ESCOLA DE DOUTORAMENTO
INTERNACIONAL DA USC

María
Alonso Pena

Tese de doutoramento

New approaches to nonparametric circular regression models

Santiago de Compostela, 2022

Programa de Doutoramento en Estatística e Investigación Operativa

TESE DE DOUTORAMENTO

NEW APPROACHES TO NONPARAMETRIC CIRCULAR REGRESSION MODELS

María Alonso Pena

ESCOLA DE DOUTORAMENTO INTERNACIONAL
DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOUTORAMENTO
EN ESTATÍSTICA E INVESTIGACIÓN OPERATIVA

SANTIAGO DE COMPOSTELA
2022

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DECLARACIÓN DA AUTORA DA TESE

Dna. **María Alonso Pena**
Título da tese: **New approaches to nonparametric circular regression models**

Presento a miña tese, seguindo o procedemento axeitado ao Regulamento, e declaro que:

1) A tese abarca os resultados da elaboración do meu traballo.
2) De ser o caso, na tese faise referencia ás colaboracións que tivo este traballo.
3) Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.
4) A tese é a versión definitiva presentada para a súa defensa e coincide coa versión enviada en formato electrónico.

E comprométome a presentar o Compromiso Documental de Supervisión no caso de que o orixinal non estea na Escola.

En **Santiago de compostela, a 4 de xullo de 2022**

María Alonso Pena

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DOS DIRECTORES/TITOR DA TESE

New approaches to nonparametric circular regression models

Dna. Rosa M. Crujeiras Casais

INFORMA:

Que a presente tese se corresponde co traballo realizado por Dna. María Alonso Pena, baixo a miña dirección/titorización, e autorizo a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como directora/titora desta non incorre nas causas de abstención establecidas na Lei 40/2015.

De acordo co indicado no Regulamento de Estudos de Doutoramento, declara tamén que a presente tese de doutoramento é idónea para ser defendida en base á modalidade de Monográfica con reproducción de publicacións, nos que a participación da doutoranda foi decisiva para a súa elaboración e as publicacións se axustan ao Plan de Investigación.

En **Santiago de Compostela, a 4 de xullo de 2022**

Rosa M. Crujeiras Casais

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Agradecementos

Quero agradecer aquí a todas as persoas que fixeron, dunha forma ou doutra, que a realización desta tese fose posible. En primeiro lugar, aos meus compañeiros e compañeiras da sala π, Alejandra, Alfredo, Arís, Brais, Diego, Ignacio, Laura D., Laura F., María, Mohammad e Ruth, por aguantar as miñas queixas día si e día tamén, e por prestarme axuda, tanto profesional como persoal, sempre que a necesitei. Fago este agradecemento estensivo ao resto de compañeiros e compañeiras da Facultade de Matemáticas cos que tiven o pracer de compartir estes últimos anos.

Grazas tamén ao persoal do departamento, e en especial aos profesores cos que compartín docencia e que fixeron que esta fose unha gran experiencia.

To Irène Gijbels, for hosting me in Leuven during three months, for her crucial guidance and for serving as an inspiration to me.

A Jose, como colaborador deste traballo, que tivo bastante que ver en que esta tese se iniciase. Pero tamén como amigo, por aconsellarme, aturarme e por ter a porta do despacho sempre aberta para min.

A Lali, Andrea, Aldara e Alba, por estar sempre aí. A Érika e Paula, por facer que este último ano me sentise realmente en casa. E a Álex, por estar cando máis o necesitaba.

Á miña nai, meu pai, meu irmán e ao resto da familia, polo apoio durante estes anos. Moi especialmente á miña prima Adela, aos meus tíos Inés e Manolo e á Abuela María, por acollerme na súa casa durante o confinamento e permitirme seguir traballando neste proxecto. E a Breixo e a Cibrán, por axudarme a desconectar, xogando.

Pero, sen dúbida, a persoa á que máis lle teño que agradecer é Rosa. Grazas por todas as túas ideas, consellos, correcións, correos e por estar sempre dispoñible. Grazas por confiar en min dende o principio, por apoiarme, por entenderme e por animarme a ser mellor cada día. Grazas por guiarme, de xeito excelente, todos estes anos.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

María Alonso Pena
Santiago de Compostela, xullo de 2022

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Funding

Research of the author was supported by Xunta de Galicia through the grant ED481A-2019/139 from the Consellería de Educación, Universidade e Formación Profesional. Part of the research done in Chapters 3 and 4 was carried out during a visit to the Department of Mathematics of KU Leuven, financed by Xunta de Galicia. This work has also been supported by Project PID2020-116587GB-100 funded by MCIN/AEI/10.13039/501100011033, the Competitive Reference Groups 2021–2024 (ED431C 2021/24) from Xunta de Galicia and partially funded by project C16/20/002 of the Research Fund KU Leuven, Belgium. The Supercomputing Center of Galicia (CESGA) is acknowledged for providing the computational resources that allowed to run most of the simulations. Finally, Professor Felicita Scapini (University of Florence) and Professor Hiroto Ogawa (Hokkaido University) are thankfully acknowledged for providing some of the data sets employed in this dissertation.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Contents

Contents i

Abstract v

Methodology xi

1 Introduction 1
- 1.1 Background on kernel regression . . . 2
  - 1.1.1 Mean regression . . . 3
  - 1.1.2 Local likelihood regression . . . 6
  - 1.1.3 Multimodal regression . . . 7
- 1.2 Background on circular data . . . 8
  - 1.2.1 Population and sample measures . . . 9
  - 1.2.2 Circular distribution models . . . 11
  - 1.2.3 Regression for circular data . . . 14
- 1.3 Objectives and manuscript distribution . . . 18

