\right] &\leq \frac{1}{n}\mathbb{E}\left[L_h^4\left(\mathbf{x},\mathbf{X}\right)\sigma^4(\mathbf{X})\right] \\
&= \frac{c_{h,q}(L)^4}{n}\int_{\Omega_q} L^4\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right)\sigma^4(\mathbf{y})f(\mathbf{y})\,\omega_q(d\mathbf{y}) \\
&= \frac{c_{h,q}(L)^4}{nc_{h,q}(L^4)}\left[\sigma^4(\mathbf{x})f(\mathbf{x})+o\left(1\right)\right] \\
&= \mathcal{O}\left((nh^{3q})^{-1}\right).
\end{aligned}$$

The remaining order is $o_{\mathbb{P}}\left(h^{-q}\right)$ because $\frac{(nh^{3q})^{-\frac{1}{2}}}{h^{-q}}=(nh^q)^{-\frac{1}{2}}\to 0$ by **A4**.

Proof of viii. By (7.2) and Lemma B.6 applied componentwise, since the functions in the integrand are vector valued, it follows that

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} L_h^2\left(\mathbf{x},\mathbf{X}_i\right)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})\sigma^2(\mathbf{X}_i)\right]\\
&\qquad = c_{h,q}(L)^2\int_{\Omega_q} L^2\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right)\mathbf{B}_{\mathbf{x}}^T(\mathbf{y}-\mathbf{x})\sigma^2(\mathbf{y})f(\mathbf{y})\,\omega_q(d\mathbf{y})\\
&\qquad = \frac{c_{h,q}(L)^2}{c_{h,q}(L^2)}\left(\mathbf{0}+o\left(\mathbf{1}\right)\right)\\
&\qquad = o\left(h^{-q}\mathbf{1}\right),\\
&\mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{i=1}^{n} L_h^2\left(\mathbf{x},\mathbf{X}_i\right)\mathbf{B}_{\mathbf{x}}^T(\mathbf{X}_i-\mathbf{x})\sigma^2(\mathbf{X}_i)\right]\\
&\qquad \leq c_{h,q}(L)^4\int_{\Omega_q} L^4\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right)\left(\mathbf{B}_{\mathbf{x}}^T(\mathbf{y}-\mathbf{x})\right)^2\sigma^4(\mathbf{y})f(\mathbf{y})\,\omega_q(d\mathbf{y})\\
&\qquad = \frac{c_{h,q}(L)^4}{nc_{h,q}(L^4)}\left(\mathbf{0}+o\left(\mathbf{1}\right)\right)\\
&\qquad = o\left((nh^{3q})^{-1}\right).
\end{aligned}$$

Proof of ix. The proof is analogous to *viii*: using (7.2) and Lemma B.6 componentwise the statement is proved trivially. □

B.2.2 Asymptotic results for the goodness-of-fit test

Lemma B.2. *Under assumptions **A1**–**A4** and **A7**, for a random sample $\{(\mathbf{X}_i,Y_i)\}_{i=1}^n$ the following statements hold:*

i. $\int_{\Omega_q}\left(\hat{m}_{h,p}(\mathbf{x})-\mathcal{L}_{h,p}m(\mathbf{x})\right)f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}) = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$.

ii. $\int_{\Omega_q}\left(\sum_{i=1}^n L_h^*\left(\mathbf{x},\mathbf{X}_i\right)g(\mathbf{X}_i)\right)^2 f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}) = \int_{\Omega_q} g(\mathbf{x})^2 f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})(1+o\left(1\right))$ $+\mathcal{O}_{\mathbb{P}}\left((nh^q)^{-1}+n^{-\frac{1}{2}}\right)$.

iii. $\int_{\Omega_q}\sum_{i=1}^n\sum_{j=1}^n L_h^*\left(\mathbf{x},\mathbf{X}_i\right)L_h^*\left(\mathbf{x},\mathbf{X}_j\right)\sigma(\mathbf{X}_i)\varepsilon_i g(\mathbf{X}_j)f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}) = \mathcal{O}_{\mathbb{P}}\left((nh^{\frac{q}{2}})^{-1}\right)$.

iv. $\int_{\Omega_q}\sum_{i=1}^n\left(L_h^*\left(\mathbf{x},\mathbf{X}_i\right)\sigma(\mathbf{X}_i)\varepsilon_i\right)^2 f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}) = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})$ $\times(1+o\left(1\right))+\mathcal{O}_{\mathbb{P}}\left((n^{\frac{3}{2}}h^q)^{-1}\right)$.

v. $\mathbb{E}[W_{ijn}^2] = n^{-2}\nu^2\left(1+o\left(1\right)\right)$, $\mathbb{E}\left[W_{ijn}W_{jkn}W_{kln}W_{lin}\right] = \mathcal{O}\left(n^{-4}h^{2q}\right)$, $\mathbb{E}[W_{ijn}^4] = \mathcal{O}\left((n^4h^q)^{-1}\right)$, $\mathbb{E}\left[W_{ijn}W_{ikn}^2W_{jkn}\right] = \mathcal{O}\left(n^{-4}\right)$, *where $\nu^2\equiv\nu_{\boldsymbol{\theta}_0}^2$ is given in Theorem 7.3.*

Proof of Lemma B.2. The proof is divided for each statement. As in Lemma B.1, Chebychev's inequality and Lemma B.6 are used repeatedly.

Proof of i. By Corollary 7.1,

$$\int_{\Omega_q}\left(\hat{m}_{h,p}(\mathbf{x})-\mathcal{L}_{h,p}m(\mathbf{x})\right)f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})$$

$$= \int_{\Omega_q} \sum_{i=1}^{n} L_h^*(\mathbf{x}, \mathbf{X}_i)(Y_i - m(\mathbf{X}_i)) f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \left(1 + o_{\mathbb{P}}(1)\right).$$

Using the properties of the conditional expectation, Fubini, relation (7.2) and Lemma B.6:

$$\begin{aligned}
&\mathbb{E}\left[\int_{\Omega_q} \sum_{i=1}^{n} L_h^*(\mathbf{x}, \mathbf{X}_i)(Y_i - m(\mathbf{X}_i)) f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x})\right] \\
&\quad = \int_{\Omega_q} \mathbb{E}\left[L_h^*(\mathbf{x}, \mathbf{X}) \mathbb{E}\left[Y - m(\mathbf{X}) | \mathbf{X}\right]\right] f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \\
&\quad = 0, \\
&\mathbb{V}\mathrm{ar}\left[\int_{\Omega_q} \sum_{i=1}^{n} L_h^*(\mathbf{x}, \mathbf{X}_i)(Y_i - m(\mathbf{X}_i)) f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x})\right] \\
&\quad = \frac{1}{nh^q \lambda_q(L)} \int_{\Omega_q} \int_{\Omega_q} L\left(\frac{1 - \mathbf{y}^T \mathbf{x}}{h^2}\right) \sigma^2(\mathbf{x})\, f(\mathbf{x}) w(\mathbf{x}) w(\mathbf{y})\, \omega_q(d\mathbf{x})\, \omega_q(d\mathbf{y})\, (1 + o(1)) \\
&\quad = \frac{1}{n} \int_{\Omega_q} \sigma^2(\mathbf{y})\, f(\mathbf{y}) w(\mathbf{y})^2\, \omega_q(d\mathbf{y})\, (1 + o(1)) \\
&\quad = \mathcal{O}\left(n^{-1}\right).
\end{aligned}$$

Then $\int_{\Omega_q} (\hat{m}_{h,p}(\mathbf{x}) - \mathcal{L}_{h,p} m(\mathbf{x})) f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})(1 + o_{\mathbb{P}}(1)) = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$.

Proof of ii. The integral can be split in two addends:

$$\begin{aligned}
&\int_{\Omega_q} \left(\sum_{i=1}^{n} L_h^*(\mathbf{x}, \mathbf{X}_i)\, g(\mathbf{X}_i)\right)^2 f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \\
&\quad = \frac{1}{n^2 h^{2q} \lambda_q(L)^2} \sum_{i=1}^{n} \int_{\Omega_q} L^2\left(\frac{1 - \mathbf{x}^T \mathbf{X}_i}{h^2}\right) \frac{g(\mathbf{X}_i)^2 w(\mathbf{x})}{f(\mathbf{x})}\, \omega_q(d\mathbf{x}) \\
&\qquad + \frac{1}{n^2 h^{2q} \lambda_q(L)^2} \sum_{i \neq j} \int_{\Omega_q} L\left(\frac{1 - \mathbf{x}^T \mathbf{X}_i}{h^2}\right) L\left(\frac{1 - \mathbf{x}^T \mathbf{X}_j}{h^2}\right) \frac{g(\mathbf{X}_i) g(\mathbf{X}_j) w(\mathbf{x})}{f(\mathbf{x})}\, \omega_q(d\mathbf{x}) \\
&\quad = I_1 + I_2.
\end{aligned}$$

Now, by applying Fubini, (7.2) and Lemma B.6,

$$\begin{aligned}
\mathbb{E}[I_1] &= \frac{1}{nh^{2q} \lambda_q(L)^2} \int_{\Omega_q} \mathbb{E}\left[L^2\left(\frac{1 - \mathbf{x}^T \mathbf{X}}{h^2}\right) g(\mathbf{X})^2\right] \frac{w(\mathbf{x})}{f(\mathbf{x})}\, \omega_q(d\mathbf{x}) \\
&= \frac{1}{nh^{2q} \lambda_q(L)^2} \int_{\Omega_q} \int_{\Omega_q} L^2\left(\frac{1 - \mathbf{x}^T \mathbf{y}}{h^2}\right) \frac{g(\mathbf{y})^2 w(\mathbf{x})}{f(\mathbf{x})} f(\mathbf{y})\, \omega_q(d\mathbf{y})\, \omega_q(d\mathbf{x}) \\
&= \frac{\lambda_q(L^2)}{nh^q \lambda_q(L)^2} \int_{\Omega_q} g(\mathbf{x})^2 w(\mathbf{x})\, \omega_q(d\mathbf{x}) (1 + o(1)) \\
&= \mathcal{O}\left((nh^q)^{-1}\right), \\
\mathbb{V}\mathrm{ar}[I_1] &\leq \frac{1}{n^3 h^{4q} \lambda_q(L)^4} \mathbb{E}\left[\left(\int_{\Omega_q} L^2\left(\frac{1 - \mathbf{x}^T \mathbf{X}}{h^2}\right) \frac{g(\mathbf{X})^2 w(\mathbf{x})}{f(\mathbf{x})}\, \omega_q(d\mathbf{x})\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\lambda_q (L^2)^2}{n^3 h^{2q} \lambda_q(L)^4} \int_{\Omega_q} \frac{g(\mathbf{y})^2 w(\mathbf{y})^2}{f(\mathbf{y})} \, \omega_q(d\mathbf{y})(1 + o(1)) \\
&= \mathcal{O}\left((n^3 h^{2q})^{-1}\right)
\end{aligned}
$$

and therefore $I_1 = \mathcal{O}_{\mathbb{P}}\left((nh^q)^{-1}\right)$. On the other hand, by Lemma B.6 and the independence of $\mathbf{X}_i$ and $\mathbf{X}_j$ if $i \neq j$:

$$
\begin{aligned}
\mathbb{E}[I_2] &= \frac{1 - n^{-1}}{h^{2q} \lambda_q(L)^2} \int_{\Omega_q} \mathbb{E}\left[L\left(\frac{1 - \mathbf{x}^T \mathbf{X}}{h^2} \right) g(\mathbf{X}) \right]^2 \frac{w(\mathbf{x})}{f(\mathbf{x})} \, \omega_q(d\mathbf{x}) \\
&= \left(1 - n^{-1}\right) \int_{\Omega_q} g(\mathbf{x})^2 f(\mathbf{x}) w(\mathbf{x}) \, \omega_q(d\mathbf{x})(1 + o(1)) \\
&= \int_{\Omega_q} g(\mathbf{x})^2 f(\mathbf{x}) w(\mathbf{x}) \, \omega_q(d\mathbf{x})(1 + o(1)), \\
\mathbb{E}\left[I_2^2\right] &= \frac{1}{n^4 h^{4q} \lambda_q(L)^4} \sum_{i \neq j} \sum_{k \neq l} \int_{\Omega_q} \int_{\Omega_q} \mathbb{E}\left[L\left(\frac{1 - \mathbf{x}^T \mathbf{X}_i}{h^2} \right) L\left(\frac{1 - \mathbf{x}^T \mathbf{X}_j}{h^2} \right) L\left(\frac{1 - \mathbf{y}^T \mathbf{X}_k}{h^2} \right) \right. \\
&\quad \left. \times L\left(\frac{1 - \mathbf{y}^T \mathbf{X}_l}{h^2} \right) g(\mathbf{X}_i) g(\mathbf{X}_j) g(\mathbf{X}_k) g(\mathbf{X}_l) \right] \frac{w(\mathbf{x}) w(\mathbf{y})}{f(\mathbf{x}) f(\mathbf{y})} \, \omega_q(d\mathbf{x}) \, \omega_q(d\mathbf{y}) \\
&= \mathcal{O}\left((n^2 h^{4q})^{-1}\right) \int_{\Omega_q} \int_{\Omega_q} \mathbb{E}\left[L\left(\frac{1 - \mathbf{x}^T \mathbf{X}}{h^2} \right) L\left(\frac{1 - \mathbf{y}^T \mathbf{X}}{h^2} \right) g(\mathbf{X})^2 \right]^2 \\
&\quad \times \frac{w(\mathbf{x}) w(\mathbf{y})}{f(\mathbf{x}) f(\mathbf{y})} \, \omega_q(d\mathbf{x}) \, \omega_q(d\mathbf{y}) \\
&\quad + \mathcal{O}\left(nh^{-4q}\right) \int_{\Omega_q} \int_{\Omega_q} \mathbb{E}\left[L\left(\frac{1 - \mathbf{x}^T \mathbf{X}}{h^2} \right) L\left(\frac{1 - \mathbf{y}^T \mathbf{X}}{h^2} \right) g(\mathbf{X})^2 \right] \\
&\quad \times \mathbb{E}\left[L\left(\frac{1 - \mathbf{x}^T \mathbf{X}}{h^2} \right) g(\mathbf{X}) \right] \mathbb{E}\left[L\left(\frac{1 - \mathbf{y}^T \mathbf{X}}{h^2} \right) g(\mathbf{X}) \right] \frac{w(\mathbf{x}) w(\mathbf{y})}{f(\mathbf{x}) f(\mathbf{y})} \, \omega_q(d\mathbf{x}) \, \omega_q(d\mathbf{y}) \\
&\quad + \frac{n^4 - \mathcal{O}(n^3)}{n^4 h^{4q} \lambda_q(L)^4} \left(\int_{\Omega_q} \mathbb{E}\left[L\left(\frac{1 - \mathbf{x}^T \mathbf{X}}{h^2} \right) g(\mathbf{X}) \right]^2 \frac{w(\mathbf{x})}{f(\mathbf{x})} \, \omega_q(d\mathbf{x}) \right)^2 \\
&= \mathcal{O}\left((n^2 h^{2q})^{-1}\right) \int_{\Omega_q} \int_{\Omega_q} L^2\left(\frac{1 - \mathbf{y}^T \mathbf{x}}{h^2} \right) g(\mathbf{x})^4 f(\mathbf{x}) \frac{w(\mathbf{x}) w(\mathbf{y})}{f(\mathbf{y})} \, \omega_q(d\mathbf{x}) \, \omega_q(d\mathbf{y}) \\
&\quad + \mathcal{O}\left((nh^q)^{-1}\right) \int_{\Omega_q} \int_{\Omega_q} L\left(\frac{1 - \mathbf{y}^T \mathbf{x}}{h^2} \right) g(\mathbf{x})^3 g(\mathbf{y}) f(\mathbf{x}) w(\mathbf{x}) w(\mathbf{y}) \, \omega_q(d\mathbf{x}) \, \omega_q(d\mathbf{y}) \\
&\quad + \left(1 - \mathcal{O}\left(n^{-1}\right)\right) \mathbb{E}[I_2]^2 \\
&= \mathcal{O}\left((n^2 h^q)^{-1}\right) + \mathcal{O}\left(n^{-1}\right) + \left(1 - \mathcal{O}\left(n^{-1}\right)\right) \mathbb{E}[I_2]^2 .
\end{aligned}
$$

Then $\mathbb{V}\mathrm{ar}[I_2] = \mathbb{E}[I_2^2] - \mathbb{E}[I_2]^2 = \mathcal{O}(n^{-1})$ and $I_2 = \int_{\Omega_q} g(\mathbf{x})^2 f(\mathbf{x}) w(\mathbf{x}) \, \omega_q(d\mathbf{x})(1 + o(1)) + \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$. Finally,

$$
I_1 + I_2 = \int_{\Omega_q} g(\mathbf{x})^2 f(\mathbf{x}) w(\mathbf{x}) \, \omega_q(d\mathbf{x})(1 + o(1)) + \mathcal{O}_{\mathbb{P}}\left((nh^q)^{-1} + n^{-\frac{1}{2}}\right).
$$

Proof of iii. By the tower property of the conditional expectation and $\mathbb{E}\left[\varepsilon|\mathbf{X}\right]=0$, the expectation is zero. By the independece between ε's and $\mathbb{E}\left[\varepsilon^2|\mathbf{X}\right]=1$, the variance is

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}\left[\int_{\Omega_q}\sum_{i=1}^{n}\sum_{j=1}^{n}L_h^*\left(\mathbf{x},\mathbf{X}_i\right)L_h^*\left(\mathbf{x},\mathbf{X}_j\right)\varepsilon_i g(\mathbf{X}_j)f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right]\\
&\quad=\frac{1}{n^4h^{4q}\lambda_q(L)^4}\sum_{i,j,k,l=1}^{n}\int_{\Omega_q}\int_{\Omega_q}\mathbb{E}\Bigg[L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right)L\left(\frac{1-\mathbf{x}^T\mathbf{X}_j}{h^2}\right)L\left(\frac{1-\mathbf{y}^T\mathbf{X}_k}{h^2}\right)\\
&\qquad\times L\left(\frac{1-\mathbf{y}^T\mathbf{X}_l}{h^2}\right)\mathbb{E}\left[\varepsilon_i\varepsilon_k|\,\mathbf{X}_i,\mathbf{X}_k\right]g(\mathbf{X}_j)g(\mathbf{X}_l)\Bigg]\frac{w(\mathbf{x})w(\mathbf{y})}{f(\mathbf{x})f(\mathbf{y})}\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y})\\
&\quad=\frac{1}{n^4h^{4q}\lambda_q(L)^4}\sum_{i,j,l=1}^{n}\int_{\Omega_q}\int_{\Omega_q}\mathbb{E}\Bigg[L\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right)L\left(\frac{1-\mathbf{x}^T\mathbf{X}_j}{h^2}\right)L\left(\frac{1-\mathbf{y}^T\mathbf{X}_i}{h^2}\right)\\
&\qquad\times L\left(\frac{1-\mathbf{y}^T\mathbf{X}_l}{h^2}\right)g(\mathbf{X}_j)g(\mathbf{X}_l)\Bigg]\frac{w(\mathbf{x})w(\mathbf{y})}{f(\mathbf{x})f(\mathbf{y})}\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y})\\
&\quad=\frac{1}{n^4h^{4q}\lambda_q(L)^4}\left\{I_1+I_2+I_3+I_4\right\},
\end{aligned}$$

where, by repeated use of Lemma B.6:

$$\begin{aligned}
I_1&=n\int_{\Omega_q}\int_{\Omega_q}\mathbb{E}\left[L^2\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)L^2\left(\frac{1-\mathbf{y}^T\mathbf{X}}{h^2}\right)g(\mathbf{X})^2\right]\frac{w(\mathbf{x})w(\mathbf{y})}{f(\mathbf{x})f(\mathbf{y})}\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y})\\
&=\mathcal{O}\left(nh^{2q}\right),\\
I_2&=\mathcal{O}\left(n^2\right)\int_{\Omega_q}\int_{\Omega_q}\mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)L\left(\frac{1-\mathbf{y}^T\mathbf{X}}{h^2}\right)\right]\\
&\quad\times\mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)L\left(\frac{1-\mathbf{y}^T\mathbf{X}}{h^2}\right)g(\mathbf{X})\right]\frac{w(\mathbf{x})w(\mathbf{y})}{f(\mathbf{x})f(\mathbf{y})}\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y})\\
&=\mathcal{O}\left(n^2h^{4q}\right),\\
I_3&=\mathcal{O}\left(n^2\right)\int_{\Omega_q}\int_{\Omega_q}\mathbb{E}\left[L^2\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)L\left(\frac{1-\mathbf{y}^T\mathbf{X}}{h^2}\right)g(\mathbf{X})\right]\mathbb{E}\left[L\left(\frac{1-\mathbf{y}^T\mathbf{X}}{h^2}\right)g(\mathbf{X})\right]\\
&\quad\times\frac{w(\mathbf{x})w(\mathbf{y})}{f(\mathbf{x})f(\mathbf{y})}\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y})\\
&=\mathcal{O}\left(n^2h^{3q}\right),\\
I_4&=\mathcal{O}\left(n^3\right)\int_{\Omega_q}\int_{\Omega_q}\mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right)L\left(\frac{1-\mathbf{y}^T\mathbf{X}}{h^2}\right)\right]\mathbb{E}\left[L\left(\frac{1-\mathbf{y}^T\mathbf{X}}{h^2}\right)g(\mathbf{X})\right]^2\\
&\quad\times\frac{w(\mathbf{x})w(\mathbf{y})}{f(\mathbf{x})f(\mathbf{y})}\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y})\\
&=\mathcal{O}\left(n^3h^{4q}\right).
\end{aligned}$$

Because $\mathcal{O}\left((n^3h^{2q})^{-1}+n^{-2}+(n^2h^q)^{-1}+n^{-1}\right)=\mathcal{O}\left((n^2h^q)^{-1}\right)$ by **A4**, we have that

$$\int_{\Omega_q}\sum_{i=1}^{n}\sum_{j=1}^{n}L_h^*\left(\mathbf{x},\mathbf{X}_i\right)L_h^*\left(\mathbf{x},\mathbf{X}_j\right)\sigma(\mathbf{X}_i)\varepsilon_i g(\mathbf{X}_j)f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})=\mathcal{O}_{\mathbb{P}}\Big(\left(nh^{\frac{q}{2}}\right)^{-1}\Big).$$

Proof of iv. Let us denote $I = \int_{\Omega_q} \sum_{i=1}^n \left(L_h^*\left(\mathbf{x}, \mathbf{X}_i\right) \sigma(\mathbf{X}_i)\varepsilon_i\right)^2 f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x})$. By the unit conditional variance of ε and the boundedness of $\mathbb{E}\left[\varepsilon^4|\mathbf{X}\right]$,

$$
\begin{aligned}
\mathbb{E}[I] &= \sum_{i=1}^n \int_{\Omega_q} \mathbb{E}\left[\left(L_h^*\left(\mathbf{x}, \mathbf{X}_i\right) \sigma(\mathbf{X}_i)\right)^2 \mathbb{E}\left[\varepsilon_i^2|\mathbf{X}_i\right]\right] f(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \\
&= \frac{1}{nh^{2q}\lambda_q(L)^2} \int_{\Omega_q}\int_{\Omega_q} L^2\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) \frac{\sigma^2(\mathbf{y}) w(\mathbf{x})}{f(\mathbf{x})} f(\mathbf{y})\, \omega_q(d\mathbf{y})\, \omega_q(d\mathbf{x}) \\
&= \frac{1}{nh^{2q}\lambda_q(L)^2} \int_{\Omega_q} \frac{1}{c_{h,q}(L^2)} \sigma^2(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x})(1+o(1)) \\
&= \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q} \int_{\Omega_q} \sigma^2(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x})(1+o(1)), \\
\mathbb{E}\left[I^2\right] &= \sum_{i=1}^n \sum_{j=1}^n \int_{\Omega_q}\int_{\Omega_q} \mathbb{E}\left[\left(L_h^*\left(\mathbf{x}, \mathbf{X}_i\right) \sigma(\mathbf{X}_i) L_h^*\left(\mathbf{y}, \mathbf{X}_j\right) \sigma(\mathbf{X}_j)\right)^2 \mathbb{E}\left[\varepsilon_i^2\varepsilon_j^2|\mathbf{X}_i, \mathbf{X}_j\right]\right] \\
&\quad \times f(\mathbf{x}) f(\mathbf{y}) w(\mathbf{x}) w(\mathbf{y})\, \omega_q(d\mathbf{x})\, \omega_q(d\mathbf{y}) \\
&= \frac{1}{n^3 h^{4q}\lambda_q(L)^4} \int_{\Omega_q}\int_{\Omega_q} \mathbb{E}\left[L^2\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right) L^2\left(\frac{1-\mathbf{y}^T\mathbf{X}}{h^2}\right) \sigma^4(\mathbf{X}) \mathbb{E}\left[\varepsilon^4|\mathbf{X}\right]\right] \\
&\quad \times \frac{w(\mathbf{x}) w(\mathbf{y})}{f(\mathbf{x}) f(\mathbf{y})}\, \omega_q(d\mathbf{x})\, \omega_q(d\mathbf{y}) \\
&\quad + \frac{1-n^{-1}}{n^2 h^{4q}\lambda_q(L)^4} \left(\int_{\Omega_q}\int_{\Omega_q} L^2\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) \frac{\sigma^2(\mathbf{y}) w(\mathbf{x})}{f(\mathbf{x})} f(\mathbf{y})\, \omega_q(d\mathbf{x})\, \omega_q(d\mathbf{y})\right)^2 \\
&= \mathcal{O}\left((n^3 h^{2q})^{-1}\right) \int_{\Omega_q} \frac{\sigma^4(\mathbf{x}) w(\mathbf{x})^2}{f(\mathbf{x})}\, \omega_q(d\mathbf{x}) + \left(1-\mathcal{O}\left(n^{-1}\right)\right) \mathbb{E}\left[I\right]^2 \\
&= \mathcal{O}\left((n^3 h^{2q})^{-1}\right) + \left(1-\mathcal{O}\left(n^{-1}\right)\right) \mathbb{E}\left[I\right]^2.
\end{aligned}
$$

Then $\mathbb{V}\mathrm{ar}\left[I\right] = \mathcal{O}\left((n^3 h^{2q})^{-1}\right) - \mathcal{O}\left(n^{-1}\right) \mathbb{E}\left[I\right]^2 = \mathcal{O}\left((n^3 h^{2q})^{-1}\right)$ and as a consequence

$$
I = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q} \int_{\Omega_q} \sigma^2(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x})(1+o(1)) + \mathcal{O}_{\mathbb{P}}\left(\left(n^{\frac{3}{2}} h^q\right)^{-1}\right).
$$

Proof of v. The computation of

$$
\begin{aligned}
\mathbb{E}\left[W_{ijn}^2\right] &= \frac{n^2 h^q}{n^4 h^{4q}\lambda_q(L)^4} \int_{\Omega_q}\int_{\Omega_q} \mathbb{E}\left[L\left(\frac{1-\mathbf{x}^T\mathbf{X}}{h^2}\right) L\left(\frac{1-\mathbf{y}^T\mathbf{X}}{h^2}\right) \sigma^2(\mathbf{X}) \mathbb{E}\left[\varepsilon^2|\mathbf{X}\right]\right]^2 \\
&\quad \times \frac{w(\mathbf{x}) w(\mathbf{y})}{f(\mathbf{x}) f(\mathbf{y})}\, \omega_q(d\mathbf{x})\, \omega_q(d\mathbf{y}) \\
&= \frac{1}{n^2 h^{3q}\lambda_q(L)^4} \int_{\Omega_q}\int_{\Omega_q} \left[\int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{z}}{h^2}\right) L\left(\frac{1-\mathbf{y}^T\mathbf{z}}{h^2}\right) \sigma^2(\mathbf{z}) f(\mathbf{z})\, \omega_q(d\mathbf{z})\right]^2 \\
&\quad \times \frac{w(\mathbf{x}) w(\mathbf{y})}{f(\mathbf{x}) f(\mathbf{y})}\, \omega_q(d\mathbf{x})\, \omega_q(d\mathbf{y})
\end{aligned} \tag{B.11}
$$

is split in the cases where $q \geq 2$ and $q = 1$. For the first one, the usual change of variables given by Lemma B.4 is applied:

$$
\mathbf{y} = s\mathbf{x} + (1-s^2)^{\frac{1}{2}} \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}, \quad \omega_q(d\mathbf{y}) = (1-s^2)^{\frac{q}{2}-1}\, \omega_{q-1}(d\boldsymbol{\xi})\, ds. \tag{B.12}
$$

Because $q \geq 2$, it is possible also to consider an extra change of variables:

$$\begin{aligned} \mathbf{z} &= t\mathbf{x} + \tau \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi} + (1-t^2-\tau^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta}, \\ \omega_q(d\mathbf{z}) &= (1-t^2-\tau^2)^{\frac{q-3}{2}}\,\omega_{q-2}(d\boldsymbol{\eta})\,dt\,d\tau, \end{aligned} \tag{B.13}$$

where $t, \tau \in (-1,1)$, $t^2+\tau^2<1$, $\boldsymbol{\eta} \in \Omega_{q-2}$ and $\mathbf{A}_{\boldsymbol{\xi}} = (\mathbf{a}_1, \ldots, \mathbf{a}_q)_{q\times(q-1)}$ is the semi-orthonormal matrix resulting from the completion of $\boldsymbol{\xi}$ to the orthonormal basis $\{\boldsymbol{\xi}, \mathbf{a}_1, \ldots, \mathbf{a}_{q-1}\}$ of $\mathbb{R}^q$. This change of variables is obtained by a recursive use of Lemma B.4:

$$\begin{aligned} \int_{\Omega_q} f(\mathbf{z})\,\omega_q(d\mathbf{z}) &= \int_{-1}^{1}\int_{\Omega_{q-1}} f\left(t\mathbf{x} + (1-t^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}'\right)(1-t^2)^{\frac{q}{2}-1}\,\omega_{q-1}(d\boldsymbol{\xi}')\,dt \\ &= \int_{-1}^{1}\int_{-1}^{1}\int_{\Omega_{q-2}} f\left(t\mathbf{x} + (1-t^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\left(s\boldsymbol{\xi} + (1-s^2)^{\frac{1}{2}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta}\right)\right) \\ &\quad \times (1-s^2)^{\frac{q-3}{2}}(1-t^2)^{\frac{q}{2}-1}\omega_{q-2}(d\boldsymbol{\eta})\,ds\,dt \\ &= \iint_{t^2+\tau^2<1}\int_{\Omega_{q-2}} f\left(t\mathbf{x} + \tau\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi} + (1-t^2-\tau^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta}\right) \\ &\quad \times \left(1-\tau^2(1-t^2)^{-1}\right)^{\frac{q-3}{2}}(1-t^2)^{\frac{q-3}{2}}\omega_{q-2}(d\boldsymbol{\eta})\,d\tau\,dt \\ &= \iint_{t^2+\tau^2<1}\int_{\Omega_{q-2}} f\left(t\mathbf{x} + \tau\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi} + (1-t^2-\tau^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta}\right) \\ &\quad \times (1-t^2-\tau^2)^{\frac{q-3}{2}}\,\omega_{q-2}(d\boldsymbol{\eta})\,d\tau\,dt, \end{aligned}$$

where in the third equality a change of variables $\tau = (1-t^2)^{\frac{1}{2}}s$ is used. The matrix $\mathbf{B}_{\mathbf{x}}\mathbf{A}_{\boldsymbol{\xi}}$ of dimension $(q+1)\times(q-1)$ can be interpreted as the one formed by the column vectors that complete the orthonormal set $\{\mathbf{x}, \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\}$ to an orthonormal basis in $\mathbb{R}^{q+1}$.

If the changes of variables (B.12) and (B.13) is applied first, after that the changes $r = \frac{1-s}{h^2}$ and

$$\begin{cases} \rho = \frac{1-t}{h^2}, \\ \theta = \tau\left[h\left(\rho(2-h^2\rho)\right)^{\frac{1}{2}}\right]^{-1}, \end{cases} \qquad \left|\frac{\partial(t,\tau)}{\partial(\rho,\theta)}\right| = h^3\left[\rho(2-h^2\rho)\right]^{\frac{1}{2}}$$

are used and, denoting

$$\begin{aligned} \boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}} &= -rh^2\mathbf{x} + \left[rh^2(2-rh^2)\right]^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}, \\ \boldsymbol{\beta}_{\mathbf{x},\boldsymbol{\xi}} &= -h^2\rho\mathbf{x} + h\left[\rho(2-h^2\rho)\right]^{\frac{1}{2}}\left[\theta\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi} + (1-\theta^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta}\right], \end{aligned}$$

then the following result is obtained employing the DCT (see Lemma 4 of García-Portugués et al. (2014) for technical details in a similar situation):

$$\begin{aligned} \text{(B.11)} = \frac{1}{n^2h^{3q}\lambda_q(L)^4}\int_{\Omega_q}\int_{\Omega_q}\Bigg[\iint_{t^2+\tau^2<1}\int_{\Omega_{q-2}} \\ \times L\left(\frac{1-t}{h^2}\right)L\left(\frac{1-\mathbf{y}^T\left(t\mathbf{x}+\tau\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}+(1-t^2-\tau^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta}\right)}{h^2}\right) \\ \times \sigma^2\left(t\mathbf{x}+\tau\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}+(1-t^2-\tau^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta}\right) \end{aligned}$$

$$
\begin{aligned}
&\times f\left(t\mathbf{x}+\tau\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}+(1-t^2-\tau^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta}\right)(1-t^2-\tau^2)^{\frac{q-3}{2}}\,\omega_{q-2}(d\boldsymbol{\eta})\,dt\,d\tau\Big]^2\\
&\times\frac{w(\mathbf{x})w(\mathbf{y})}{f(\mathbf{x})f(\mathbf{y})}\,\omega_q(d\mathbf{x})\,\omega_q(d\mathbf{y})\\
=&\frac{1}{n^2h^{3q}\lambda_q(L)^4}\int_{-1}^{1}\int_{\Omega_{q-1}}\int_{\Omega_q}\Bigg[\iint_{t^2+\tau^2<1}\int_{\Omega_{q-2}}L\left(\frac{1-t}{h^2}\right)L\left(\frac{1-st-\tau(1-s^2)^{\frac{1}{2}}}{h^2}\right)\\
&\times\sigma^2\left(t\mathbf{x}+\tau\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}+(1-t^2-\tau^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta}\right)\\
&\times f\left(t\mathbf{x}+\tau\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}+(1-t^2-\tau^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\eta}\right)(1-t^2-\tau^2)^{\frac{q-3}{2}}\,\omega_{q-2}(d\boldsymbol{\eta})\,dt\,d\tau\Big]^2\\
&\times\frac{w(\mathbf{x})w\left(s\mathbf{x}+(1-s^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\right)}{f(\mathbf{x})f\left(s\mathbf{x}+(1-s^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\right)}\,\omega_q(d\mathbf{x})(1-s^2)^{\frac{q}{2}-1}\,\omega_{q-1}(d\boldsymbol{\xi})\,ds\\
=&\frac{1}{n^2\lambda_q(L)^4}\int_0^{2h^{-2}}\int_{\Omega_{q-1}}\int_{\Omega_q}\Bigg[\int_0^{2h^{-2}}\int_{-1}^{1}\int_{\Omega_{q-2}}L(\rho)\\
&\times L\left(r+\rho-h^2r\rho-\theta\left[r\rho(2-h^2r)(2-h^2\rho)\right]^{\frac{1}{2}}\right)\sigma^2\left(\mathbf{x}+\boldsymbol{\beta}_{\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}}\right)f\left(\mathbf{x}+\boldsymbol{\beta}_{\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}}\right)\\
&\times(1-\theta^2)^{\frac{q-3}{2}}\rho^{\frac{q}{2}-1}(2-h^2\rho)^{\frac{q}{2}-1}\,\omega_{q-2}(d\boldsymbol{\eta})\,dt\,d\tau\Big]^2\frac{w(\mathbf{x})w\left(\mathbf{x}+\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right)}{f(\mathbf{x})f\left(\mathbf{x}+\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right)}\\
&\times\omega_q(d\mathbf{x})\,r^{\frac{q}{2}-1}(2-h^2r)^{\frac{q}{2}-1}\,\omega_{q-1}(d\boldsymbol{\xi})\,dr\\
=&\frac{(1+o(1))}{n^2\lambda_q(L)^4}\int_0^{\infty}\int_{\Omega_{q-1}}\int_{\Omega_q}\Bigg[\int_0^{\infty}\int_{-1}^{1}\int_{\Omega_{q-2}}L(\rho)\,L\left(r+\rho-2\theta(r\rho)^{\frac{1}{2}}\right)\sigma^2(\mathbf{x})\,f(\mathbf{x})\\
&\times(1-\theta^2)^{\frac{q-3}{2}}\rho^{\frac{q}{2}-1}2^{\frac{q}{2}-1}\omega_{q-2}(d\boldsymbol{\eta})\,dt\,d\tau\Big]^2\frac{w(\mathbf{x})^2}{f(\mathbf{x})^2}\,\omega_q(d\mathbf{x})\,r^{\frac{q}{2}-1}2^{\frac{q}{2}-1}\,\omega_{q-1}(d\boldsymbol{\xi})\,dr\\
=&(1+o(1))\,\frac{\omega_{q-1}\omega_{q-2}^2\,2^{\frac{3q}{2}-3}}{n^2\lambda_q(L)^4}\int_{\Omega_q}\sigma^4(\mathbf{x})\,w(\mathbf{x})^2\,\omega_q(d\mathbf{x})\\
&\times\int_0^{\infty}r^{\frac{q}{2}-1}\left\{\int_0^{\infty}\rho^{\frac{q}{2}-1}L(\rho)\int_{-1}^{1}(1-\theta^2)^{\frac{q-3}{2}}L\left(r+\rho-2\theta(r\rho)^{\frac{1}{2}}\right)d\theta\,d\rho\right\}^2dr\\
=&n^{-2}\nu^2(1+o(1)).
\end{aligned}
$$