2 Nonparametric inference for circular regression 21
- 2.1 Nonparametric estimation of the mean regression function . . . 22
  - 2.1.1 Circular covariate and real-valued response . . . 22
  - 2.1.2 Circular response . . . 24
- 2.2 A test of no-effect . . . 25
  - 2.2.1 Circular covariate and real-valued response . . . 25
  - 2.2.2 Circular response . . . 27
- 2.3 A nonparametric analysis of covariance model . . . 28
  - 2.3.1 Circular covariate and real-valued response . . . 28
  - 2.3.2 Circular response . . . 31
- 2.4 Simulation study . . . 33
  - 2.4.1 Results for the no-effect tests . . . 34
  - 2.4.2 Results for the ANCOVA tests . . . 37
- 2.5 Real data illustration . . . 39

3 A general framework for circular regression 47
- 3.1 Local likelihood estimation for circular regression . . . 48
- 3.2 Bias and variance of the estimator . . . 51
  - 3.2.1 Bias of the estimator . . . 51
  - 3.2.2 Variance of the estimator . . . 52
- 3.3 Selection of the smoothing parameter . . . 53
  - 3.3.1 Selection of the pilot concentration: least-squares case . . . 55
  - 3.3.2 Selection of the pilot concentration: general case . . . 56
- 3.4 Simulation study . . . 58

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3.5 Real data illustration . . . 66
3.6 Extensions to partially linear and additive models . . . 71

4 Joint estimation for over/under-dispersed regression models 73
4.1 Modeling the dispersion . . . 75
4.1.1 The Conway-Maxwell-Poisson model . . . 75
4.1.2 The double exponential family . . . 76
4.1.3 Double exponential regression model for circular covariate . . . 77
4.2 Joint estimation via circular local likelihood . . . 78
4.3 Simulation study . . . 81
4.3.1 Simulation scenarios . . . 81
4.3.2 Simulation results . . . 82
4.3.3 Joint visualization of mean and dispersion functions . . . 84
4.4 Analysis of the number of spikes in a macaque's neuron . . . 85

5 Multimodal circular regression 89
5.1 On the use of mean shift for circular modal regression . . . 91
5.1.1 Real-valued response . . . 92
5.1.2 Circular response . . . 94
5.2 Some theoretical results . . . 97
5.3 Smoothing parameter selection . . . 100
5.3.1 Modal cross-validation for circular regression . . . 102
5.4 Simulation study . . . 102
5.4.1 Simulation scenarios . . . 103
5.4.2 Simulation results . . . 103
5.4.3 The unknown boundary condition . . . 107
5.5 Analysis of the escape responses in larval zebrafish . . . 110

6 Conclusions, discussion and further research 115

Appendices 121

A Real data sets 123
A.1 Flywheels . . . 123
A.2 Sand hoppers . . . 124
A.3 Crickets . . . 125
A.4 Human motor resonance . . . 126
A.5 Spike count . . . 126
A.6 Pm10 particles . . . 127
A.7 Zebrafish . . . 127

USC DE SANTIAGO DE COMPOSTELA

B Additional simulation results **131**
B.1 Supplement to Chapter 2 . . . 131
B.2 Supplement to Chapter 5 . . . 140

C Theoretical results **143**
C.1 Proofs of Chapter 2 . . . 143
C.2 Proofs of Chapter 3 . . . 145
C.3 Assumptions and proofs of Chapter 5 . . . 153

D Software: update of the `NPCirc` package **181**
D.1 The `NPCirc` package . . . 181
D.2 Documentation for the new functions . . . 181

Resumo en galego **213**

Resumen en castellano **223**

Further information **233**

References **239**

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Abstract

The direction in which an animal escapes from a threat is an example of a random variable which is not supported on a euclidean space, but on a unit circumference. Other examples of such variables include wind or wave directions, angles of bone fractures or variables with time or calendar effects. Periodic observations, such as angles and directions, are known as circular data, and are a particular case of more complex data structures designated as directional data.

Classical statistical methods are not adequate for the correct modeling of these observations, and a suitable analysis must take into account the particular nature of the variables, which should not depend on the selected zero direction or the chosen sense of rotation.

The number of statistical works dealing with the appropriate analysis of circular data has considerably increased in recent years. New methodologies have been proposed in order to perform a number of inferential tasks. An interesting problem is the study of possible dependencies between a circular random variable and other variables of different types, which can be carried out from a regression perspective. The most common approach for modeling regression functions, usually considered as the conditional mean of the variables, is the assumption of parametric forms and the subsequent estimation of the model's parameters. However, more flexible ways of estimating a regression curve can be obtained by employing nonparametric methods, such as kernel-type estimators.

The objective of this dissertation is the proposal and study of kernel estimators for the regression function when at least one of the variables of interest presents a circular nature. The regression function is first considered as the classical conditional mean of the response variable, given the value of the covariate. A more general setting is also considered, where the regression curves are regarded as general characteristics of the conditional distribution and the estimation is carried out by local maximum likelihood. Finally, the estimation of the conditional local modes is conducted, constituting a circular multimodal regression approach. The performance of all the considered models is studied and they are applied to real data sets.

USC UNIVERSIDADE

KEYWORDS: Circular data; Kernel regression; Local likelihood; Multimodal regression.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Resumo

A dirección na que un animal escapa dunha ameaza é un exemplo de variable aleatoria con soporte na circunferencia unidade, que non é un espazo euclídeo. Outros exemplos inclúen direccións de vento ou das ondas do mar, ángulos de fracturas en ósos ou variables con efecto do tempo ou do calendario. Estas observacións periódicas, como ángulos ou direccións, son coñecidas como datos circulares, e constitúen un caso particular de estruturas de datos máis complexas denominadas datos direccionais.

Os métodos estatísticos clásicos non resultan axeitados para o correcto modelado destas observacións, e a súa análise estatística debe ter en conta a natureza periódica das variables, que non debería depender da dirección orixe seleccionada ou do sentido de rotación elixido.

O número de traballos na literatura estatística que tratan da análise apropiada de datos circulares ten aumentado de xeito considerable nos últimos anos. Para realizar distintas tarefas inferenciais, novas metodoloxías teñen sido propostas. Un problema interesante é o estudo da posible dependencia entre unha variable aleatoria circular e outras variables de diferente tipo, que pode ser tratado dende a perspectiva da regresión. O enfoque máis común para modelar funcións de regresión, normalmente consideradas como a media condicional das variables, é a suposición de formas paramétricas e a posterior estimación dos parámetros do modelo. Porén, poden obterse vías de estimación da curva de regresión máis flexibles empregando métodos non paramétricos, como os estimadores tipo núcleo.