For $q=1$, define the change of variables:

$$
\begin{aligned}
\mathbf{y}&=s\mathbf{x}+(1-s^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi},\quad \omega_1(d\mathbf{y})=(1-s^2)^{-\frac{1}{2}}\,\omega_0(d\boldsymbol{\xi})\,ds,\\
\mathbf{z}&=t\mathbf{x}+(1-t^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\eta},\quad \omega_1(d\mathbf{z})=(1-t^2)^{-\frac{1}{2}}\,\omega_0(d\boldsymbol{\eta})\,dt,
\end{aligned}
$$

where $\boldsymbol{\xi},\boldsymbol{\eta}\in\Omega_0=\{-1,1\}$. Note that as $q=1$ and $\mathbf{x}^T(\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})=\mathbf{x}^T(\mathbf{B}_{\mathbf{x}}\boldsymbol{\eta})=0$, then necessarily $\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}=\mathbf{B}_{\mathbf{x}}\boldsymbol{\eta}$ or $\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}=-\mathbf{B}_{\mathbf{x}}\boldsymbol{\eta}$. These changes of variables are applied first, later $\rho=\frac{1-t}{h^2}$ and finally $r=\frac{1-s}{h^2}$, using that:

$$
\frac{1-st-(1-s^2)^{\frac{1}{2}}(1-t^2)^{\frac{1}{2}}(\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T\mathbf{B}_{\mathbf{x}}\boldsymbol{\eta}}{h^2}
$$

$$= r + \rho - h^2 r\rho - \Big(r\rho(2-h^2 r)(2-h^2\rho)\Big)^{\frac{1}{2}} (\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T \mathbf{B}_{\mathbf{x}}\boldsymbol{\eta}.$$

Finally, considering

$$\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}} = -rh^2\mathbf{x} + \Big[rh^2(2-rh^2)\Big]^{\frac{1}{2}} \mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}, \quad \boldsymbol{\beta}_{\mathbf{x},\boldsymbol{\eta}} = -\rho h^2\mathbf{x} + \Big[\rho h^2(2-\rho h^2)\Big]^{\frac{1}{2}} \mathbf{B}_{\mathbf{x}}\boldsymbol{\eta},$$

it follows by the use of the DCT:

$$\begin{aligned}
\text{(B.11)} = & \frac{1}{n^2 h^3 \lambda_q(L)^4} \int_{\Omega_1}\int_{\Omega_1} \Bigg[\int_{-1}^{1}\int_{\Omega_0} L\left(\frac{1-t}{h^2}\right) L\left(\frac{1-\mathbf{y}^T(t\mathbf{x}+(1-t^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\eta})}{h^2}\right) \\
& \times \sigma^2\left(t\mathbf{x}+(1-t^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\eta}\right) f\left(t\mathbf{x}+(1-t^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\eta}\right)(1-t^2)^{-\frac{1}{2}}\,\omega_0(d\boldsymbol{\eta})\,dt\Bigg]^2 \\
& \times \frac{w(\mathbf{x})w(\mathbf{y})}{f(\mathbf{x})f(\mathbf{y})}\,\omega_1(d\mathbf{x})\,\omega_1(d\mathbf{y}) \\
= & \frac{1}{n^2 h^3 \lambda_q(L)^4} \int_{-1}^{1}\int_{\Omega_0}\int_{\Omega_1} \Bigg[\int_{-1}^{1}\int_{\Omega_0} \\
& \times L\left(\frac{1-t}{h^2}\right) L\left(\frac{1-st-(1-t^2)^{\frac{1}{2}}(1-s^2)^{\frac{1}{2}}(\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T(\mathbf{B}_{\mathbf{x}}\boldsymbol{\eta})}{h^2}\right) \\
& \times \sigma^2\left(t\mathbf{x}+(1-t^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\right) f\left(t\mathbf{x}+(1-t^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\right)(1-t^2)^{-\frac{1}{2}}\,\omega_0(d\boldsymbol{\eta})\,dt\Bigg]^2 \\
& \times \frac{w(\mathbf{x})w\left(s\mathbf{x}+(1-s^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\right)}{f(\mathbf{x})f\left(s\mathbf{x}+(1-s^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi}\right)}\,\omega_1(d\mathbf{x})\,(1-s^2)^{-\frac{1}{2}}\,\omega_0(d\boldsymbol{\xi})\,ds \\
= & \frac{1}{n^2 \lambda_q(L)^4} \int_{0}^{2h^{-2}} \int_{\Omega_0}\int_{\Omega_1} \Bigg[\int_{0}^{2h^{-2}}\int_{\Omega_0} \\
& \times L(\rho)\, L\left(r+\rho-h^2 r\rho-\Big(r\rho(2-h^2 r)(2-h^2\rho)\Big)^{\frac{1}{2}}(\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T\mathbf{B}_{\mathbf{x}}\boldsymbol{\eta}\right) \\
& \times \sigma^2\left(\mathbf{x}+\boldsymbol{\beta}_{\mathbf{x},\boldsymbol{\eta}}\right) f\left(\mathbf{x}+\boldsymbol{\beta}_{\mathbf{x},\boldsymbol{\eta}}\right)\rho^{-\frac{1}{2}}(2-h^2\rho)^{-\frac{1}{2}}\,\omega_0(d\boldsymbol{\eta})\,d\rho\Bigg]^2 \frac{w(\mathbf{x})w\left(\mathbf{x}+\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right)}{f(\mathbf{x})f\left(\mathbf{x}+\boldsymbol{\alpha}_{\mathbf{x},\boldsymbol{\xi}}\right)} \\
& \times \omega_1(d\mathbf{x})\, r^{-\frac{1}{2}}(2-h^2 r)^{-\frac{1}{2}}\,\omega_0(d\boldsymbol{\xi})\,dr \\
= & \frac{2^{-1}\left(1+o(1)\right)}{n^2\lambda_q(L)^4} \int_{0}^{\infty}\int_{\Omega_0}\int_{\Omega_1} \Bigg[\int_{0}^{\infty}\int_{\Omega_0} L(\rho)\, L\left(r+\rho-2(r\rho)^{\frac{1}{2}}(\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})^T\mathbf{B}_{\mathbf{x}}\boldsymbol{\eta}\right) \\
& \times \sigma^2(\mathbf{x})\, f(\mathbf{x})\,\rho^{-\frac{1}{2}}\,\omega_0(d\boldsymbol{\eta})\,d\rho\Bigg]^2 \frac{w(\mathbf{x})^2}{f(\mathbf{x})^2}\,\omega_1(d\mathbf{x})\, r^{-\frac{1}{2}}\,\omega_0(d\boldsymbol{\xi})\,dr \\
= & \frac{\omega_0 2^{-\frac{3}{2}}\left(1+o(1)\right)}{n^2\lambda_q(L)^4} \int_{\Omega_1} \sigma^4(\mathbf{x})\, w(\mathbf{x})^2\,\omega_1(d\mathbf{x}) \\
& \times \int_{0}^{\infty} r^{-\frac{1}{2}}\Bigg\{\int_{0}^{\infty}\rho^{-\frac{1}{2}} L(\rho)\left[L\left(r+\rho-2(r\rho)^{\frac{1}{2}}\right)+L\left(r+\rho+2(r\rho)^{\frac{1}{2}}\right)\right]d\rho\Bigg\}^2 dr \\
= & n^{-2}\nu^2\left(1+o(1)\right).
\end{aligned}$$

The rest of the results are provided by the recursive use of Lemma B.6, bearing in mind that the indexes are pairwise different:

$$
\begin{aligned}
&\mathbb{E}[W_{ijn}^4] \\
&\quad = \frac{n^4h^{2q}}{n^8h^{8q}\lambda_q(L)^8}\int_{\Omega_q}\times\overset{4}{\cdots}\times\int_{\Omega_q}\mathbb{E}\left[\prod_{k=1}^{4}L\left(\frac{1-\mathbf{x}_k^T\mathbf{X}}{h^2}\right)\sigma^4(\mathbf{X})\mathbb{E}\left[\varepsilon^4|\mathbf{X}\right]\right]^2\prod_{k=1}^{4}\frac{w(\mathbf{x}_k)}{f(\mathbf{x}_k)}\,\omega_q(d\mathbf{x}_k) \\
&\quad = \mathcal{O}\left((n^4h^{4q})^{-1}\right)\int_{\Omega_q}\times\overset{4}{\cdots}\times\int_{\Omega_q}\prod_{k=2}^{4}L^2\left(\frac{1-\mathbf{x}_k^T\mathbf{X}}{h^2}\right)\sigma^8(\mathbf{x}_1)f(\mathbf{x}_1)\prod_{k=1}^{8}\frac{w(\mathbf{x}_k)}{f(\mathbf{x}_k)}\,\omega_q(d\mathbf{x}_k) \\
&\quad = \mathcal{O}\left((n^4h^{q})^{-1}\right), \\
&\mathbb{E}\Big[W_{ijn}W_{jkn}W_{kln}W_{lin}\Big] \\
&\quad = \frac{n^4h^{2q}}{n^8h^{8q}\lambda_q(L)^8}\int_{\Omega_q}\times\overset{4}{\cdots}\times\int_{\Omega_q}\mathbb{E}\left[L\left(\frac{1-\mathbf{x}_1^T\mathbf{X}}{h^2}\right)L\left(\frac{1-\mathbf{x}_4^T\mathbf{X}}{h^2}\right)\sigma^2(\mathbf{X})\right] \\
&\qquad \times\mathbb{E}\left[L\left(\frac{1-\mathbf{x}_1^T\mathbf{X}}{h^2}\right)L\left(\frac{1-\mathbf{x}_2^T\mathbf{X}}{h^2}\right)\sigma^2(\mathbf{X})\right]\mathbb{E}\left[L\left(\frac{1-\mathbf{x}_2^T\mathbf{X}}{h^2}\right)L\left(\frac{1-\mathbf{x}_3^T\mathbf{X}}{h^2}\right)\sigma^2(\mathbf{X})\right] \\
&\qquad \times\mathbb{E}\left[L\left(\frac{1-\mathbf{x}_3^T\mathbf{X}}{h^2}\right)L\left(\frac{1-\mathbf{x}_4^T\mathbf{X}}{h^2}\right)\sigma^2(\mathbf{X})\right]\prod_{k=1}^{8}\frac{w(\mathbf{x}_k)}{f(\mathbf{x}_k)}\,\omega_q(d\mathbf{x}_k) \\
&\quad = \mathcal{O}\left((n^4h^{2q})^{-1}\right)\int_{\Omega_q}\times\overset{4}{\cdots}\times\int_{\Omega_q}L\left(\frac{1-\mathbf{x}_4^T\mathbf{x}_1}{h^2}\right)L\left(\frac{1-\mathbf{x}_2^T\mathbf{x}_1}{h^2}\right)L\left(\frac{1-\mathbf{x}_2^T\mathbf{x}_3}{h^2}\right) \\
&\qquad \times L\left(\frac{1-\mathbf{x}_4^T\mathbf{x}_3}{h^2}\right)\sigma^4(\mathbf{x}_1)\sigma^4(\mathbf{x}_3)\frac{f(\mathbf{x}_1)f(\mathbf{x}_3)}{f(\mathbf{x}_2)f(\mathbf{x}_3)}\prod_{k=1}^{4}w(\mathbf{x}_k)\,\omega_q(d\mathbf{x}_k) \\
&\quad = \mathcal{O}\left(n^{-4}h^{2q}\right), \\
&\mathbb{E}\Big[W_{ijn}W_{ikn}^2W_{jkn}\Big] \\
&\quad = \frac{n^4h^{2q}}{n^8h^{8q}\lambda_q(L)^8}\int_{\Omega_q}\times\overset{4}{\cdots}\times\int_{\Omega_q}\mathbb{E}\left[L\left(\frac{1-\mathbf{x}_1^T\mathbf{X}}{h^2}\right)L\left(\frac{1-\mathbf{x}_2^T\mathbf{X}}{h^2}\right)\sigma^2(\mathbf{X})\right] \\
&\qquad \times\mathbb{E}\left[L\left(\frac{1-\mathbf{x}_1^T\mathbf{X}}{h^2}\right)L\left(\frac{1-\mathbf{x}_3^T\mathbf{X}}{h^2}\right)L\left(\frac{1-\mathbf{x}_4^T\mathbf{X}}{h^2}\right)\sigma^3(\mathbf{X})\mathbb{E}\left[\varepsilon^3|\mathbf{X}\right]\right]^2\prod_{k=1}^{4}\frac{w(\mathbf{x}_k)}{f(\mathbf{x}_k)}\,\omega_q(d\mathbf{x}_k) \\
&\quad = \mathcal{O}\left((n^4h^{3q})^{-1}\right)\int_{\Omega_q}\times\overset{4}{\cdots}\times\int_{\Omega_q}L\left(\frac{1-\mathbf{x}_2^T\mathbf{x}_1}{h^2}\right)L^2\left(\frac{1-\mathbf{x}_3^T\mathbf{x}_1}{h^2}\right)L^2\left(\frac{1-\mathbf{x}_4^T\mathbf{x}_1}{h^2}\right) \\
&\qquad \times\sigma^8(\mathbf{x}_1)\frac{f(\mathbf{x}_1)^2}{f(\mathbf{x}_2)f(\mathbf{x}_3)f(\mathbf{x}_4)}\prod_{k=1}^{4}w(\mathbf{x}_k)\,\omega_q(d\mathbf{x}_k) \\
&\quad = \mathcal{O}\left(n^{-4}\right).
\end{aligned}
$$

□

Lemma B.3. *Under assumptions **A1**–**A6** and **A9**, for a random sample $\{(\mathbf{X}_i,Y_i)\}_{i=1}^n$ the following statements hold:*

i. $\int_{\Omega_q}\sum_{i=1}^n\left(W_n^p\left(\mathbf{x},\mathbf{X}_i\right)\hat{\varepsilon}_iV_i^*\right)^2\hat{f}_h(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}) = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2_{\boldsymbol{\theta}_1}(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})$ $\times(1+o_{\mathbb{P}}(1))+\mathcal{O}_{\mathbb{P}^*}\left((n^{\frac{3}{2}}h^q)^{-1}\right)$.

ii. $2n^2h^q\sum_{i\neq j}\left[\int_{\Omega_q} W_n^p\left(\mathbf{x},\mathbf{X}_i\right) W_n^p\left(\mathbf{x},\mathbf{X}_j\right)\hat{\varepsilon}_i\hat{\varepsilon}_j\hat{f}_h(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right]^2 = 2\nu^2_{\boldsymbol{\theta}_1}\left(1+o_{\mathbb{P}}\left(1\right)\right)$, $\mathbb{E}^*[W^*_{ijn}W^*_{jkn}W^*_{kln}W^*_{lin}] = \mathcal{O}_{\mathbb{P}}\left(n^{-4}h^{2q}\right)$, $\mathbb{E}^*[W^{*4}_{ijn}] = \mathcal{O}_{\mathbb{P}}\left((n^4h^q)^{-1}\right)$ *and* $\mathbb{E}^*[W^*_{ijn}W^{*2}_{ikn}W^*_{jkn}] = \mathcal{O}_{\mathbb{P}}\left(n^{-4}\right)$.

Proof of Lemma B.3. The proof is divided in the evaluation of each statement.