O obxectivo desta tese de doutoramento é a proposta e estudo de estimadores tipo núcleo para a función de regresión cando, ao menos, unha das variables de interés presenta natureza circular. A función de regresión considérase, primeiramente, como a media condicional da variable resposta, dado o valor da covariable. Tamén se estuda un escenario máis xeral, onde as curvas de regresión son conceptuadas como características xerais da distribución condicional, e a estimación realízase mediante máxima verosimilitude local. Finalmente, lévase a cabo a estimación das modas locais condicionais, establecendo o tratamento da regresión multimodal circular. Para concluír, estúdase o funcionamento de todos os modelos considerados e estes son aplicados a datos reais.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PALABRAS CHAVE: Datos circulares; Regresión multimodal; Regresión tipo núcleo; Verosimilitude local.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Resumen

La dirección en la que un animal escapa de una amenaza es un ejemplo de variable aleatoria con soporte en la circunferencia unidad, que no es un espacio euclídeo. Otros ejemplos incluyen direcciones del viento o de las olas del mar, ángulos de fracturas óseas o variables con efectos del tiempo o del calendario. Estas observaciones periódicas, como ángulos o direcciones, son conocidas como datos circulares, y constituyen un caso particular de estructuras de datos máis complejas designadas como datos direccionales.

Los métodos estadísticos clásicos no resultan adecuados para el correcto modelado de estas observaciones, y su análisis estadístico debe tener en cuenta la naturaleza periódica de las variables, que no debería depender de la dirección origen seleccionada o del sentido de rotación elegido.

El número de trabajos en la literatura estadística que tratan el análisis apropiado de datos circulares ha aumentado considerablemente en los últimos años. Para realizar distintas tareas inferenciales, nuevas metodologías han sido propuestas. Un problema interesante es el estudio de la posible dependencia entre una variable aleatoria circular y otras variables de diferente tipo, que puede ser tratado desde el punto de vista de la regresión. El enfoque más común para modelar funciones de regresión, normalmente consideradas como la media condicional de las variables, es la suposición de formas paraméteicas y la posterior estimación de los parámetros del modelo. Sin embargo, pueden obtenerse vías de estimación de la curva de regresión más flexibles empleando métodos no paramétricos, como los estimadores tipo núcleo.

El objetivo de esta tesis doctoral es la propuesta y estudio de estimadores tipo núcleo para la función de regresión cuando, por lo menos, una de las variables de interés presenta naturaleza circular. La función de regresión se considera, en primer lugar, como la media condicional de la variable respuesta, dado el valor de la covariable. También se estudia un escenario más general, donde las curvas de regresión son conceptuadas como características generales de la distribución condicional, y el proceso de estimación se realiza mediante máxima verosimilitud local. Finalmente, se lleva a cabo la estimación de las modas locales condicionales, estableciendo el tratamiento de la regresión multimodal circular. Para concluir, se estudia el funcionamiento de todos los modelos considerados y éstos son aplicados a datos reales.

PALABRAS CLAVE: Datos circulares; Regresión multimodal; Regresión tipo núcleo; Verosimilitud local.

USC UNIVERSIDADE DE SANTIAGO

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Methodology

This thesis follows the classical research methodology in the field of statistics. In general, the research process begins by carrying out a comprehensive study of the topics to be addressed and by reviewing some classical and recent bibliographical references. The following stages include the proposal of new regression models and hypothesis testing tools, and their theoretical and/or computational analysis, for which the specific methodology employed in each chapter is detailed hereafter.

Chapter 2. Nonparametric inference for circular regression.

- Proposal of test statistics for the no-effect tests and calibration strategies.
- Proposal of test statistics for equality and parallelism and calibration strategies.
- Conceptualization and design of simulation experiments.
- Illustration with real data.

Chapter 3. A general framework for circular regression.

- Proposal of the estimation strategy in the circular local likelihood setting.
- Derivation of the estimator's properties.
- Proposal and study of a selection criterion for the smoothing parameter.
- Conceptualization and design of simulation experiments.
- Illustration with real data.

Chapter 4. Joint estimation for over/under-dispersed regression models.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

- Proposal of the joint estimation strategy.
- Conceptualization and design of simulation experiments.
- Application to real data.

Chapter 5. Multimodal circular regression.

- Proposal of the multimodal regression estimators.
- Derivation of the estimators' properties.
- Proposal of a selection criterion for the smoothing parameters.
- Conceptualization and design of simulation experiments.
- Application to real data.

In all the works that compose this dissertation, the design of simulation studies to empirically assess the performance of the new methods was necessary. For such aim, the use of `R` and `C++` software was essential. In addition, `R` software was employed to produce the graphical representations displayed in this manuscript, as well as to apply the derived statistical methods to real data sets.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Chapter 1
Introduction

Regression models are meant to investigate the relationship between a response variable Y and, in the simplest case, an explanatory variable X. Once this relationship is modeled, the regression analysis can be used to predict future values of the response given the predictor. The regression function is generally regarded as the expectation of the response conditioned on the value of the covariate, although considering other conditional characteristics (such as quantiles or modes) may also be of interest for explaining the underlying relationship.

Regardless of how the regression function is defined, inference with regression models is usually carried out under a parametric perspective, but more flexible models can be considered by using nonparametric regression. Over the last decades, there have been numerous contributions dealing with the nonparametric estimation of the regression function, but most works deal with the analysis of data supported on euclidean spaces. However, modern measurement tools allow to obtain data that, for a complete analysis, require embedding in other spaces beyond the euclidean context (Patrangenaru and Ellingson, 2016). This is the case of circular data, which are observations defined on the unit circumference. The periodicity of this kind of data makes classical statistical techniques (including regression methods) not suitable for a proper analysis.

The goal of this thesis is to propose new flexible methods for inference in the context of regression with circular data and showing the applicability of such techniques to real data. Section 1.1 gives a review on kernel regression methods and Section 1.2 presents an overview on circular data, including circular distribution and regression models. A detailed description of the objectives of this work and the manuscript distribution are provided in Section 1.3.

Contents

1.1 Background on kernel regression 2
1.1.1 Mean regression 3
1.1.2 Local likelihood regression 6
1.1.3 Multimodal regression 7
1.2 Background on circular data 8
1.2.1 Population and sample measures 9
1.2.2 Circular distribution models 11
1.2.3 Regression for circular data 14
1.3 Objectives and manuscript distribution 18

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1.1 Background on kernel regression

The regression function is usually defined as the expectation of the response variable conditioned on the predictor, *i.e.*,

$$m(x_0) = \mathbb{E}(Y|X = x_0). \tag{1.1}$$

The estimation of $m(\cdot)$ is important to explain the relationship between the variables and to predict future values of the response given the covariate. Classical linear regression assumes that $m(\cdot)$ is a straight line, where the parameters of the line are estimated. Other parametric shapes can be considered to obtain more flexibility when modeling the relationship between the variables, but its usage leads to several problems in practice, such as overfitting issues or the loss of interpretability of the parameters, which makes the nonparametric estimation of $m(\cdot)$ an appealing alternative.

Different methods can be employed to obtain a nonparametric estimator of the regression function, such as splines or wavelets, but kernel methods are a widely studied alternative (Wand and Jones, 1995; Fan and Gijbels, 1996). The estimation of regression functions through kernel methods allows to construct a wide range of inferential tools, including nonparametric tests to ascertain the effect of the predictor variable (no-effect tests) or tests conceived to compare regression curves across different groups (Bowman and Azzalini, 1997).

While the early statistical literature on kernel regression is focused on the estimation of the mean function (1.1), these methods have been extended to more general frameworks. Fan et al. (1995) generalized local polynomial fitting to the case of generalized linear models, a problem that was also approached by Staniswalis (1989), extending the local likelihood idea of Tibshirani and Hastie (1987) to include kernel weights in the local likelihood function. This approach allows to nonparametrically estimate conditional characteristics of a response variable Y given the covariate X by specifying the distribution of the response variable (which can be either discrete or continuous). For asymmetric conditional distributions, the estimation of the conditional median (and, more generally, of the conditional quantiles) has also been addressed by Chaudhuri (1991a) and Chaudhuri (1991b) as a nonparametric alternative to the quantile regression idea of Koenker and Bassett (1978). While these are robust methods, there has also been abundant interest on the estimation of the conditional (global) mode, as a more robust measure of location than the median (see Lee, 1989, 1993; Yao and Li, 2014). Kernel-type estimators of the modal regression function have been proposed (Sager and Thisted, 1982; Yao et al., 2012).

Although quantile and modal regression are specially useful when the conditional distribution of the data is asymmetric, different methods are needed when the conditional distribution presents multiple modes. For this setting, the nonparametric estimation of finite mixtures of regressions has been studied (Huang et al., 2013), although the joint selection of the smoothing parameter and the number of mixture components presents a difficult issue. A more flexible and completely data-driven approach is multimodal nonparametric regression (Einbeck and Tutz, 2006; Chen et al., 2016a), where the conditional local modes are esti-

mated by obtaining the local maxima of the estimated conditional density, based on kernel smoothers.

1.1.1 Mean regression

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an independent and identically distributed sample of (X, Y). A regression model to explain the dependence between Y and X can be written as

$$Y = m(X) + \sigma\varepsilon, \quad \text{where } \mathbb{E}(\varepsilon|X = x_0) = 0,\ \mathbb{E}(\varepsilon^2|X = x_0) = 1,$$

$\sigma > 0$ is the standard deviation of $Y|X = x_0$ and $m(\cdot)$ is defined as in (1.1). Under some regularity conditions for $m(\cdot)$, including that the regression function is *smooth*, Nadaraya (1964) and Watson (1964) proposed estimating $m(x_0)$ as

$$\hat{m}(x_0) = \frac{\sum_{i=1}^n L_h(X_i - x_0)Y_i}{\sum_{i=1}^n L_h(X_i - x_0)}, \tag{1.2}$$

where $L_h(\cdot) = L(\cdot/h)/h$, with $L(\cdot)$ being a kernel function (usually a symmetric around zero density) and h is known as the bandwidth or smoothing parameter. Generally, the standard normal density is used as a kernel in practice (although for theoretical derivations, compactly supported kernels are employed).

For any point x_0 from the support of X, the Nadaraya-Watson estimator is a weighted average of the responses, where the weights depend on the kernel function, assigning higher weights to the observations closer to x_0. On top of this interpretation, the Nadaraya-Watson estimator can also be thought as the result of fitting horizontal lines locally in a neighborhood of x_0, assigning weights to each observation. This approach can be extended by fitting polynomials of a predetermined degree locally (Fan and Gijbels, 1996). Assuming that the regression function $m(\cdot)$ is at least p times differentiable at x_0, for data points X_i in a neighborhood of x_0, the regression function can be expanded as

$$m(X_i) \approx m(x_0) + m'(x_0)(X_i - x_0) + \ldots + \frac{m^{(p)}(x_0)}{p!}(X_i - x_0)^p,$$

where $m^{(p)}(\cdot)$ denotes the pth derivative of $m(\cdot)$. Then, one can consider $\beta_j = m^{(j)}(x_0)/j!$, with $j = 0, \ldots, p$, as local parameters. The estimators of the local parameters are obtained by minimizing the following weighted least-squares problem:

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$\sum_{i=1}^n \left[Y_i - \sum_{j=0}^p \beta_j (X_i - x_0)^j \right]^2 L_h(X_i - x_0). \tag{1.3}$$

Then, the nonparametric estimator of the regression function at $X = x_0$ is $\hat{m}(x_0) = \hat{\beta}_0$ and the other estimated coefficients $\hat{\beta}_j$ allow for the estimation of the derivatives of order j.