Proof of i. Using that the V_i^*'s are iid and independent with respect to the sample,

$$\begin{aligned}\mathbb{E}^*\left[\int_{\Omega_q}\sum_{i=1}^n W_n^p(\mathbf{x},\mathbf{X}_i)^2\hat{\varepsilon}_i^2V_i^{*2}\hat{f}_h(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right]&\\ = \int_{\Omega_q}\sum_{i=1}^n W_n^p(\mathbf{x},\mathbf{X}_i)^2(Y_i-m_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i))^2\hat{f}_h(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})&\\ = \int_{\Omega_q}\sum_{i=1}^n L_h^*(\mathbf{x},\mathbf{X}_i)^2(Y_i-m_{\boldsymbol{\theta}_1}(\mathbf{X}_i))^2 f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})(1+o_{\mathbb{P}}\left(1\right)) &\qquad \text{(B.14)}\end{aligned}$$

where the last equality holds because by assumptions **A5** and **A6**, $m_{\hat{\boldsymbol{\theta}}}(\mathbf{x}) - m_{\boldsymbol{\theta}_1}(\mathbf{x}) = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$ uniformly in $\mathbf{x}\in\Omega_q$. By applying the tower property of the conditional expectation as in *iii* from Lemma B.1, it is easy to derive from *iv* in Lemma B.2 that

$$\text{(B.14)} = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2_{\boldsymbol{\theta}_1}(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})(1+o_{\mathbb{P}}\left(1\right)).$$

The order of the variance is obtained applying the same idea, i.e., first deriving the variance with respect to the V_i^*'s and then applying the order computation given in the proof of *iv* in Lemma B.2 (adapted via the conditional expectation):

$$\begin{aligned}\mathbb{V}\text{ar}^*\left[\int_{\Omega_q}\sum_{i=1}^n L_h^*(\mathbf{x},\mathbf{X}_i)^2(Y_i-m_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i))^2V_i^{*2}f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right]&\\ = \sum_{i=1}^n\left(\int_{\Omega_q}L_h^*(\mathbf{x},\mathbf{X}_i)^2(Y_i-m_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i))^2f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right)^2\mathbb{V}\text{ar}^*\left[V_i^{*2}\right]&\\ = \mathcal{O}\left((n^4h^{4q})^{-1}\right)\sum_{i=1}^n\left(\int_{\Omega_q}L^2\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h^2}\right)(Y_i-m_{\boldsymbol{\theta}_1}(\mathbf{X}_i))^2\frac{w(\mathbf{x})}{f(\mathbf{x})}\,\omega_q(d\mathbf{x})\right)^2&\\ = \mathcal{O}_{\mathbb{P}}\left((n^3h^{2q})^{-1}\right).&\end{aligned}$$

The statement holds by Chebychev's inequality with respect to the probability law $\mathbb{P}^*$.

Proof of ii. First, by Corollary 7.1, the expansion for the kernel density estimate and the fact $m_{\hat{\boldsymbol{\theta}}}(\mathbf{x}) - m_{\boldsymbol{\theta}_1}(\mathbf{x}) = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$ uniformly in $\mathbf{x}\in\Omega_q$, we have

$$\begin{aligned}I_n &= 2n^2h^q\sum_{i\neq j}\left[\int_{\Omega_q}W_n^p\left(\mathbf{x},\mathbf{X}_i\right)W_n^p\left(\mathbf{x},\mathbf{X}_j\right)\hat{\varepsilon}_i\hat{\varepsilon}_j\hat{f}_h(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right]^2\\ &= 2\sum_{i\neq j}I_{ijn}(1+o_{\mathbb{P}}\left(1\right)),\end{aligned}$$

where

$$I_{ijn} = n^2 h^q \int_{\Omega_q}\int_{\Omega_q} L_h^*(\mathbf{x},\mathbf{X}_i)\, L_h^*(\mathbf{x},\mathbf{X}_j)\, L_h^*(\mathbf{y},\mathbf{X}_i)\, L_h^*(\mathbf{y},\mathbf{X}_j) \\ \times (Y_i - m_{\boldsymbol{\theta}_1}(\mathbf{X}_i))^2 (Y_j - m_{\boldsymbol{\theta}_1}(\mathbf{X}_j))^2 f(\mathbf{x}) f(\mathbf{y}) w(\mathbf{x}) w(\mathbf{y})\, \omega_q(d\mathbf{x})\, \omega_q(d\mathbf{y}).$$

By the tower property of the conditional expectation and *iv* in Lemma B.2, $\mathbb{E}[I_{ijn}] = \mathbb{E}[W_{ijn}^2] = n^{-2}\nu_{\boldsymbol{\theta}_1}^2(1+o(1))$ (considering that the W_{ijn}'s are defined with respect to $\boldsymbol{\theta}_1$ instead of $\boldsymbol{\theta}_0$). To prove that $I_n \xrightarrow{p} 2\nu_{\boldsymbol{\theta}_1}^2$, consider $\widetilde{I}_n = 2\sum_{i\neq j} I_{ijn}$ and, by (7.18) and (7.19),

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\widetilde{I}_n\right] &= \mathbb{E}\left[\left(2\sum_{i\neq j} I_{ijn}\right)^2\right] - 4n^2(n-1)^2\mathbb{E}[I_{ijn}]^2 \\
&= 4\sum_{i\neq j}\sum_{k\neq l}\mathbb{E}\left[W_{ijn}^2 W_{kln}^2\right] - 4n^2(n-1)^2\mathbb{E}\left[W_{ijn}^2\right]^2 \\
&= \frac{1}{3}\mathbb{E}\left[W_n^4\right] - \mathbb{V}\mathrm{ar}[W_n]^2 + o(1) \\
&= \mathbb{V}\mathrm{ar}[W_n]^2\left(\frac{1}{3}\mathbb{V}\mathrm{ar}[W_n]^{-2}\,\mathbb{E}\left[W_n^4\right] - 1\right) + o(1) \\
&= 2\nu_{\boldsymbol{\theta}_1}^2(1+o(1))o(1) + o(1) \\
&= o(1),
\end{aligned}$$

because, as it was shown in the proof of Theorem 7.2, conditions *b)* and *d)* hold. Then, $\widetilde{I}_n - \mathbb{E}[\widetilde{I}_n]$ converges to zero in squared mean, which implies that it converges in probability and therefore

$$I_n = \widetilde{I}_n(1+o_{\mathbb{P}}(1)) = \left(\widetilde{I}_n - \mathbb{E}[\widetilde{I}_n] + 2\nu_{\boldsymbol{\theta}_1}^2 + o(1)\right)(1+o_{\mathbb{P}}(1)) = 2\nu_{\boldsymbol{\theta}_1}^2 + o_{\mathbb{P}}(1),$$

which proofs the first statement.

Second, it follows straightforwardly that $\mathbb{E}^*[W_{ijn}^{*4}] = \mathcal{O}_{\mathbb{P}}(W_{ijn}^4)$, $\mathbb{E}^*[W_{ijn}^* W_{jkn}^* W_{kln}^* W_{lin}^*] = \mathcal{O}_{\mathbb{P}}(W_{ijn}W_{jkn}W_{kln}W_{lin})$ and $\mathbb{E}^*[W_{ijn}^* W_{ikn}^{*2} W_{jkn}^*] = \mathcal{O}_{\mathbb{P}}(W_{ijn}W_{ikn}^2 W_{jkn})$. The idea now is to use that, for a random variable X_n and by the Markov's inequality, $X_n = \mathbb{E}[X_n] + \mathcal{O}_{\mathbb{P}}(\mathbb{E}[|X_n|])$. The expectations of the variables are given in *v* from Lemma B.2. The orders of the absolute expectations are the same: in the definition of W_{ijn} the only factor with sign is $\varepsilon_i\varepsilon_j$, which is handled by the assumption of boundedness of $\mathbb{E}[|\varepsilon|^3|\mathbf{X}]$. Therefore, $W_{ijn}^4 = \mathcal{O}_{\mathbb{P}}((n^4h^q)^{-1})$, $W_{ijn}W_{jkn}W_{kln}W_{lin} = \mathcal{O}_{\mathbb{P}}(n^{-4}h^{2q})$ and $W_{ijn}W_{ikn}^2 W_{jkn} = \mathcal{O}_{\mathbb{P}}(n^{-4})$, so the statement is proved. □

B.2.3 General purpose lemmas

Lemma B.4 (Tangent-normal change of variables)**.** *Let f be a function defined in Ω_q and $\mathbf{x}\in\Omega_q$. Then $\int_{\Omega_q} f(\mathbf{z})\,\omega_q(d\mathbf{z}) = \int_{-1}^1\int_{\Omega_{q-1}} f(t\mathbf{x} + (1-t^2)^{\frac{1}{2}}\mathbf{B}_{\mathbf{x}}\boldsymbol{\xi})(1-t^2)^{\frac{q}{2}-1}\,\omega_{q-1}(d\boldsymbol{\xi})\,dt$, where $\mathbf{B}_{\mathbf{x}} = (\mathbf{b}_1,\ldots,\mathbf{b}_q)_{(q+1)\times q}$ is the projection matrix given in Section 7.2.*

Proof of Lemma B.4. See Lemma 2 of García-Portugués et al. (2013). □

Lemma B.5. *Set $\mathbf{x} = (x_1,\ldots,x_{q+1})\in\Omega_q$. For all $i,j,k = 1,\ldots,q+1$, $\int_{\Omega_q} x_i\,\omega_q(d\mathbf{x}) = 0$, $\int_{\Omega_q} x_i x_j\,\omega_q(d\mathbf{x}) = \delta_{ij}\frac{\omega_q}{q+1}$ and $\int_{\Omega_q} x_i x_j x_k\,\omega_q(d\mathbf{x}) = 0$.*

Proof of Lemma B.5. Apply Lemma B.4 considering $\mathbf{x} = \mathbf{e}_i \in \Omega_q$. Then $\int_{\Omega_q} x_i \, \omega_q(d\mathbf{x}) = \omega_{q-1} \int_{-1}^{1} t(1-t^2)^{\frac{q}{2}-1} \, dt = 0$ as the integrand is an odd function. As a consequence, and applying the same change of variables, for $i \neq j$:

$$\int_{\Omega_q} x_i x_j \, \omega_q(d\mathbf{x}) = \int_{-1}^{1} (1-t^2)^{\frac{q-1}{2}} \, dt \int_{\Omega_{q-1}} \mathbf{e}_j^T \mathbf{B}_{\mathbf{x}} \boldsymbol{\xi} \, \omega_{q-1}(d\boldsymbol{\xi}) = 0.$$

For $i = j$, $\int_{\Omega_q} x_i^2 \, \omega_q(d\mathbf{x}) = \frac{1}{q+1} \int_{\Omega_q} \sum_{j=1}^{q} x_j^2 \, \omega_q(d\mathbf{x}) = \frac{\omega_q}{q+1}$. For the trivariate case,

$$\begin{aligned}
\int_{\Omega_q} x_i^3 \, \omega_q(d\mathbf{x}) &= \omega_{q-1} \int_{-1}^{1} t^3 (1-t^2)^{\frac{q}{2}-1} \, dt = 0, \\
\int_{\Omega_q} x_i^2 x_j \, \omega_q(d\mathbf{x}) &= \int_{-1}^{1} t^2 (1-t^2)^{\frac{q-1}{2}} \, dt \int_{\Omega_{q-1}} \mathbf{e}_j^T \mathbf{B}_{\mathbf{x}} \boldsymbol{\xi} \, \omega_{q-1}(d\boldsymbol{\xi}) = 0, \; i \neq j, \\
\int_{\Omega_q} x_i x_j x_k \, \omega_q(d\mathbf{x}) &= \int_{-1}^{1} t (1-t^2)^{\frac{q}{2}} \, dt \int_{\Omega_{q-1}} \mathbf{e}_j^T \mathbf{B}_{\mathbf{x}} \boldsymbol{\xi} \mathbf{e}_k^T \mathbf{B}_{\mathbf{x}} \boldsymbol{\xi} \, \omega_{q-1}(d\boldsymbol{\xi}) = 0, \; i \neq j \neq k,
\end{aligned}$$

using that the integrand is odd and the first statement. □

Lemma B.6 (Bai et al. (1988))**.** *Let $\varphi : \Omega_q \longrightarrow \mathbb{R}$ be a continuous function and denote $L_h\varphi(\mathbf{x}) = c_{h,q}(L) \int_{\Omega_q} L\left(\frac{1-\mathbf{x}^T\mathbf{y}}{h^2}\right) \varphi(\mathbf{y}) \, \omega_q(d\mathbf{y})$. Under assumptions **A3**–**A4**, $L_h\varphi(\mathbf{x}) = \varphi(\mathbf{x}) + o(1)$, where the remaining order is uniform for any $\mathbf{x} \in \Omega_q$.*

Proof of Lemma B.6. This corresponds to Lemma 5 in Bai et al. (1988), but with slightly different conditions and notation. Assumptions **A1** and **A3** imply conditions (a), (b), (c_1) and (d) stated in Theorem 1 of the aforementioned paper. □

B.3 Further simulation results

Some extra simulation results are given to provide a better understanding of the design of the simulation study presented in the paper and a deeper insight into the empirical performance of the goodness-of-fit tests for different significance levels and sample sizes.

Graphical representations of the densities considered for the directional predictor $\mathbf{X}$ are shown in Figure B.1. These densities aim to capture simple designs like the uniform and more challenging ones with *holes* in the support.

The deviations from the null hypothesis, Δ_1 and Δ_2, are shown in Figure B.2, jointly with the conditional standard deviation function used to generate data with heteroskedastic noise.

The coefficients δ for obtaining deviations $\delta\Delta_1$ and $\delta\Delta_2$ in each scenario were chosen such that the density of the response $Y = m_{\boldsymbol{\theta}_0}(\mathbf{X}) + \delta\Delta(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$ under H_0 ($\delta = 0$) and under H_1 ($\delta \neq 0$) were similar. Figure B.3 shows the densities of Y under the null and the alternative for the four scenarios and dimensions considered. This is a graphical way of ensuring that the deviation is not trivial to detect and hence is not straightforward to reject H_0. Note that, due to the design of the deviations, it may be easier to detect them on a particular dimension.
The empirical sizes of the test for significance levels $\alpha = 0.01, 0.05, 0.10$ are given in Figures B.4, B.5 and B.6, corresponding to sample sizes $n = 100, 250$ and 500. Nominal levels are respected in most scenarios. Finally, the empirical powers for $n = 100, 250$ and 500 are given in Figure

B.7 and, as it can be seen, the rejection rates increase with n. A final remark is that it seems harder to detect the alternative in S4 for $q = 2$ due to the shape of the parametric model, Δ_2 and the design density.

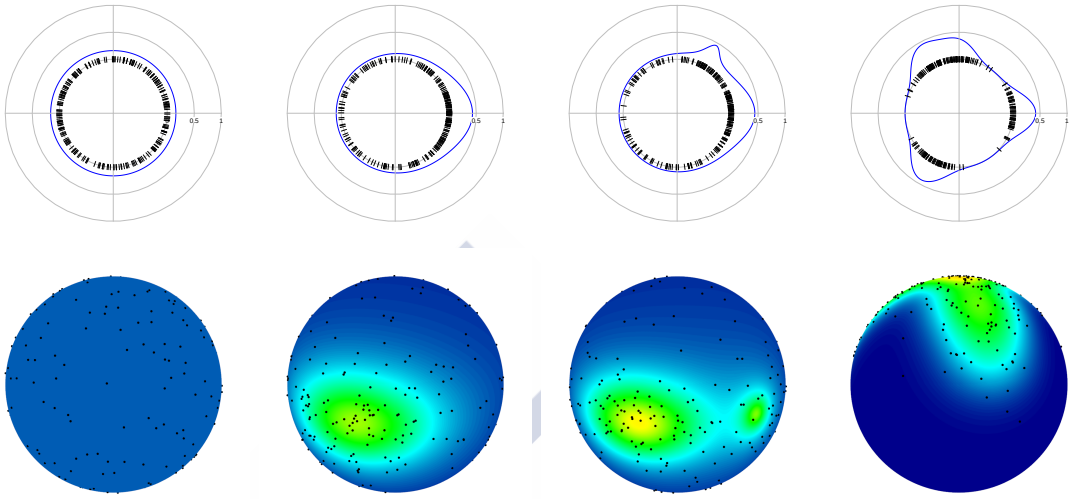


Figure B.1: From left to right: directional densities for scenarios S1 to S4 for circular and spherical cases.

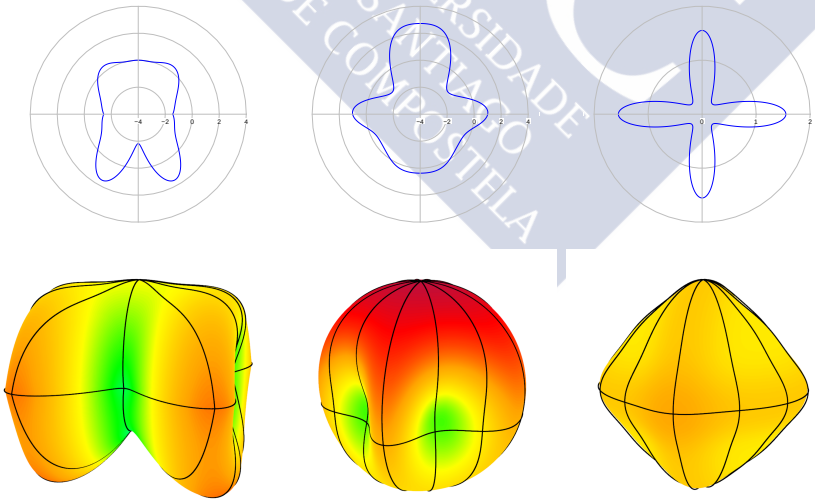


Figure B.2: From left to right: deviations Δ_1 and Δ_2 and conditional standard deviation function σ_2 for circular and spherical cases.