When the degree of the polynomial is $p = 0$, the local polynomial estimator coincides with the Nadaraya-Watson estimator in (1.2).

The selection of the parameter h is of great importance, since small values of h lead to smaller bias but larger variance, while larger values of the smoothing parameter give rise to more bias and less variance. In addition, if $h \to 0$ the curve estimation tends to the interpolation of the data. On the contrary, as $h \to \infty$, $\hat{m}(\cdot)$ tends to a global fit. An expression for an optimal bandwidth can be derived by minimizing the Mean Integrated Squared Error (MISE) of the estimator, but it is not usable in practice, since it depends on unknown quantities. Numerous methods have been proposed in order to select an optimal smoothing parameter in practice, in the sense that bias and variance are balanced out. Plug-in methods (see Ruppert et al., 1995) try to estimate the unknown quantities of the optimal bandwidth. Other approaches such as cross-validation techniques or the AICc criterion of Hurvich et al. (1988) are also widely used in practice, the latter obtaining a good performance especially for small sample sizes.

Local polynomial estimators can be used to construct hypothesis testing tools. For example, kernel estimators may be used to test the significance of the covariate (no-effect test). In addition, further complexities arise when introducing different types of covariates, and more "complex" models are necessary. This is the case of analysis of covariance (ANCOVA) models when an additional categorical covariate is incorporated. Testing tools to determine simplifying hypotheses for nonparametric regression can also be constructed by making use of kernel estimators. In what follows, the test of no-effect and analysis of covariance (ANCOVA) models are reviewed.

Assessing the effect of the covariate

The pursued data-driven character of kernel methods makes it difficult to determine which features of the estimation correspond to the underlying regression function and which ones are just sample noise. As a graphical tool, Chaudhuri and Marron (1999) introduced the SiZer map, which allows to visualize the regions where the regression (or density) curve is significantly increasing or decreasing over a range of smoothing parameters. Therefore, the SiZer map can be a useful tool in order to determine the significance of the predictor variable. Nevertheless, it is important to formally assess the shape of the curve through hypothesis testing. For such purposes, Bowman and Azzalini (1997, Chap. 5) introduced a no-effect test in the nonparametric setting, which aims to analyze the significance of the predictor variable without predetermining a parametric form.

When trying to examine the evidence of the response and explanatory variables being related, two competing models are considered. The next hypotheses statement is used for the significance test:

$$
\begin{aligned}
&H_0 : Y_i = \mu + \varepsilon_i, \quad \forall\, i = 1, \ldots, n, \\
&H_1 : Y_i = m(X_i) + \varepsilon_i, \quad m(X_i) \neq \mu \text{ for some } i = 1, \ldots, n.
\end{aligned}
$$

The errors ε_i are assumed to be independent and $N(0, \sigma^2)$ distributed, and also independent from X, but just smoothness assumptions on $m(\cdot)$ are required. Under the null hypothesis, μ is estimated as the sample mean of the responses and, under H_1, the regression function is estimated with a kernel-type estimator. In order to construct the test statistic, Bowman and Azzalini (1997) proposed using

$$L_1 = \frac{RSS_0 - RSS}{RSS}, \tag{1.4}$$

with RSS_0 being the residual sum of squares under H_0 and RSS being the residual sum of squares under H_1:

$$RSS_0 = \sum_{i=1}^{n}(Y_i - \hat{\mu})^2, \quad \text{and} \quad RSS = \sum_{i=1}^{n}[Y_i - \hat{m}(X_i)]^2.$$

The statistic L_1 in (1.4) is a ratio between the difference of residual sums of squares and the residual sum of squares under H_1, and because of this the effect of the error variance σ^2 is scaled out. The distribution of L_1 is approximated by a shifted and scaled χ^2 with the first three cumulants matching those of the real distribution of L_1.

It is important to remark that for the implementation of the no-effect test it is necessary to select a smoothing parameter. The outcome of the test is very influenced by the bandwidth and, in practice, the test is usually carried out over a range of smoothing parameters, obtaining the trace of the test.

Analysis of covariance models

Consider now a regression model where there is also a categorical covariate with J different categories, so each observation belongs to one out of J groups. The subscript ji will be used to denote the ith observation in the jth group, where n_j is the sample size in group j. In this scenario, an analysis of covariance (ANCOVA) model can be formulated, by allowing each group to have a different regression function:

$$Y_{ji} = m_j(X_{ji}) + \varepsilon_{ji}, \quad j = 1, \ldots, J, \; i = 1, \ldots, n,$$

where the ε_{ji} are assumed to follow a $N(0, \sigma^2)$ distribution. The local polynomial estimator may be used to estimate the regression curves for each group, but hypothesis testing is needed to ascertain if the regression curves are actually different or if the differences between the estimator are an artifact of the data. The equality test is carried out with the following hypotheses statement:

$$\begin{aligned} H_0 &: Y_{ji} = m(X_{ji}) + \varepsilon_{ji}, \quad \forall\, j = 1, \ldots, J, \\ H_1 &: Y_{ji} = m_j(X_{ji}) + \varepsilon_{ji}, \quad \exists\, j, k = 1, \ldots, J \mid m_j(\cdot) \neq m_k(\cdot). \end{aligned}$$

Young and Bowman (1995) proposed the following test statistic for the equality test:

$$L_2 = \frac{1}{\hat{\sigma}^2} \sum_{j=1}^{J} \sum_{i=1}^{n_j} [\hat{m}_j(X_{ji}) - \hat{m}(X_{ji})]^2,$$

where $\hat{m}(\cdot)$ and $\hat{m}_j(\cdot)$ are the local polynomial estimators of $m(\cdot)$ and $m_j(\cdot)$ respectively and $\hat{\sigma}^2$ is an estimator of the variance.