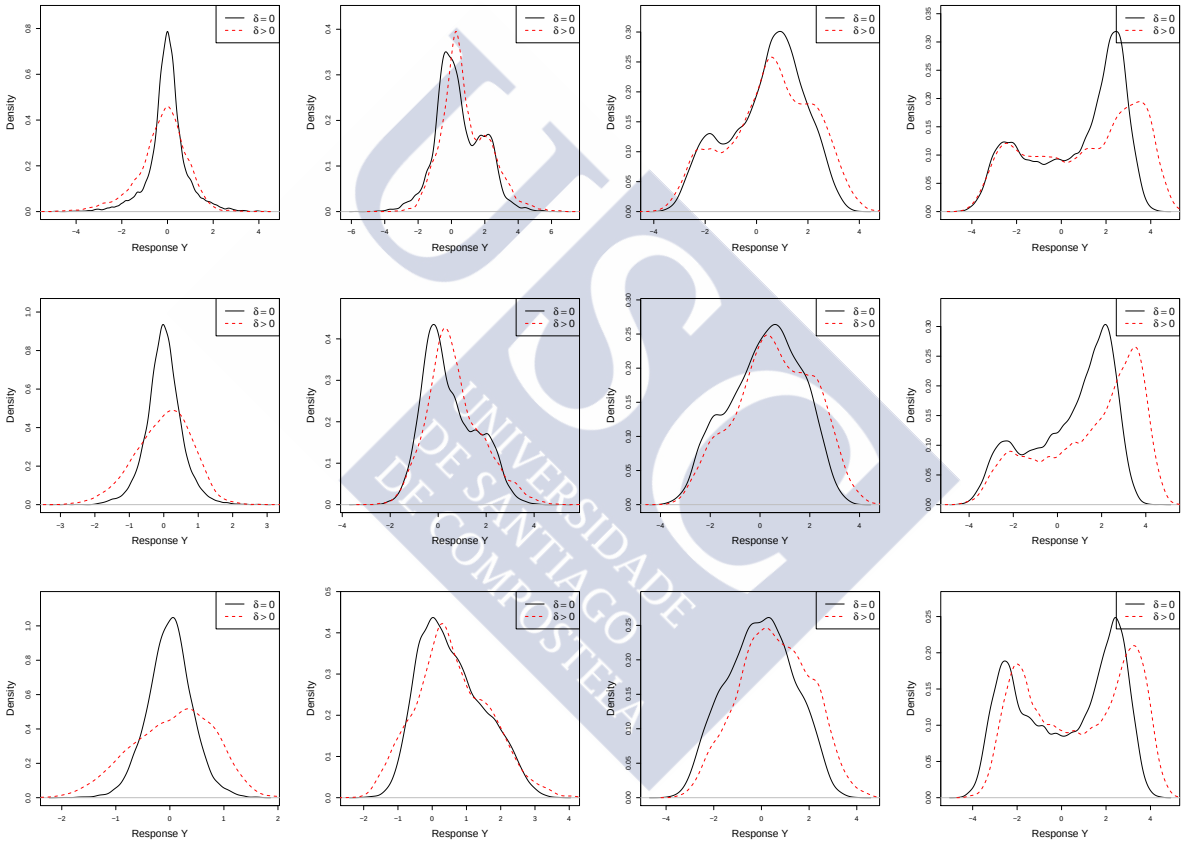


Figure B.3: Densities of the response Y under the null (solid line) and under the alternative (dashed line) for scenarios S1 to S4 (columns, from left to right) and dimensions $q = 1, 2, 3$ (rows, from top to bottom).

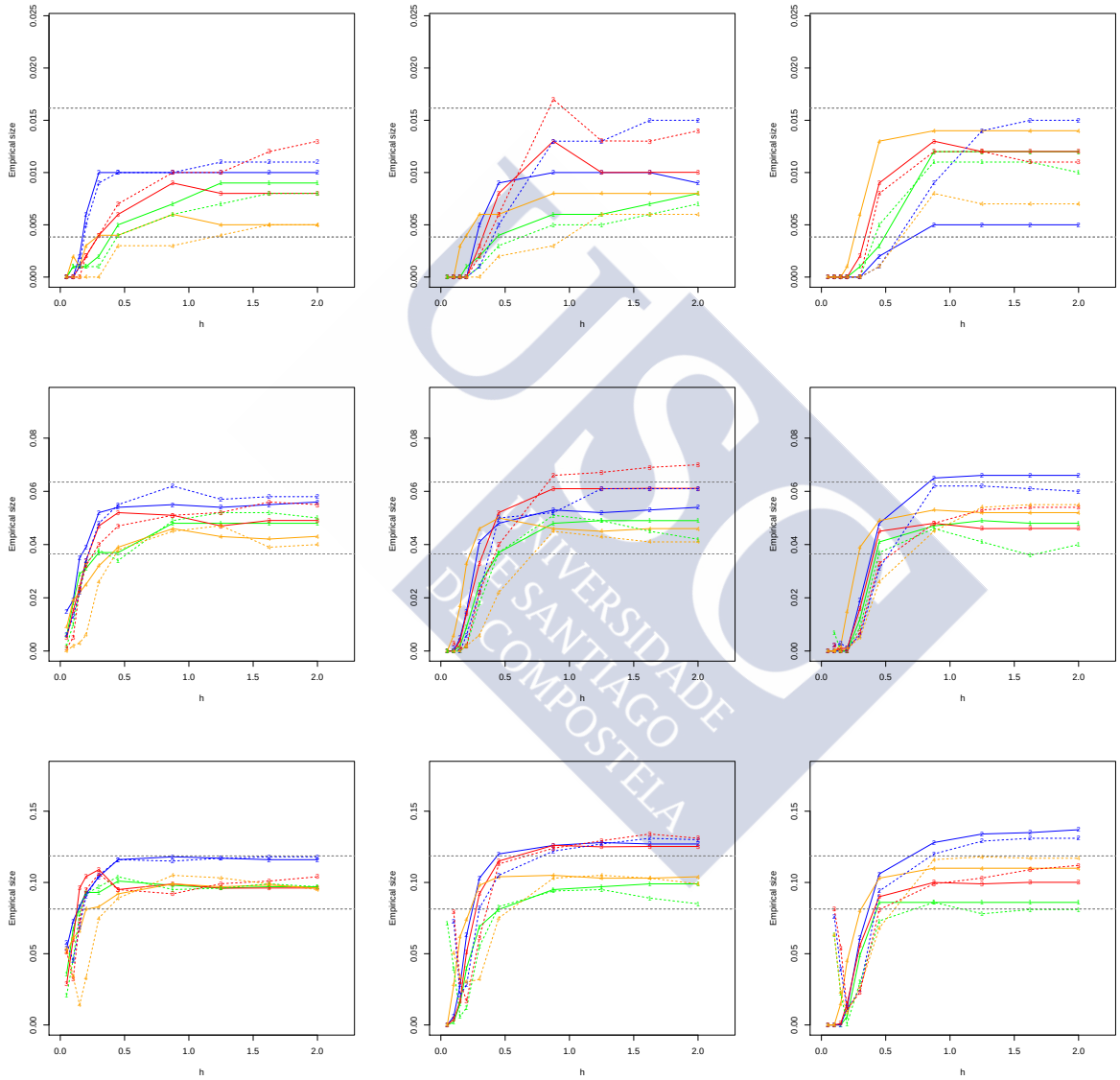


Figure B.4: Empirical sizes for $\alpha = 0.01$ (first row), $\alpha = 0.05$ (second row) and $\alpha = 0.10$ (third row) for the different scenarios, with $p = 0$ (solid line) and $p = 1$ (dashed line). From left to right, columns represent dimensions $q = 1, 2, 3$ with sample size $n = 100$.

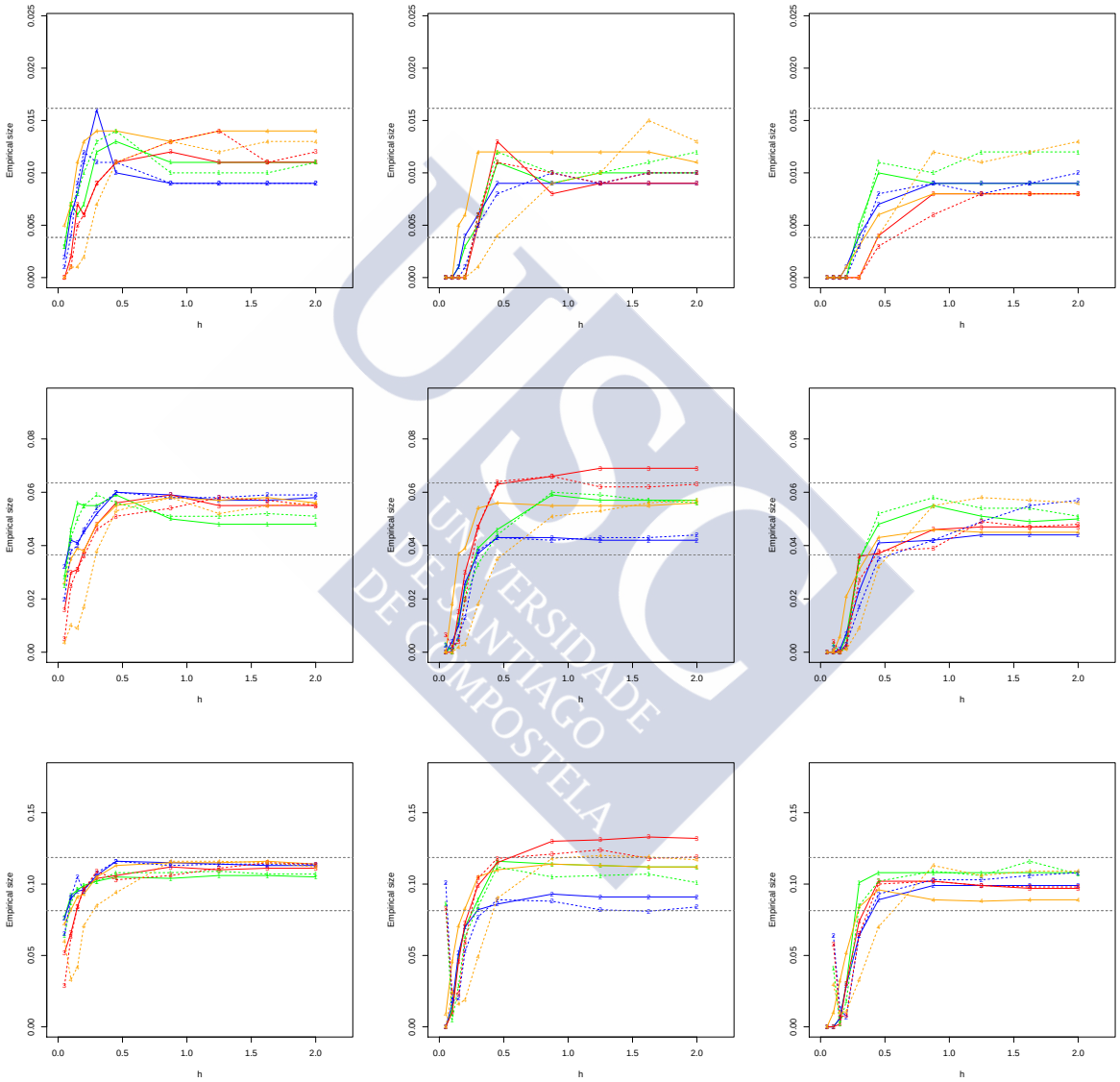


Figure B.5: Empirical sizes for $\alpha = 0.01$ (first row), $\alpha = 0.05$ (second row) and $\alpha = 0.10$ (third row) for the different scenarios, with $p = 0$ (solid line) and $p = 1$ (dashed line). From left to right, columns represent dimensions $q = 1, 2, 3$ with sample size $n = 250$.

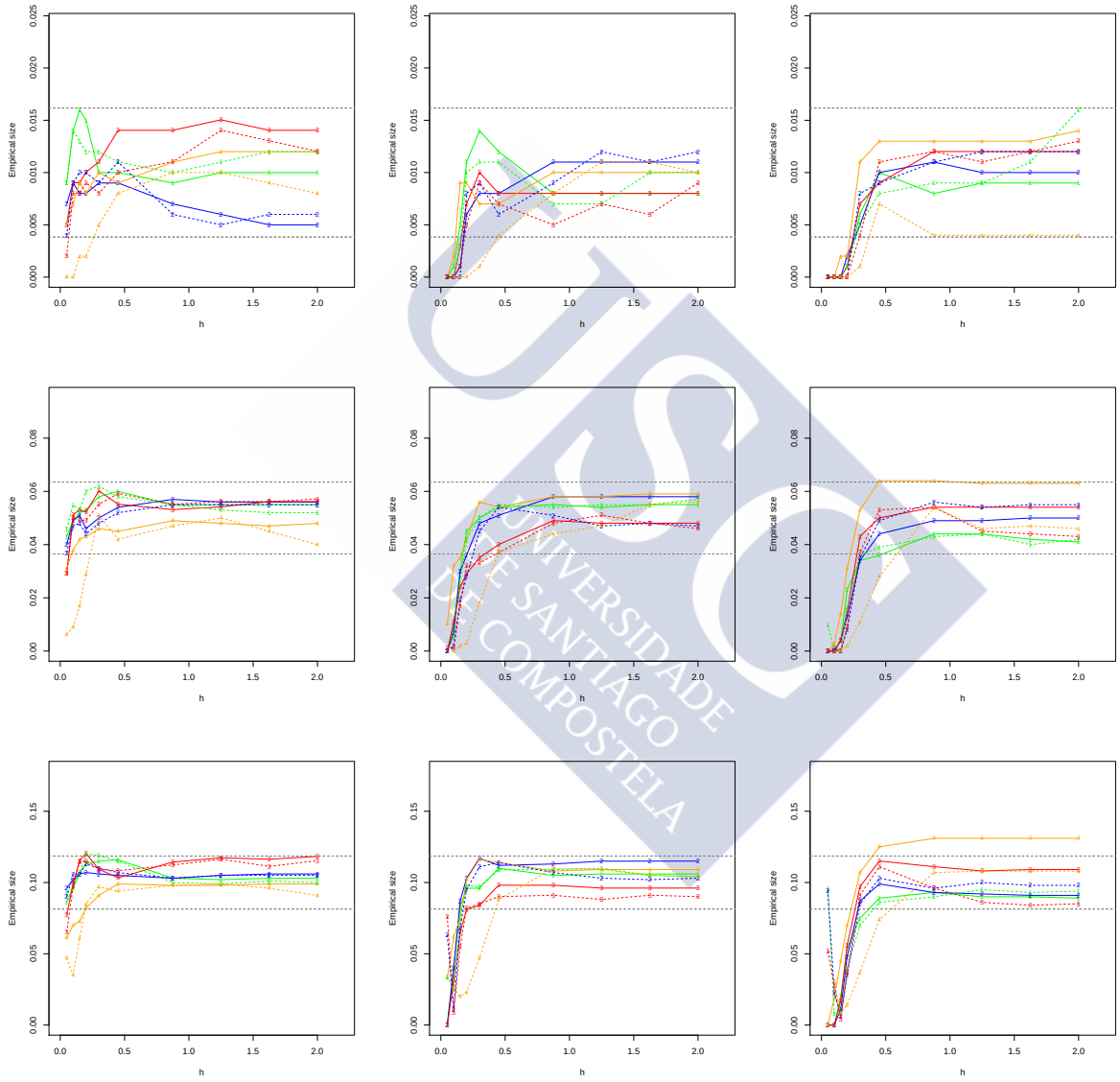


Figure B.6: Empirical sizes for $\alpha = 0.01$ (first row), $\alpha = 0.05$ (second row) and $\alpha = 0.10$ (third row) for the different scenarios, with $p = 0$ (solid line) and $p = 1$ (dashed line). From left to right, columns represent dimensions $q = 1, 2, 3$ with sample size $n = 500$.

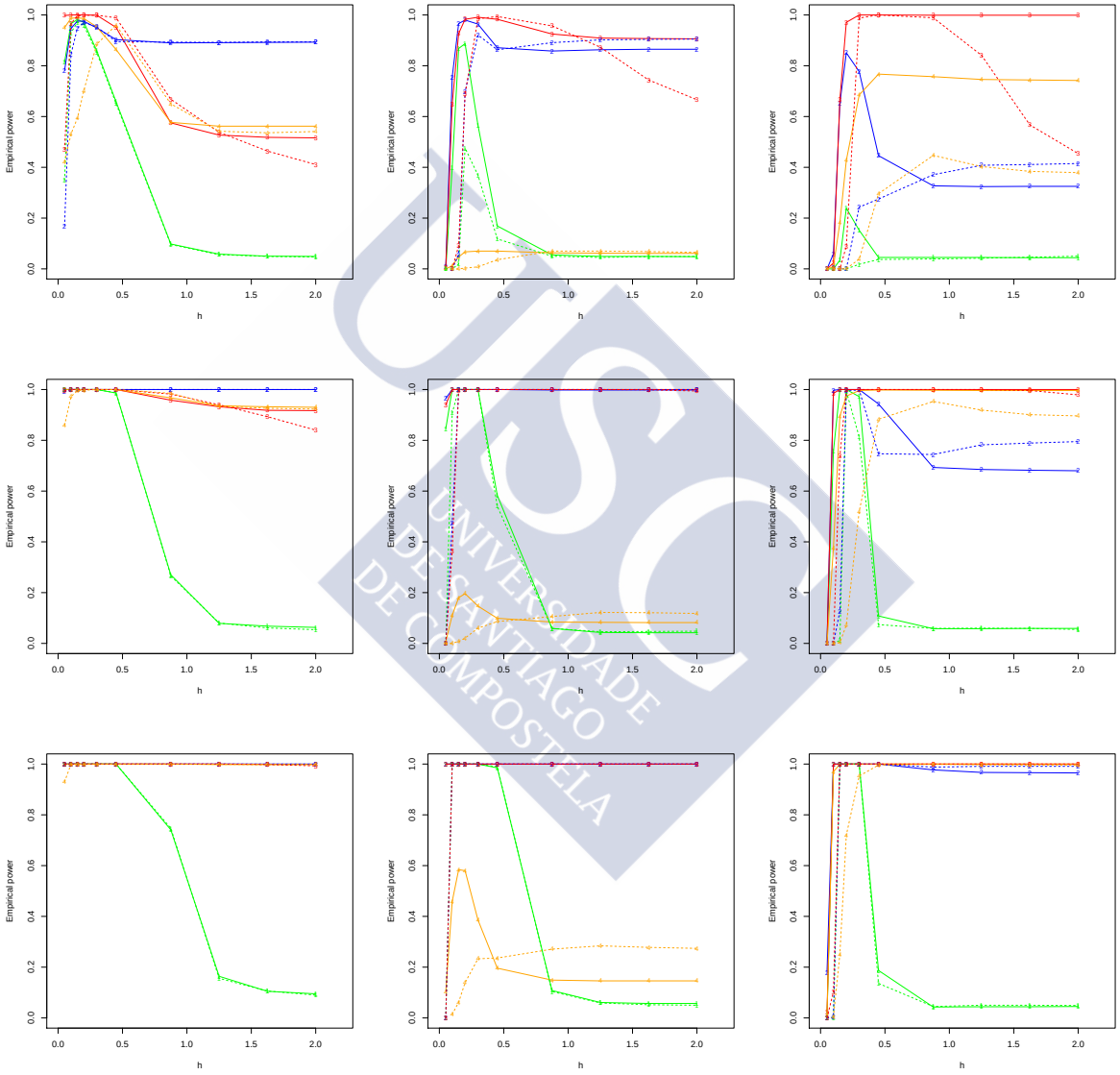


Figure B.7: Empirical powers for the different scenarios, with $p = 0$ (solid line) and $p = 1$ (dashed line). From top to bottom, rows represent sample sizes $n = 100, 250, 500$ and from left to right, columns represent dimensions $q = 1, 2, 3$.

References

Bai, Z. D., Rao, C. R., and Zhao, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.*, 27(1):24–39.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.*, 79(19):2066–2075.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2014). Nonparametric regression for spherical data. *J. Amer. Statist. Assoc.*, 109(506):748–763.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013). Kernel density estimation for directional-linear data. *J. Multivariate Anal.*, 121:152–175.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2014). Central limit theorems for directional and linear data with applications. *Statist. Sinica*, to appear.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Wang, X., Zhao, L., and Wu, Y. (2000). Distribution free laws of the iterated logarithm for kernel estimator of regression function based on directional data. *Chinese Ann. Math. Ser. B*, 21(4):489–498.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Resumen en castellano

¿Qué son los datos direccionales?

El término *datos direccionales* (*directional data* en inglés) fue introducido en el libro de Mardia (1972) para referirse al tipo de datos cuyo soporte se encuentra en una circunferencia, en una esfera o, en general, en una hiperesfera de dimensión arbitraria. Este tipo de datos aparece de forma natural en diversas disciplinas aplicadas: proteómica (ángulos en la estructura de las proteínas, ver por ejemplo Hamelryck et al. (2012)); ciencias medioambientales (dirección del viento (Johnson y Wehrly, 1978), dirección de las olas (Jona-Lasinio et al., 2012)); biología (orientación de animales, ver Batschelet (1981) para distintos ejemplos); fenómenos cíclicos (tiempos de llegada a una unidad médica (Fisher, 1993, página 239), estacionalidad en congelación y deshielo (Oliveira et al., 2013)); astronomía (posición de las estrellas, ver Secciones 1.2.8 y 1.5.3 de Perryman (1997)); análisis de imágenes (Dryden, 2005) o incluso en minería de textos (*text mining* en inglés; análisis de frecuencias de palabras en textos, ver por ejemplo Banerjee et al. (2005)). La colección de técnicas estadísticas diseñadas para analizar datos direccionales fue bautizada como *estadística para datos direccionales* (*directional statistics* en inglés) debido al libro homónimo de Mardia y Jupp (2000), una reedición revisada de Mardia (1972).

Las observaciones direccionales se representan como puntos en una hiperesfera euclídea de dimensión q, $\Omega_q = \{\mathbf{x} \in \mathbb{R}^{q+1} : ||\mathbf{x}|| = 1\}$ (también denotada por $\mathbb{S}^q$), donde los casos más sencillos se corresponden con la circunferencia ($q = 1$) y la esfera ($q = 2$). La inferencia con datos direccionales es, de hecho, *inferencia bajo restricciones*, ya que todos los métodos utilizados para el análisis estadístico deben de tener en cuenta la especial naturaleza de Ω_q, algo que no es requerido en el análisis de datos *lineales* (*i.e.* euclídeos). Un ejemplo pedagógico que ilustra este problema es la definición de una media direccional adecuada para la situación más sencilla: cuando se dispone de dos observaciones $\mathbf{X}_1$ y $\mathbf{X}_2$ en la circunferencia Ω_1. Un primer intento para definir esta media podría ser mediante la media euclídea $\bar{\mathbf{X}} = \frac{\mathbf{X}_1+\mathbf{X}_2}{2}$, pero el problema es que no se garantiza que $\bar{\mathbf{X}}$ pertenezca a Ω_1. Otra posibilidad consiste en considerar coordenadas polares, calcular la media usual de los correspondientes ángulos $\theta_1, \theta_2 \in [0, 2\pi)$ y después establecer la media direccional como el punto $(\cos\bar{\theta}, \sin\bar{\theta})$, donde $\bar{\theta} = \frac{\theta_1+\theta_2}{2}$. El problema con esta aproximación es que si, por ejemplo, $\theta_1 = \frac{\pi}{4}$ y $\theta_2 = \frac{7\pi}{4}$, entonces $\bar{\theta} = \pi$, lo que produce un resultado en la dirección opuesta a la media obvia, que se corresponde con $\bar{\theta} = 0$. Una definición razonable para la media direccional se obtiene mediante $\frac{\bar{\mathbf{X}}}{||\bar{\mathbf{X}}||}$ (si $\bar{\mathbf{X}} \neq 0$, en otro caso no está definida), véase Mardia y Jupp (2000) para más detalles.

En la literatura estadística se han considerado principalmente dos enfoques para analizar datos direccionales, dependiendo del tipo de representación empleada. El primero se basa en las coordenadas polares y esféricas para desarrollar métodos específicamente diseñados para tratar

datos circulares y esféricos, respectivamente, los más comunes en la práctica. Este es el enfoque seguido en los libros de Fisher (1993), Jammalamadaka y SenGupta (2001) y Pewsey et al. (2013) para datos circulares y de Fisher et al. (1993) para esféricos. Desafortunadamente, las extensiones de estos métodos a una dimensión arbitraria q no son sencillas debido a la naturaleza de las coordenadas esféricas en dimensiones superiores. El segundo enfoque hace uso de las coordenadas Cartesianas de los puntos en Ω_q, sin asumir ninguna dimensión en particular y por tanto asegurando una mayor generalidad. Esta es la aproximación seguida en esta tesis, excepto en el Capítulo 2, donde se hace uso de las coordenadas polares.

Quizás la distribución direccional más conocida sea la densidad *von Mises-Fisher* (ver Watson (1983) y Mardia y Jupp (2000)), también llamada simplemente von Mises. La densidad von Mises, denotada por $\text{vM}(\boldsymbol{\mu}, \kappa)$ (o por $\text{vM}(\mu, \kappa)$, si $q = 1$ y $\boldsymbol{\mu} = (\cos\mu, \sin\mu)$), viene dada por

$$f_{\text{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = C_q(\kappa) \exp\left\{\kappa \mathbf{x}^T \boldsymbol{\mu}\right\}, \quad C_q(\kappa) = \frac{\kappa^{\frac{q-1}{2}}}{(2\pi)^{\frac{q+1}{2}} \mathcal{I}_{\frac{q-1}{2}}(\kappa)},$$

siendo $\boldsymbol{\mu} \in \Omega_q$ la media direccional, $\kappa \geq 0$ el parámetro de concentración alrededor de la media ($\kappa = 0$ proporciona la distribución uniforme en Ω_q) y $\mathcal{I}_\nu$ la función modificada de Bessel de primera especie y orden ν, la cuál se puede escribir como (ver ecuación 10.32.2 de Olver et al. (2010))

$$\mathcal{I}_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\pi^{1/2}\Gamma\left(\nu + \frac{1}{2}\right)} \int_{-1}^{1} (1 - t^2)^{\nu - \frac{1}{2}} e^{zt}\, dt.$$

Esta distribución está considerada como la análoga de la gaussiana para datos direccionales por dos razones principales. Primero, presenta la misma caracterización en términos del Estimador de Máxima Verosimilitud (EMV) que la distribución gaussiana tiene en el caso euclídeo: es la única distribución direccional cuyo EMV del parámetro de localización es la media muestral direccional (ver Bingham y Mardia (1975) para una demostración). Segundo, la densidad von Mises se puede obtener a partir de un vector con distribución normal condicionado a que tenga norma unitaria. Esto es, para el vector aleatorio normal

$$\mathbf{X} \sim \mathcal{N}_{q+1}\left(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_{q+1}\right), \text{ con } \boldsymbol{\mu} \in \mathbb{R}^{q+1} \backslash \{\mathbf{0}\} \text{ y } \sigma^2 > 0,$$

se tiene que

$$\mathbf{Y} = (\mathbf{X} \,|\, ||\mathbf{X}|| = 1) \sim \text{vM}\left(\frac{\boldsymbol{\mu}}{||\boldsymbol{\mu}||}, \frac{||\boldsymbol{\mu}||}{\sigma^2}\right).$$

Este resultado muestra que el inverso del parámetro de concentración κ de una von Mises se puede identificar con la varianza de una normal multivariante cuya matriz de covarianzas sea proporcional a la identidad. Véase Gatto (2011) para una prueba de este resultado en una situación más general.

Otra distribución direccional destacable es la introducida por Jones y Pewsey (2005), la cual es denotada por $\text{JP}(\boldsymbol{\mu}, \kappa, \psi)$. Motivada originalmente para el caso circular, su densidad se puede definir también en Ω_q para una dimensión arbitraria q:

$$f_{\text{JP}}(\mathbf{x}; \boldsymbol{\mu}, \kappa, \psi) = \frac{|\text{senh}(\kappa\psi)|^{\frac{q-1}{2}} \left(\cosh(\kappa\psi) + \text{senh}(\kappa\psi)\mathbf{x}^T\boldsymbol{\mu}\right)^{\frac{1}{\psi}}}{2^{\frac{q-1}{2}}\Gamma\left(\frac{q+1}{2}\right) P^{-\frac{q-1}{2}}_{\frac{1}{\psi}+\frac{q-1}{2}}(\cosh(\kappa\psi))},$$

donde $\boldsymbol{\mu} \in \Omega_q$ es el parámetro de localización, $\kappa \geq 0$ es la concentración alrededor de $\boldsymbol{\mu}$, $\psi \in \mathbb{R}$ es el parámetro forma que controla una especie de curtosis negativa con respecto a una $\text{vM}(\boldsymbol{\mu}, \kappa)$ ($\psi < 0$ produce densidades más apuntadas mientras que las resultantes con $\psi > 0$ son más achatadas) y P_ν^μ es la función de Legendre de primera especie, orden μ y grado ν (ver ecuación 14.12.4 de Olver et al. (2010)). Esta familia paramétrica tiene la interesante propiedad de que contiene como casos particulares a la $\text{vM}(\boldsymbol{\mu}, \kappa)$ (correspondiente a $\psi \to 0$) y, con $q = 1$ y tomando coordenadas polares, la Cardioide ($\psi = 1$), la *Wrapped Cauchy* ($\psi = -1$) y la *Cartwright's power-of-cosine* ($\psi < 0$, $\kappa \to \infty$). Véase la Sección 2 de Jammalamadaka y SenGupta (2001) para más detalles sobre estas distribuciones.

En muchas situaciones las variables aleatorias direccionales aparecen junto a una variable lineal o direccional, siendo las variables circular-lineal (soporte en el cilindro $\Omega_1 \times \mathbb{R}$) y circular-circular (soporte en el toro $\Omega_1 \times \Omega_1$) las más comunes. En estos contextos, Johnson y Wehrly (1978) y Wehrly y Johnson (1979) propusieron un modelo semiparamétrico para la densidad conjunta:

$$f_{\Theta,X}(\theta, x) = 2\pi g\left(2\pi(F_\Theta(\theta) \pm F_X(x))\right) \times f_\Theta(\theta) f_X(x), \tag{B.15}$$

siendo Θ una variable circular con densidad f_Θ y función de distribución F_Θ y X una variable lineal o circular con f_X y F_X las funciones de densidad y distribución asociadas. g es una densidad circular que actúa como una función enlace entre las densidades marginales, dadas por f_Θ y f_X, considerando un signo positivo (dependencia negativa) o negativo (dependencia positiva) en $\pm$. g se puede interpretar en términos de cópulas (ver Nelsen (2006) para una introducción sobre el tema), ya que la densidad cópula de (Θ, X) está dada por $c_{\Theta,X}(u, v) = 2\pi g\left(2\pi(u \pm v)\right)$. De esta estructura semiparamétrica se pueden obtener distintos modelos paramétricos, como por ejemplo el modelo von Mises bivariante dado en Shieh y Johnson (2005).

Sin embargo, no todas las densidades paramétricas en este contexto satisfacen (1.1). Por ejemplo, la densidad circular-lineal de Mardia y Sutton (1978) (denotada por $\text{MS}(\mu, \kappa, m, \rho_1, \rho_2, \sigma)$) o la densidad circular-circular *Wrapped Normal Torus* dada en el Ejemplo 7.3 de Johnson y Wehrly (1977) (denotada por $\text{WNT}(m_1, m_2, \sigma_1, \sigma_2, \rho)$) son dos densidades paramétricas que no verifican la relación (1.1). Las expresiones de estas densidades son, respectivamente:

$$f_{\text{MS}}(\theta, x; \mu, \kappa, m, \sigma, \rho_1, \rho_2) = f_{\text{vM}}(\theta; \mu, \kappa) \times f_{\mathcal{N}}(z; m(\theta; \mu, \kappa, m, \sigma, \rho_1, \rho_2), \sigma(1 - \rho_1 - \rho_2)),$$

$$f_{\text{WNT}}(\theta, \psi; m_1, m_2, \sigma_1, \sigma_2, \rho) = \sum_{p_1=-\infty}^{\infty} \sum_{p_2=-\infty}^{\infty} f_{\mathcal{N}}(\theta + 2\pi p_1, \psi + 2\pi p_2; m_1, m_2, \sigma_1, \sigma_2, \rho),$$

donde $m(\theta; \mu, \kappa, m, \sigma, \rho_1, \rho_2) = m + \sigma\kappa^{\frac{1}{2}}\{\rho_1(\cos(\theta) - \cos(\mu)) + \rho_2(\text{sen}(\theta) - \text{sen}(\mu))\}$, $f_{\mathcal{N}}(\cdot; m, \sigma)$ es la densidad de una $\mathcal{N}(m, \sigma^2)$ y $f_{\mathcal{N}}(\cdot, \cdot; m_1, m_2, \sigma_1, \sigma_2, \rho)$ es la de una normal bivariante con vector media $(m_1, m_2)^T$, varianzas marginales σ_1^2 y σ_2^2 y coeficiente de correlación ρ. Estas distribuciones serán empleadas a lo largo de los diferentes capítulos. Por ejemplo, las dos últimas, entre otras, aparecen en el Capítulo 6 y la relación (1.1) es crucial para el Capítulo 2.

Por último, merece la pena destacar que los datos direccionales también pueden considerarse como casos particulares de espacios más generales, como ocurre por ejemplo en el *análisis estadístico de formas* (ver Dryden y Mardia (1998) y Kendall et al. (1999) para una revisión sobre el tema) o en la *estadística en variedades Riemannianas* (ver Bhattacharya y Bhattacharya (2012) y sus referencias).

Contribuciones de la tesis

Los métodos paramétricos han jugado un papel predominante en el desarrollo de la inferencia estadística para el análisis de datos direccionales (ver Mardia (1972) y Watson (1983)). Publicaciones surgidas después de estos libros, como Fisher (1993), Fisher et al. (1993), Mardia y Jupp (2000), Jammalamadaka y SenGupta (2001) y Pewsey et al. (2013) también se centraron fundamentalmente en el uso de técnicas paramétricas. Estos métodos asumen en mayor o menor medida que en el proceso estocástico generador de los datos se verifica una hipótesis paramétrica. Por ejemplo, la inferencia sobre la densidad desconocida de una variable aleatoria direccional se suele hacer mediante la asunción de una cierta densidad dependiente de parámetros desconocidos y que son estimados a partir de los datos. Mientras que este procedimiento lleva a resultados óptimos (en términos de eficiencia) si la hipótesis paramétrica es correcta, la estimación puede ser totalmente errónea si la hipótesis no se verifica.

Por otro lado, los métodos noparamétricos no requieren de fuertes hipótesis en el proceso estocástico generador, a excepción de ciertas condiciones de regularidad. La principal ventaja de los métodos noparamétricos es que siempre proporcionan soluciones razonables para la inferencia en general, al margen de supuestos paramétricos. Obviamente, un método noparamétrico no es óptimo en comparación con su competidor paramétrico diseñado *ad hoc* para un escenario paramétrico, pero aún así sigue siendo muy útil. Por ejemplo, la comparación de un ajuste paramétrico y uno noparamétrico da lugar a los conocidos *contrastes de bondad de ajuste*, que permiten contrastar de manera formal si una hipótesis paramétrica es plausible a partir de la información proporcionada por la muestra.

El propósito de esta tesis es el de proporcionar nuevas herramientas metodológicas para realizar inferencia noparamétrica con datos direccionales y lineales. Específicamente, se presentan métodos noparamétricos tanto para estimación como para contrastes de hipótesis, para las curvas de densidad y regresión, en situaciones donde están presentes variables aleatorias direccionales. Este planteamiento comprende los casos direccional, direccional-lineal y direccional-direccional. A continuación, se proporcionan breves *estados del arte* en cada una de estas temáticas junto con las contribuciones de la tesis, refiriéndose a los artículos que proporcionan los Capítulos 2–7 y que conforman el núcleo principal de este trabajo.

A. **Función de densidad**. Sean $\mathbf{X}$ e $\mathbf{Y}$ dos variables direccionales y Z una variable lineal. Sean $f_{\mathbf{X}}$, $f_{\mathbf{X},Z}$ y $f_{\mathbf{X},\mathbf{Y}}$ sus funciones de densidad direccional, direccional-lineal y direccional-direccional asociadas, respectivamente.

A.1. **Estimación**. El objetivo es la estimación de $f_{\mathbf{X}}$, $f_{\mathbf{X},Z}$ y $f_{\mathbf{X},\mathbf{Y}}$ mediante técnicas de suavizado tipo núcleo. La densidad de la cópula también se estima para los casos circular-lineal y circular-circular.

- **Estado del arte**. La estimación núcleo de la densidad $f_{\mathbf{X}}$ fue inicialmente considerada por Hall et al. (1987) y Bai et al. (1988), quienes establecieron las propiedades asintóticas básicas del estimador, y fue estudiada más tarde por Klemelä (2000). Taylor (2008) y Oliveira et al. (2012) propusieron selectores del ancho de banda para el caso circular, basándose la última referencia en la expresión asintótica del error establecida en Di Marzio et al. (2009). Para la estimación de las densidades $f_{\mathbf{X},Z}$ y $f_{\mathbf{X},\mathbf{Y}}$ con variables circulares, Fernández-Durán (2007) introdujo un método paramétrico considerando la estructura de la cópula dada

por Johnson y Wehrly (1978). Una estimación noparamétrica de la densidad de la cópula se propone en Carnicero et al. (2013) mediante el uso de polinomios de Bernstein.