There may be cases where the regression curves for each of the groups are not the same but differ of each other just by a constant shift. This assumption may be analyzed through the test of parallelism. In this case, the considered hypotheses are

$$H_0 : Y_{ji} = \gamma_j + m(X_{ji}) + \varepsilon_{ji}, \quad \gamma_1 = 0,\ \gamma_j \in \mathbb{R},\ \forall\, j = 1, \ldots, J,$$
$$H_1 : Y_{ji} = m_j(X_{ji}) + \varepsilon_{ji}, \quad m_j(\cdot) \neq m_k(\cdot) + \gamma \text{ for some } j, k = 1, \ldots, J \text{ and } \forall\, \gamma \in \mathbb{R}.$$

In this case, the model under the alternative hypothesis is estimated as in the test of equality, but under the null hypothesis a semiparametric method is used to estimate both the regression function and the shift parameters. The test statistic used by Young and Bowman (1995) is given by

$$L_3 = \frac{1}{\hat{\sigma}^2} \sum_{j=1}^{J} \sum_{i=1}^{n_j} [\hat{\gamma}_j + \hat{m}(X_{ji}) - \hat{m}_j(X_{ji})]^2,$$

where $\hat{\sigma}^2$ is an estimator of the variance, $\hat{m}(\cdot)$ is an estimator of the regression function and $\hat{\gamma}_j$ is the estimator of the shift parameter for the jth group.

As in the no-effect test, the cumulants of the distribution of L_2 and L_3 are obtained and the distribution of the statistics L_2 and L_3 is approximated to a shifted and scaled χ^2 with the same cumulants.

1.1.2 Local likelihood regression

The kernel estimators reviewed in the previous section allow to estimate the conditional mean of Y given X. This approach is useful when Y is a continuous variable and the expectation of the error is zero. In addition, as it happens for the no-effect, equality and parallelism tests, it is usually assumed that the errors follow a normal distribution.

A broader class of kernel estimators which accounts for more general responses was derived by Staniswalis (1989) and Fan et al. (1995), by extending the idea of local likelihood proposed by Tibshirani and Hastie (1987). In what follows, the local polynomial fitting approach of Fan et al. (1995) will be reviewed. In this setting, the response variable Y can be either discrete or continuous and the target function to estimate is not necessarily the conditional mean, but any conditional parameter or a transformation of it. Note that the conditional distribution of $Y|X$ needs to be specified. Assume that the density function of $Y|X = x_0$ is given by $f[y; \lambda(x_0)]$, where $\lambda(x_0)$ is a quantity of interest derived from the conditional density. Now let $g(x_0) = \eta[\lambda(x_0)]$, with $\eta(\cdot)$ being a link function and

$\ell[g(x_0), y] = \log f[y; \lambda(x_0)]$. Fan et al. (1995) proposed estimating $g(x_0)$ by maximizing the local log-likelihood, given by

$$\mathcal{L}_p(\boldsymbol{\beta}; h, x_0) = \sum_{i=1}^{n} \ell \left[\sum_{j=0}^{p} \beta_j (X_i - x_0)^j, Y_i \right] L_h(X_i - x_0),$$

where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^T$ (with ^{T} denoting the transpose operator) is a vector of local parameters and p is the degree of the local polynomial used. The estimator of $g(x_0)$ is then given by $\hat{g}(x_0) = \hat{\beta}_0$. Note that this approach depends on the selection of a smoothing parameter h, which plays a crucial role in the outcome of the estimation process. Fan et al. (1998) also studied the use of the local likelihood methodology for inference purposes, proposing a method for estimating the bias and variance of the estimated parameters and a unified approach for selecting the bandwidth in practice.

This methodology allows to estimate a parameter of the conditional distribution, which in the particular case of belonging to the exponential family it is usually a transformation of the conditional mean. Note that in the particular case of the normal distribution, the maximization problem is equivalent to the minimization problem in (1.3).

1.1.3 Multimodal regression

When dealing with highly asymmetric data sets, modal regression presents a useful alternative, where the regression function of Y conditioned on the value of X is regarded as the mode of the conditional density (Sager and Thisted, 1982; Yao et al., 2012). However, such proposals assume that the conditional density of $Y|X$ is unimodal. When the density of $Y|X$ presents multiple modes, a recent fully nonparametric approach to study the relationship between the variables is multimodal regression (see Chen, 2018; Alonso-Pena, 2020, for a review). The original idea of estimating a multimodal regression model was first introduced by Scott (1992, Chap. 8). Afterwards, Einbeck and Tutz (2006) proposed using the conditional mean shift algorithm to estimate the local conditional modes.

Assume that the conditional density function of the response Y given $X = x$, namely $f(y|x)$, may present more than one mode (although the number of local modes is unspecified), and that all the modes belong to the interior of the support of X. Consider now a bivariate sample of (X, Y), namely $\{(X_i, Y_i)\}_{i=1}^n$. If the conditional density is multimodal, the data cannot be described by a regression function, but it can be described by a multifunction, *i.e.*, for each value x in the support of X, there might be several values of the response variable that maximize the conditional density. This can be interpreted as a scenario where the data $\{(X_i, Y_i)\}_{i=1}^n$ belong to different groups, like in an ANCOVA model, but without these groups being specified in advanced.

The goal of multimodal regression is estimating the regression multifunction, defined as the local maxima of the conditional density function. More formally, the regression multi-

function for x in the support of X is defined as the set

$$M(x) = \left\{ y : \frac{\partial}{\partial y} f(y|x) = 0, \frac{\partial^2}{\partial y^2} f(y|x) < 0 \right\}. \tag{1.5}$$

Recall that the conditional density function can be written as the ratio between the joint density, which with an abuse of notation will be denoted $f(x, y)$, and the marginal density of the explanatory variable, $f(x)$, as long as it is non null. A simple approach to estimate the regression multifunction (1.5) is to use an indirect nonparametric method, without assuming any parametric form. First, the conditional density is estimated via a kernel method (see Hyndman et al., 1996) as

$$\hat{f}(y|x) = \frac{\frac{1}{n}\sum_{j=1}^{n} L_{h_1}(X_i - x) L_{h_2}(Y_i - y)}{\frac{1}{n}\sum_{j=1}^{n} L_{h_1}(X_i - x)}$$

where h_1 and h_2 are smoothing parameters. Then, $M(x)$ in (1.5) is estimated as the collection of local modes of the estimated density, *i.e.*,

$$\widehat{M}(x) = \left\{ y : \frac{\partial}{\partial y} \hat{f}(y|x) = 0, \frac{\partial^2}{\partial y^2} \hat{f}(y|x) < 0 \right\}.$$

However, it is not possible to obtain a closed form expression of $\widehat{M}(x)$, so numerical methods are needed for the computation of the estimated regression multifunction. Einbeck and Tutz (2006) proposed using a conditional version of the mean shift algorithm (Fukunaga and Hostetler, 1975; Cheng, 1995; Comaniciu and Meer, 2002) to estimate the regression multifunction based on a kernel-type conditional density estimator. The theoretical properties of this approach were studied by Chen et al. (2016a) and extended to measurement error problems by Zhou and Huang (2016). Although, initially, Einbeck and Tutz (2006) selected the smoothing parameters with techniques tailored for conditional density estimation, recent proposals for selecting h_1 and h_2 specifically in the modal regression context were brought up by Chen et al. (2016a) and Zhou and Huang (2019).

1.2 Background on circular data

Circular data refers to observations that can be represented as points on the circumference of the unit circle. Once a zero direction and a sense of rotation have been chosen, these observations can be expressed as angles. Most examples of this type of data come from compass or clock measurements, such as directions or observations with time or calendar effects. Circular data can be found in many different fields: biology (orientation of the red wood ants in reaction to different stimuli; Jander (1957)), geology (cross-beds; SenGupta and Rao (1966) and Mardia (1972)), environmetrics and oceanography (wind and waves directions; Jona-Lasinio et al. (2012) and Oliveira et al. (2014a)), medicine (sudden infant death syndrome; Mooney

et al. (2003)) or ecology (wildfire occurrences in the Iberian Peninsula; Ameijeiras-Alonso et al. (2019)).

Due to the periodicity of the data, classical statistical techniques might not be suitable to handle circular observations, so different methods must be applied to this kind of data. Jammalamadaka and SenGupta (2001), Pewsey et al. (2013) and Ley and Verdebout (2017) present a detailed review of circular statistics. Circular data can also be viewed as a particular case of directional data, defined in a hypersphere of arbitrary dimension (see Mardia and Jupp, 2000). Hence, methods for spherical data can be adapted to the circle.

This section aims to give some background on circular data: Section 1.2.1 gives an introduction to the main population and sample measures for circular data. Section 1.2.2 reviews the most important parametric circular distributions and parametric regression models for this kind of data will be presented in Section 1.2.3.

1.2.1 Population and sample measures

The present section will be devoted to the introduction of different measures of a circular population, such as location, concentration and dispersion measures, as well as their sample analogues. For that aim it is necessary to define, first, the concept of a circular density function.

When considering a circular random variable Θ with support in $\mathbb{T} = [0, 2\pi)$, a function $f(\cdot)$ will be a circular density if it verifies:

Condition 1. $f(\theta) \geqslant 0$ for $\theta \in \mathbb{T}$,

Condition 2. $f(\theta + 2\pi k) = f(\theta)$ for $\theta \in \mathbb{T}$ and $k \in \mathbb{Z}$,

Condition 3. $\int_0^{2\pi} f(\theta)d\theta = 1$.

Therefore, the main difference with a linear density function is the addition of Condition 2.

Population measures

Consider the circular variable Θ with density function $f(\cdot)$. Its first cosine and sine moments are defined, respectively, as

$$C = \mathbb{E}(\cos\Theta) = \int_0^{2\pi} \cos(\theta) f(\theta) d\theta \quad \text{and} \quad S = \mathbb{E}(\sin\Theta) = \int_0^{2\pi} \sin(\theta) f(\theta) d\theta.$$

The basic circular measure of location is the population mean direction, defined as

$$\mu = \text{atan2}(S, C), \tag{1.6}$$

where atan2(S, C) is an operator returning the angle between the x-axis and the vector from the origin to (C, S). It is defined as

$$\text{atan2}(a,b) = \begin{cases} \arctan(b/a) & \text{if} \quad a > 0, \\ \arctan(b/a) + \pi & \text{if} \quad b \geqslant 0,\ a < 0, \\ \arctan(b/a) - \pi & \text{if} \quad b < 0,\ a < 0, \\ \pi/2 & \text{if} \quad b > 0,\ a = 0, \\ -\pi/2 & \text{if} \quad b < 0,\ a = 0, \\ \text{undefined} & \text{if} \quad b = 0,\ a = 0. \end{cases} \tag{1.7}$$

Note that the mean direction does not always exist, since it is undefined when both S and C are zero, which means null concentration. An alternative measure of location is the mode, which, as in the real line, is defined as the maximizer of its density function. For a multimodal density, it can be of interest to consider all the local modes, defined as the set $\{\theta \in \mathbb{T} : f'(\theta) = 0, f''(\theta) < 0\}$.

In the circle, given the nature of the variable's support, namely $\mathbb{T} = [0, 2\pi)$, measures of concentration are more frequently used than measures of dispersion. The fundamental measure of concentration is the mean resultant length, which is defined as

$$\rho = \left(S^2 + C^2\right)^{1/2} \in [0, 1].$$

The mean resultant length always exists and the case when $\rho = 0$ corresponds to the situation where μ is undefined.