- **Contribuciones**. En García-Portugués et al. (2013a) se proporciona un nuevo procedimiento para estimar densidades circular-lineales y circular-circulares a partir de la estimación de la estructura en forma de cópula de Johnson y Wehrly (1978). El método considera, entre otros, una modificación del estimador tipo núcleo de Gijbels y Mielniczuk (1990). Debido a que este procedimiento es difícilmente extensible a dimensiones superiores, en García-Portugués et al. (2013b) se propone un nuevo estimador núcleo la densidad $f_{\mathbf{X},Z}$, evitando realizar la estimación mediante cópulas. Se obtienen expresiones del error exactas tanto para los estimadores de la densidad $f_{\mathbf{X},Z}$ como para la densidad $f_{\mathbf{X}}$. Estas expresiones son el punto de partida de García-Portugués (2013) para establecer nuevos selectores del ancho de banda para el estimador núcleo de la densidad $f_{\mathbf{X}}$.

A.2. **Contrastes de hipótesis**. Los dos objetivos son: *i*) contrastar si $\mathbf{X}$ y Z son independientes, *i.e.*, contrastar si $H_0 : f_{\mathbf{X},Z}(\cdot,\cdot) = f_{\mathbf{X}}(\cdot)f_Z(\cdot)$ se verifica; *ii*) contrastar si $f_{\mathbf{X},Z}$ tiene una forma paramétrica particular, *i.e.*, si $H_0 : f_{\mathbf{X},Z} \in \{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ se verifica. De forma similar, con $f_{\mathbf{X},\mathbf{Y}}$ en vez de $f_{\mathbf{X},Z}$.

- **Estado del arte**. Hasta lo que el autor conoce, el único contraste de bondad de ajuste para densidades paramétricas direccionales fue propuesto en Boente et al. (2014). Este contraste está basado en el Teorema Central del Límite (TCL) de Zhao y Wu (2001) para el Error Cuadrático Integrado (ECI) del estimador núcleo de la densidad para $f_{\mathbf{X}}$. Estos artículos se pueden considerar como los análogos direccionales de Fan (1994) (o Bickel y Rosenblatt (1973)) y Hall (1984), respectivamente. Existen también test basados en coeficientes de correlación para detectar asociación circular-lineal, como los propuestos por Mardia (1976), Johnson y Wehrly (1977) y Fisher y Lee (1981).
- **Contribuciones**. En García-Portugués et al. (2014a) se presenta un test para contrastar la independencia entre una variable direccional y otra lineal (que también es adaptable al caso direccional-direccional). El estadístico de contraste se puede ver como un análogo de Rosenblatt y Wahlen (1992) (o Rosenblatt (1975)), ya que considera la distancia cuadrática entre el estimador de $f_{\mathbf{X},Z}$ y el producto de los estimadores de $f_{\mathbf{X}}$ y f_Z. El TCL del ECI para los estimadores de $f_{\mathbf{X},Z}$ y $f_{\mathbf{X},\mathbf{Y}}$ se obtiene en García-Portugués et al. (2014b), resultando fundamental para obtener la distribución asintótica del test de independencia y también la de los test de bondad de ajuste para densidades paramétricas direccional-lineales y direccional-direccionales. Además, se prueba la consistencia de un método de remuestreo boootstap para el calibrado del test.

B. **Función de regresión**. Sea $\mathbf{X}$ una variable aleatoria direccional e Y y ε dos variables lineales. Se asume que $\mathbf{X}$ e Y se relacionan mediante el modelo de regresión $Y = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$, con $m(\cdot) = \mathbb{E}[Y|\mathbf{X} = \cdot]$ y $\sigma^2(\cdot) = \mathbb{V}\mathrm{ar}[Y|\mathbf{X} = \cdot]$.

B.1. **Estimación**. El objetivo es la estimación de m mediante suavizado tipo núcleo utilizando un estimador local lineal.

- **Estado del arte**. Una adaptación del estimador de Nadaraya-Watson para la función de regresión m fue dado por Wang et al. (2000), donde se obtuvo su ley del

logaritmo iterado. Di Marzio et al. (2009) presentaron un estimador polinómico local para m en el caso el que el predictor fuese circular. Esta aproximación fue considerada más tarde por Di Marzio et al. (2013) para la regresión con respuesta circular. En Di Marzio et al. (2014) se propuso un estimador local lineal diferente con predictor o respuesta direccional basado en expansiones de Taylor construídas a partir de la descomposición tangente normal. Una definición anterior del estimador local con respuesta direccional y predictor lineal apareció en Boente y Fraiman (1991).

- **Contribuciones**. En García-Portugués et al. (2014) se introduce un estimador proyectado local lineal para la función de regresión m. El estimador está motivado por una expansión de Taylor modificada y diseñada para evitar la sobreparametrización que surge al considerar el estimador local lineal clásico. Se obtienen expresiones para el sesgo y la varianza del estimador, así como su normalidad asintótica y su formulación en términos del núcleo equivalente. Casos particulares de este estimador incluyen a Wang et al. (2000) y al estimador local lineal con predictor circular de Di Marzio et al. (2009).

B.2. **Contrastes de hipótesis**. El objetivo es contrastar si m pertenece a una clase de funciones de regresión paramétricas, *i.e.*, si la hipótesis $H_0 : m \in \{m_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ se verifica.

- **Estado del arte**. Deschepper et al. (2008) propusieron un test para la significación de una variable lineal sobre un predictor circular, el cual es, hasta lo que el autor conoce, el único test noparamétrico en el contexto de la regresión con datos direccionales. Un mecanismo de remuestreo para la calibración del estadístico se propone también en la citada referencia.
- **Contribuciones**. En García-Portugués et al. (2014) se presenta un contraste de bondad de ajuste para modelos de regresión paramétricos con predictor direccional y respuesta lineal. El test usa el estimador proyectado local lineal para construir un estadístico que mide la distancia cuadrática entre el estimador noparamétrico y una suavización del paramétrico (similar al de Härdle y Mammen (1993)), usando ajustes locales constantes o lineales. Se obtienen la distribución asintótica del estadístico y su potencia frente a alternativas locales, además de presentarse un método de remuestreo consistente.

Conjuntos de datos reales

A lo largo de esta tesis se han considerado diferentes conjuntos de datos para motivar e ilustrar las nuevas metodologías. En esta sección se proporciona una exhaustiva descripción de los conjuntos de datos reales, la mayoría de ellos originales (excepto los ángulos de las proteínas y parte de los datos relativos a incendios). Colecciones de conjuntos de datos direccionales clásicos se encuentran disponibles en los libros de Fisher (1993), Fisher et al. (1993) y Mardia y Jupp (2000).

Dirección del viento

La planta térmica de As Pontes (7° 51' 45" W, 43° 26' 26" N), localizada en el noroeste de España, es una de las centrales eléctricas con mayor capacidad de producción de todo el país.

La planta térmica es capaz de generar hasta 2200 megavatios de potencia, pero a expensas de una considerable emisión de polución a la atmósfera. A partir de 2005 se implementaron una serie de medidas para reducir las serias consecuencias que produce en el medio ambiente la lluvia ácida, surgida a partir de altas concentraciones de dióxido de azufre (SO_2). La medición de la concentración de los distintos contaminantes, incluido el SO_2, está controlada por una red de estaciones meteorológicas situadas alrededor de la planta. Entre las variables que se miden en las estaciones se encuentra la dirección en la cual el viento sopla, ya que juega un papel predominante en la diseminación de los polutos.

La aplicación a datos reales del Capítulo 2 está centrada en la relación entre la concentración de SO_2 y la dirección del viento en una estación de control localizada al noreste de la planta térmica (estación B2, 7° 44' 10" W, 43° 32' 05" N). El objetivo de la aplicación es chequear si el viento procedente de la planta térmica lleva asociadas concentraciones elevadas de SO_2 y la efectividad de las medidas de implementadas para reducir la contaminación. Para este propósito, dos conjuntos de datos fueron obtenidos para los meses de enero de 2004 y 2011 a partir de observaciones minutales en la estación B2. Para la obtención de la muestra finalmente analizada se realizaron los siguientes pasos: 1) las observaciones no disponibles fueron omitidas; 2) los datos fueron promediados por horas para reducir la dependencia temporal; 3) se aplicó una ligera perturbación para evitar datos repetidos que aparecían sistemáticamente debido a limitaciones en los aparatos medidores; 4) la muestra de SO_2 fue transformada usando una transformación de Box-Cox para mitigar su asimetría. Este procedimiento resulta en un par de conjuntos de datos con 736 y 743 observaciones para 2004 y 2011, respectivamente.

Por otro lado, la aplicación a datos reales del Capítulo 4 incluye el estudio de la dirección del viento en una estación cercana a la planta (estación de A Mourela, 7° 51' 21,91" W, 43° 25' 52,35" N), con el fin de determinar las direcciones en las cuales los polutos tienen más probabilididad de ser transportados. La adquisición de estos datos fue realizada según los pasos anteriores pero midiendo solamente la dirección del viento, omitiendo los pasos 3) y 4) y considerando datos de junio de 2012.

Conjunto de datos Hipparcos

La misión de astrometría Hipparcos fue llevada a cabo por la Agencia Espacial Europea entre 1989–1993 para localizar la posición de más de cien mil estrellas. Esta masiva enumeración de estrellas fue recogida en el catálogo Hipparcos (Perryman, 1997) y hecha accesible públicamente. Una década después, una nueva revisión de los datos originales fue llevada a cabo por van Leeuwen (2007). Durante este período, los avances en las técnicas de medición permitieron determinar con un mayor grado de precisión la posición exacta del satélite durante la misión. En consecuencia, la nueva revisión de los datos presenta una mejora significativa en la fiabilidad del catálogo astrométrico. Este es el conjunto de datos que ha sido considerado en la aplicación del Capítulo 4 y que puede ser descargado desde el servicio de catálogos astronómicos VizieR (Ochsenbein et al., 2000). El número de estrellas en el conjunto de datos es de $n = 117955$.

Debido a que las estrellas son objetos en continuo movimiento, en el catálogo Hipparcos las mediciones fueron realizadas con respecto a la posición que ocupaban en un marco temporal común. Este concepto se conoce como *epoch* (Ep) en astronomía y fue fijado en el año mediano con respecto a la duración de la misión, Ep 1991,5. La posición de las estrellas se mide con

respecto a su proyección en la esfera celeste, *i.e.*, la posición en la superficie terrestre idealizada (perfectamente esférica) que surge de la intersección con la línea imaginaria que une el centro de la tierra con la estrella. Esta posición se parametriza por un par de ángulos (λ, β), $\lambda \in [-\pi, \pi)$, $\beta \in [0, \pi)$, de tal forma que

$$\begin{cases} x_1 = \cos\beta\cos\lambda, \\ x_2 = \cos\beta\,\mathrm{sen}\lambda, \\ x_3 = \mathrm{sen}\beta. \end{cases}$$

El centro de la esfera celeste se sitúa en el centro de la Vía Láctea y su ecuador corresponde al *plano galáctico*, es decir, al plano de rotación de la galaxia. Este sistema de coordenadas se conoce como *coordenadas galácticas* y se encuentra entre los más habituales en astronomía debido a su fácil interpretación.

La forma habitual de representar superficies esféricas en el catálogo Hipparcos (y en astronomía en general) es mediante la proyección de Aitoff. Esta transformación proyecta la superficie de la esfera en el interior de una elipse cuyo semieje mayor es el doble del semieje menor, de longitud R. El punto (x, y) dentro de la elipse está dado por

$$\begin{cases} x = -2R\cos\beta\,\mathrm{sen}\,(\lambda/2)\,/\sqrt{1+\cos\beta\cos\,(\lambda/2)}, \\ y = R\,\mathrm{sen}\beta/\sqrt{1+\cos\beta\cos\,(\lambda/2)}. \end{cases}$$

Esta proyección no conserva distancias pero sí preserva el área (las proporciones entre las áreas de las regiones en la esfera y las áreas de las regiones proyectadas se mantienen constantes). En el catálogo Hipparcos se muestra la proyección de Aitoff del histograma del número de estrellas (Figura 3.2.1). Una versión suavizada de esta representación se puede obtener reemplazando el histograma por un estimador núcleo para datos esféricos que tome en consideración la naturaleza continua de las observaciones. Esto es lo que se hace en el Capítulo 4 empleando un ancho de banda adecuado para el estimador núcleo.

Incendios en Portugal

Los Capítulos 5 y 6 analizan datos direccionales provenientes de las orientaciones principales de los incendios ocurridos en Portugal desde 1985 hasta 2005. Esta colección de datos contiene $n = 26870$ mediciones de perímetros de incendios junto a sus (log) áreas quemadas y fue obtenida a partir de las imágenes proporcionadas por satélites del tipo Landsat. Antes y después de la temporada de incendios se tomaron imágenes por satélite de la parte continental de Portugal y mediante su comparación se obtuvo un retrato fidedigno de los incendios ocurridos en dicha temporada. Los perímetros de los incendios ocurridos durante el año se extrajeron a partir de un procedimiento semi-automático que empieza con una supervisión clasificada de las imágenes y continúa con una edición manual (Barros et al., 2012). La unidad cartográfica más pequeña (en inglés *Minimum Mapping Unit*, MMU) que el satélite es capaz de medir es de 5 hectáreas, que, aunque no es capaz de capturar los incendios más pequeños, sí que permite registrar hasta el 90 % de la superficie total quemada.

Las cuencas juegan un papel importante en el estudio de los incendios y su orientación. Barros et al. (2012) delimitaron 102 cuencas en las que se agrupan los incendios, estudiando para cuales de ellas se muestra un alineamiento preferencial con la orientación de los incendios. La orientación de los perímetros de los distintos objetos (cuencas o incendios) se determina mediante la primera componente principal (PC1) obtenida a partir de los puntos que constituyen

la frontera del objeto, tanto en el espacio bidimensional definido por la latitud y longitud de cada uno de los vértices, como en un espacio tridimensional, tomando también en consideración la altitud. Por lo tanto, la PC1 corresponde a un eje que pasa a través del centro de masa de un objeto y que maximiza la varianza de los vértices proyectados, representados en $\mathbb{R}^2$ o en $\mathbb{R}^3$.

En el caso bidimensional, la orientación es un objeto *axial* (la orientación N/S es también S/N). Estas orientaciones se pueden codificar como una variable angular $\theta \in [0, \pi)$ con período π, con lo que 2θ es una variable circular usual. Con esta codificación, los ángulos 0, $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$ representan las orientaciones E/W, NE/SW, N/S y NW/SE, respectivamente. En el caso tridimensional, la orientación se representa por medio de un par de ángulos (θ, φ) usando coordenadas esféricas, donde $\theta \in [0, \pi)$ juega el mismo papel que en el caso anterior y $\varphi \in [0, \frac{\pi}{2}]$ mide la inclinación ($\Phi = \frac{\pi}{2}$ para pendientes planas y $\varphi = 0$ para las verticales; sólo se consideran los ángulos positivos ya que los negativos miden la misma pendiente). Por lo tanto, los puntos con coordenadas esféricas $(2\theta, \varphi)$, que yacen en la semiesfera superior, pueden ser considerados como realizaciones de una variable esférica.

En el Capítulo 5, la independencia entre orientaciones circulares o esféricas de los diferentes objetos y las áreas quemadas es contrastada sobre todo el conjunto de datos. El Capítulo 6 analiza una parte reducida de los datos obtenida considerando la orientación circular promedio de los incendios en cada una de las cuencas y la media de las áreas quemadas, proporcionando un conjunto de datos de tamaño $n = 102$. En dicho capítulo se aplica un contraste de bondad de ajuste para contrastar si un modelo paramétrico circular-lineal es adecuado para explicar el conjunto de datos.

Ángulos en las proteínas

Un campo científico donde la estadística de datos direccionales está llamada a jugar un papel importante es la proteómica. Las estructuras biomoleculares como las proteínas se expresan a menudo en términos de los ángulos dihédricos que describen las rotaciones de la estructura principal en torno a los enlaces entre los átomos N-C_α (ángulo ϕ) y C_α-C (ángulo Ψ). El diagrama de dispersión de estos pares de ángulos en una proteína, conocido como el *gráfico de Ramachadran*, proporciona una forma sencilla de ver las posibles torsiones de la misma. La distribución de los ángulos dihédricos y su modelado es un paso clave en el estudio del conocido como *problema del plegamiento de proteínas*, uno de los mayores problemas abiertos en biología en la actualidad. Véase Hamelryck et al. (2012) y sus referencias para más información sobre proteómica y sobre los métodos direccionales usados en el campo.

Los datos analizados en el Capítulo 6 contienen pares de ángulos dihédricos en la estructura primaria de 1932 proteínas. Estos datos fueron estudiados inicialmente en Fernández-Durán (2007) mediante el uso de la estructura cópula de Wehrly y Johnson (1979) y de modelos paramétricos para las densidades marginales y la función enlace. La idoneidad de estos modelos paramétricos se validará mediante un contraste de bondad de ajuste. El conjunto de datos está formado por 233 pares de ángulos de segmentos del tipo alanina-alanina-alanina en aminoácidos, extraídos a partir de una muestra representativa de 1932 proteínas obtenida de la lista recomendada de proteínas de julio de 2003 del Banco de Datos de Proteínas (Berman et al., 2000). El conjunto de datos está disponible en el objeto `ProteinsAAA` del paquete `CircNNTSR` (Fernández-Durán y Gregorio-Domínguez, 2013) de `R`.

Minería de textos

La estadística de datos direccionales tiene también interesantes aplicaciones en contextos de alta dimensión. Un buen ejemplo de ello es la minería de textos, donde los documentos son usualmente representados como vectores normalizados en Ω_{D-1} siendo D el número de palabras en un diccionario. Este concepto es el conocido *modelo de espacio vectorial* (*vector space model* en inglés): una colección de documentos (denominada *corpus*) $\mathbf{d}_1, \ldots, \mathbf{d}_n$ se representa por un conjunto de vectores $\{(d_{i1}, \ldots, d_{iD})\}_{i=1}^{n}$ (*matriz documento-término*) con respecto a un diccionario $\{w_1, \ldots, w_D\}$, tal que d_{ij} representa la frecuencia de la j-ésima palabra del diccionario en el documento d_i. En esta situación se requiere de alguna normalización para evitar desequilibrios entre documentos grandes y pequeños. Por ejemplo, en el caso donde un documento se crea copiando y pegando N veces otro documento ambos vectores tendrían la misma dirección, pero la longitud del primero sería N veces más grande, por tanto parecerían elementos diferentes en $\mathbb{R}^D$, aunque ambos contengan la misma información. Tomando la norma euclídea, $\mathbf{d}_i / ||\mathbf{d}_i|| \in \Omega_{D-1}$ y el corpus se puede ver como una muestra de datos direccionales.