Although measures of concentration are far more used in the circular context, it is also possible to define measures of dispersion in the circle. For instance, the population circular variance is defined in Mardia and Jupp (2000, Ch. 3) as

$$V = 1 - \rho \in [0, 1]. \tag{1.8}$$

Sample measures

Given a set of circular observations $\{\theta_1, ..., \theta_n\}$, it is possible to apply some statistics to summarize the data. For unimodal data, it is of interest to calculate the sample mean direction, which can be obtained as

$$\bar{\theta} = \text{atan2}(\bar{S}, \bar{C}),$$

where

$$\bar{S} = \frac{1}{n}\sum_{i=1}^{n} \sin(\theta_i), \quad \bar{C} = \frac{1}{n}\sum_{i=1}^{n} \cos(\theta_i).$$

Therefore, $\bar{C}$ and $\bar{S}$ are the sample analogues of C and S, respectively. Note that the sample mean direction is not defined when $\bar{S}$ and $\bar{C}$ are zero. As for the global mode direction, analogously to the global mode of a real-valued random variable, it can be estimated through an

indirect procedure, by first estimating (nonparametrically) the density function of the random variable and afterwards obtaining the global maximum. The same method can be used to estimate the set of local modes. The nonparametric estimation of the circular density function can be carried out with the circular kernel density estimator (see Section 1.2.2).

It is also possible to define the sample mean resultant length in terms of $\bar{S}$ and $\bar{C}$, as

$$\bar{R} = \left(\bar{S}^2 + \bar{C}^2\right)^{1/2}.$$

The sample mean resultant length $\bar{R}$ equals 1 only when all the data points are located at the same point on the unit circle. Then, a value close to 1 means that the data are closely congregated around the mean direction. On the other hand, a value close to 0 does not necessarily indicate that the data are evenly distributed around the circle. For example, the set of observations $\{\theta_1, ..., \theta_{n/2}, \theta_1 + \pi, ..., \theta_{n/2} + \pi\}$, for even n, has a mean resultant length of $\bar{R} = 0$.

Furthermore, sample measures of dispersion on the circle are easily obtained. The sample circular variance is defined as

$$\bar{V} = 1 - \bar{R},$$

which is also defined in $[0, 1]$. The distance between two points in the circle ω and ϕ is usually measured by the circular distance, also known as cosine distance,

$$d(\omega, \phi) = 1 - \cos(\omega - \phi). \tag{1.9}$$

Thus, a measure of dispersion of the sample $\{\theta_1, ..., \theta_n\}$ about a direction Ψ is given by

$$d_c(\Psi) = \frac{1}{n}\sum_{i=1}^{n}[1 - \cos(\theta_i - \Psi)],$$

which is minimized when $\Psi = \bar{\theta}$. Hence, it holds that $d_c(\bar{\theta}) = 1 - \bar{R} = \bar{V}$.

1.2.2 Circular distribution models

Many parametric models for circular distributions have been proposed in literature. In this section, the most widely used circular models will be briefly described.

One of the most important and useful circular distributions is the von Mises, introduced by Von Mises (1918). The von Mises distribution $\text{vM}(\mu, \kappa)$, is a symmetric and unimodal distribution determined by a mean direction $\mu \in [0, 2\pi)$ and a concentration parameter $\kappa \geqslant 0$, with density function

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \mu)], \quad \theta \in \mathbb{T},$$

where $I_j(\cdot)$ denotes the modified Bessel function of the first kind and order j, which in this

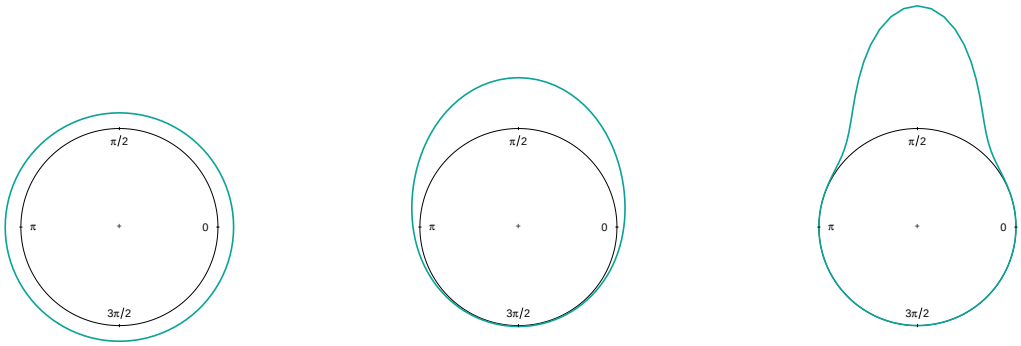


Figure 1.1: Polar representations of von Mises densities with mean direction $\mu = \pi/2$ and concentration parameters $\kappa = 0$ (left), $\kappa = 2$ (center) and $\kappa = 10$ (right).

case it is used as a normalizing constant. The von Mises distribution is the unique continuous circular distribution for which the sample mean direction, $\bar{\theta}$, is the maximum likelihood estimate of the population mean direction, μ. For this reason, the von Mises distribution is sometimes known as the circular normal, as noted by Gumbel et al. (1953), who studied its properties and remarked its similarities with the normal distribution. The spread of the Gaussian distribution is controlled by the standard deviation σ, while in the case of the von Mises this is controlled by the concentration parameter κ. The larger κ is, the larger the concentration, hence the lower the spread of the distribution. When $\kappa \to 0$, the von Mises density function tends to

$$f(\theta) = \frac{1}{2\pi}, \quad \theta \in \mathbb{T},$$

which corresponds to the density function of the circular uniform distribution. First row in Figure 1.1 shows representations of the von Mises density functions with mean direction $\mu = \pi/2$ and different concentration parameters. The density in the left panel corresponds to the circular uniform distribution.

A simple way of obtaining circular densities is to "wrap" a linear density around the unit circle. If X is a random variable in $\mathbb{R}$ with density function $g(\cdot)$, then $\Theta = X \pmod{2\pi}$ and has density function

$$f(\theta) = \sum_{k=-\infty}^{\infty} g(\theta + 2\pi k), \quad \theta \in \mathbb{T}.$$

USC

The most widely used examples of wrapped densities are the wrapped normal and the wrapped Cauchy, for which the infinite sum can be simplified to a more compact expression. Note that the von Mises distribution, the uniform distribution, the wrapped Cauchy and many other important distributions are just a particular case of a three parameter family of symmetric circular distributions known as the Jones-Pewsey family (Jones and Pewsey, 2005).

Circular distributions can also be obtained by radial projection of bivariate