La aplicación del Capítulo 7 tiene que ver con un corpus extraído del agregador de noticias *Slashdot* (`wwww.slashdot.org`): se contrasta la adecuación de un modelo lineal para explicar la popularidad de las noticias a partir de su contenido. Esta página web publica noticias/historias diarias que son enviadas y evaluadas por usuarios. La web está principalmente dedicada a la tecnología y a la ciencia, pero también cubre noticias relacionadas con política o con derechos digitales. Su estructura principal se puede entender como una serie de entradas de noticias, donde cada una incluye un título, un resumen corto de la noticia e información con respecto a su envío (usuario, fecha, palabras clave,...). Bajo el resumen se encuentra un enlace que muestra el número de comentarios en la sección de debate anexa a la entrada. Obviamente, el grado de participación en la sección de debate varía notablemente dependiendo de la temática de la noticia. Por ejemplo, noticias relacionadas con política o con software propietario tienden a ser más controvertidas y generar más discusión, mientras que noticias con un contenido científico muy específico a menudo presentan un número menor de comentarios.

El conjunto de datos fue obtenido siguiendo los siguientes pasos. Primero, los títulos, resúmenes y números de comentarios en cada una de las noticias aparecidas en 2013 fueron descargados automáticamente *parseando* el archivo de noticias, obteniéndose una colección de $n = 8121$ documentos. Después de eso, los siguientes pasos fueron llevados a cabo utilizando el paquete `tm` (Meyer et al., 2008) de minería de textos en `R` y código autoprogramado: 1) fusionar los títulos y los resúmenes en un mismo documento, omitiendo los detalles del envío; 2) eliminación del código `HTML`; 3) conversión del texto a minúsculas; 4) eliminación de palabras vacías (en inglés *stop words*; obtenidas de las listas de stop words en inglés incluidas en `tm` y `MySQL`), puntuación, espacios y números; 5) *stemming* de palabras para reducirlas a su raíz; 6) eliminación de palabras demasiado raras o frecuentes (más del 50 % de las palabras procesadas sólo aparecían en un único documento). El último paso fue dado considerando sólo las palabras que aparecían ente 58 y 1096 documentos, resultando $D = 1508$. Estas cantidades corresponden a los cuantiles 95 % y 99,95 % de la distribución empírica de la *frecuencia de documento* (*i.e.*, el número de documentos que contienen una palabra en particular). Finalmente, el corpus fue guardado como una matriz documento-término usando el diccionario formado por las D palabras seleccionadas.

Organización de la memoria

El núcleo principal de esta tesis son los Capítulos 2 a 7. Cada uno de ellos contiene a un artículo original sobre un tema específico relacionado con el tema conductor de la tesis: inferencia noparamétrica con datos direccionales y lineales. Por lo tanto, cada capítulo se presenta como un artículo autocontenido con su abstract, secciones, apéndices y referencias propias, justo en la misma forma en la que fue publicado, aceptado para su publicación o sometido. La referencia del artículo correspondiente se incluye al comienzo de cada capítulo. En el momento de la presentación de este manuscrito, los artículos de los Capítulos 2–5 han sido publicados, el artículo del Capítulo 6 ha sido aceptado y el del Capítulo 7 ha sido sometido para su publicación. A continuación se presentan breves resúmenes de los capítulos, junto con sus artículos asociados.

Capítulo 1: Introducción. En este primer capítulo se presenta una introducción al campo de la estadística con datos direccionales. Se describe el estado del arte y las referencias principales en los temas en los que la tesis aporta nuevas contribuciones, incluyendo un diagrama explicativo de dichas contribuciones. También se describen los conjuntos de datos utilizados a lo largo del manuscrito y la estructura de la tesis.

Capítulo 2: Estimación de la densidad mediante cópulas circular-lineales (García-Portugués et al., 2013a). Este capítulo presenta diferentes enfoques de la estimación de la densidad circular-lineal o circular-circular mediante el uso de cópulas y la estructura de dependencia dada por Johnson y Wehrly (1978). Para estudiar la relación entre la dirección del viento y la concentración de SO_2 en una estación meteorológica cercana a la central térmica de As Pontes se utiliza un método noparamétrico. El método se aplica para evaluar la eficacia de las medidas de reducción de emisiones llevadas a cabo en la central.

Capítulo 3: Estimación núcleo de la densidad con datos direccional-lineales (García-Portugués et al., 2013b). Una alternativa natural al método noparamétrico presentado en el capítulo anterior es un estimador núcleo de la densidad direccional-lineal directamente aplicado en los datos, *i.e.* sin requerir el uso de funciones cópulas. En este capítulo se presenta dicho estimador, proporcionando resultados para el sesgo, la varianza y la normalidad asintótica, entre otros. Las expresiones del error exacto son obtenidas para el estimador núcleo de la densidad direccional-lineal pero también para el direccional, sentando las bases para el Capítulo 4.

Capítulo 4: Selectores del ancho de banda para la estimación núcleo de la densidad con datos direccionales (García-Portugués, 2013). A partir de las expresiones del error asintóticas y exactas dadas en el Capítulo 3, se proponen tres nuevos selectores del ancho de banda. El primer selector es un análogo natural del selector circular dado en Taylor (2008), mientras que los otros dos surgen a partir de la combinación de mixturas de densidades von Mises con los criterios de error asintóticos y exactos. El funcionamiento de los selectores se compara en un extenso estudio de simulación, ilustrando el mejor de ellos con los conjuntos de dirección del viento y del satélite Hipparcos.

Capítulo 5: Un test no paramétrico para la independencia direccional-lineal (García-Portugués et al., 2014a). Usando el estimador dado en el Capítulo 3, se construye un test de independencia basado en la distancia cuadrática entre el estimador núcleo conjunto y el producto de los estimadores núcleo marginales direccional y lineal. Se obtiene una expresión cerrada

para el estadístico y se utiliza un método de remuestreo basado en permutaciones para calibrar el estadístico en la práctica. El funcionamiento del test se analiza en un estudio de simulación bajo una variedad de situaciones y se aplica para estudiar la influencia de la orientación de los incendios sobre su tamaño en el conjunto de datos sobre incendios en Portugal.

Capítulo 6: Teoremas centrales del límite para variables direccionales y lineales con aplicaciones (García-Portugués et al., 2014b). Este capítulo está dedicado a obtener un TCL para el ECI del estimador del Capítulo 3. El resultado es usado para establecer la convergencia en distribución del test de independencia dado en el Capítulo 5 y para obtener un nuevo contraste de bondad de ajuste para familias paramétricas de densidades direccional-lineales y direccional-direccionales. Se propone un método de remuestreo bootstrap consistente para calibrar el contraste de bondad de ajuste y su funcionamiento se ilustra en un extenso estudio de simulación. El test se aplica a los conjuntos de datos sobre los ángulos de las proteínas y los incendios en Portugal.

Capítulo 7: Contraste de modelos paramétricos para la regresión direccional-lineal (García-Portugués et al., 2014). Se propone un nuevo estimador local lineal para estimar la función de regresión con predictor direccional y respuesta lineal, estableciendo sus distintas propiedades como sesgo, varianza y normalidad asintótica. Basándose en este estimador, se construye un contraste de bondad de ajuste para la hipótesis nula de que la función de regresión desconocida pertenece a una cierta familia paramétrica. Se obtiene la distribución asintótica del estadístico, junto con la potencia para alternativas locales. También se proporciona un algoritmo bootstrap consistente para la calibración práctica del mismo. El test se ilustra en un estudio de simulación y se aplica al conjunto de datos de Slashdot.

Capítulo 8: Investigación futura. En este capítulo se esquematizan distintas ideas sobre proyectos futuros: nuevos selectores del ancho de banda para la regresión noparamétrica lineal-direccional, un estimador núcleo para datos direccionales bajo simetría rotacional, un paquete de `R` implementando los métodos descritos en la tesis y un contraste de bondad de ajuste para la estructura de cópula de Johnson y Wehrly (1978).

Apéndice A: Suplemento al Capítulo 6. Este suplemento contiene las pruebas de los lemas técnicos usados en el Capítulo 6, detalles exhaustivos sobre el estudio de simulación, resultados adicionales sobre el test de independencia y una extensión de la aplicación a datos reales.

Apéndice B: Suplemento al Capítulo 7. Casos particulares del estimador local lineal, las pruebas de los lemas técnicos y resultados adicionales para el estudio de simulación se incluyen en este apéndice.

Bibliografía del resumen

Bai, Z. D., Rao, C. R., y Zhao, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.*, 27(1):24–39.

Banerjee, A., Dhillon, I. S., Ghosh, J., y Sra, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions. *J. Mach. Learn. Res.*, 6:1345–1382.

Barros, A. M. G., Pereira, J. M. C., y Lund, U. J. (2012). Identifying geographical patterns of wildfire orientation: a watershed-based analysis. *Forest. Ecol. Manag.*, 264:98–107.

Batschelet, E. (1981). *Circular statistics in biology*. Academic Press, Inc., London-New York.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., y Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Res.*, 28(1):235–242.

Bhattacharya, A. y Bhattacharya, R. (2012). *Nonparametric inference on manifolds*, volumen 2 de *Institute of Mathematical Statistics Monographs*. Cambridge University Press, Cambridge.

Bickel, P. J. y Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.*, 1(6):1071–1095.

Bingham, M. S. y Mardia, K. V. (1975). Maximum likelihood characterization of the von Mises distribution. En Patil, G. P., Kotz, S., y Ord, J. K., editores, *A modern course on statistical distributions in scientific work*, volumen 17 de *NATO Advanced Study Institutes Series. Series C, Mathematical and Physical Sciences.*, páginas 387–398. D. Reidel, Dordrecht.

Boente, G. y Fraiman, R. (1991). Nonparametric regression for directional data. *Trabajos de Matemática*, 176.

Boente, G., Rodríguez, D., y González-Manteiga, W. (2014). Goodness-of-fit test for directional data. *Scand. J. Stat.*, 41(1):259–275.

Carnicero, J. A., Ausín, M. C., y Wiper, M. P. (2013). Non-parametric copulas for circular-linear and circular-circular data: an application to wind directions. *Stoch. Environ. Res. Risk Assess.*, 27(8):1991–2002.

Deschepper, E., Thas, O., y Ottoy, J. P. (2008). Tests and diagnostic plots for detecting lack-of-fit for circular-linear regression models. *Biometrics*, 64(3):912–920.

Di Marzio, M., Panzera, A., y Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.*, 79(19):2066–2075.

Di Marzio, M., Panzera, A., y Taylor, C. C. (2013). Non-parametric regression for circular responses. *Scand. J. Stat.*, 40(2):238–255.

Di Marzio, M., Panzera, A., y Taylor, C. C. (2014). Nonparametric regression for spherical data. *J. Amer. Statist. Assoc.*, 109(506):748–763.

Dryden, I. L. (2005). Statistical analysis on high-dimensional spheres and shape spaces. *Ann. Statist.*, 33(4):1643–1665.

Dryden, I. L. y Mardia, K. V. (1998). *Statistical shape analysis*. Wiley Series in Probability and Statistics. Probability and Statistics. John Wiley & Sons, Chichester.

Fan, Y. (1994). Testing the goodness of fit of a parametric density function by kernel method. *Economet. Theor.*, 10(2):316–356.

Fernández-Durán, J. J. (2007). Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics*, 63(2):579–585.

Fernández-Durán, J. J. y Gregorio-Domínguez, M. M. (2013). *CircNNTSR: an R package for the statistical analysis of circular data using nonnegative trigonometric sums (NNTS) models.* R package version 2.1.

Fisher, N. I. (1993). *Statistical analysis of circular data.* Cambridge University Press, Cambridge.

Fisher, N. I. y Lee, A. J. (1981). Nonparametric measures of angular-linear association. *Biometrika*, 68(3):629–636.

Fisher, N. I., Lewis, T., y Embleton, B. J. J. (1993). *Statistical analysis of spherical data.* Cambridge University Press, Cambridge.

García-Portugués, E. (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron. J. Stat.*, 7:1655–1685.

García-Portugués, E., Barros, A. M. G., Crujeiras, R. M., González-Manteiga, W., y Pereira, J. (2014a). A test for directional-linear independence, with applications to wildfire orientation and size. *Stoch. Environ. Res. Risk Assess.*, 28(5):1261–1275.

García-Portugués, E., Crujeiras, R. M., y González-Manteiga, W. (2013a). Exploring wind direction and SO_2 concentration by circular-linear density estimation. *Stoch. Environ. Res. Risk Assess.*, 27(5):1055–1067.

García-Portugués, E., Crujeiras, R. M., y González-Manteiga, W. (2013b). Kernel density estimation for directional-linear data. *J. Multivariate Anal.*, 121:152–175.

García-Portugués, E., Crujeiras, R. M., y González-Manteiga, W. (2014b). Central limit theorems for directional and linear data with applications. *Statist. Sinica*, aceptado.

García-Portugués, E., Van Keilegom, I., Crujeiras, R. M., y González-Manteiga, W. (2014). Testing parametric models in linear-directional regression. *arXiv:1409.0506*.

Gatto, R. (2011). The generalized von Mises-Fisher distribution. En Wells, M. T. y SenGupta, A., editores, *Advances in directional and linear statistics*, páginas 51–68. Physica, Heidelberg.

Gijbels, I. y Mielniczuk, J. (1990). Estimating the density of a copula function. *Comm. Statist. Theory Methods*, 19(2):445–464.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.*, 14(1):1–16.

Hall, P., Watson, G. S., y Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762.

Hamelryck, T., Mardia, K., y Ferkinghoff-Borg, J., editores (2012). *Bayesian methods in structural bioinformatics.* Statistics for Biology and Health. Springer, Berlin.

Härdle, W. y Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, 21(4):1926–1947.

Jammalamadaka, S. R. y SenGupta, A. (2001). *Topics in circular statistics*, volumen 5 de *Series on Multivariate Analysis.* World Scientific Publishing, River Edge.

Johnson, R. A. y Wehrly, T. (1977). Measures and models for angular correlation and angular-linear correlation. *J. Roy. Statist. Soc. Ser. B*, 39(2):222–229.

Johnson, R. A. y Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *J. Amer. Statist. Assoc.*, 73(363):602–606.

Jona-Lasinio, G., Gelfand, A., y Jona-Lasinio, M. (2012). Spatial analysis of wave direction data using wrapped gaussian processes. *Ann. Appl. Stat.*, 6(4):1478–1498.

Jones, M. C. y Pewsey, A. (2005). A family of symmetric distributions on the circle. *J. Amer. Statist. Assoc.*, 100(472):1422–1428.

Kendall, D. G., Barden, D., Carne, T. K., y Le, H. (1999). *Shape and shape theory*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester.

Klemelä, J. (2000). Estimation of densities and derivatives of densities with directional data. *J. Multivariate Anal.*, 73(1):18–40.

Mardia, K. V. (1972). *Statistics of directional data*, volumen 13 de *Probability and Mathematical Statistics*. Academic Press, London.

Mardia, K. V. (1976). Linear-circular correlation coefficients and rhythmometry. *Biometrika*, 63(2):403–405.

Mardia, K. V. y Jupp, P. E. (2000). *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, segunda edición.

Mardia, K. V. y Sutton, T. W. (1978). A model for cylindrical variables with applications. *J. Roy. Statist. Soc. Ser. B*, 40(2):229–233.

Meyer, D., Hornik, K., y Feinerer, I. (2008). Text mining infrastructure in R. *J. Stat. Softw.*, 25(5):1–54.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, segunda edición.

Ochsenbein, F., Bauer, P., y Marcout, J. (2000). The VizieR database of astronomical catalogues. *Astron. Astrophys. Supp. Ser.*, 143(23):23–32.

Oliveira, M., Crujeiras, R. M., y Rodríguez-Casal, A. (2012). A plug-in rule for bandwidth selection in circular density estimation. *Comput. Statist. Data Anal.*, 56(12):3898–3908.

Oliveira, M., Crujeiras, R. M., y Rodríguez-Casal, A. (2013). Nonparametric circular methods for exploring environmental data. *Environ. Ecol. Stat.*, 20(1):1–17.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F., y Clark, C. W., editores (2010). *NIST handbook of mathematical functions*. Cambridge University Press, Cambridge.

Perryman, M. A. C. (1997). *The Hipparcos and Tycho catalogues*, volumen 1200 de *ESA SP*. ESA Publication Division, Noordwijk.

Pewsey, A., Neuhäuser, M., y Ruxton, G. D. (2013). *Circular statistics in R*. Oxford University Press, Oxford.

Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.*, 3(1):1–14.

Rosenblatt, M. y Wahlen, B. E. (1992). A nonparametric measure of independence under a hypothesis of independent components. *Statist. Probab. Lett.*, 15(3):245–252.

Shieh, G. S. y Johnson, R. A. (2005). Inferences based on a bivariate distribution with von Mises marginals. *Ann. Inst. Statist. Math.*, 57(4):789–802.

Taylor, C. C. (2008). Automatic bandwidth selection for circular density estimation. *Comput. Statist. Data Anal.*, 52(7):3493–3500.

van Leeuwen, F. (2007). *Hipparcos, the new reduction of the raw data*, volumen 350 de *Astrophysics and Space Science Library*. Springer, Dordrecht.

Wang, X., Zhao, L., y Wu, Y. (2000). Distribution free laws of the iterated logarithm for kernel estimator of regression function based on directional data. *Chinese Ann. Math. Ser. B*, 21(4):489–498.

Watson, G. S. (1983). *Statistics on spheres*, volumen 6 de *University of Arkansas Lecture Notes in the Mathematical Sciences*. John Wiley & Sons, New York.

Wehrly, T. E. y Johnson, R. A. (1979). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, 67(1):255–256.

Zhao, L. y Wu, C. (2001). Central limit theorem for integrated square error of kernel estimators of spherical density. *Sci. China Ser. A*, 44(4):474–483.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA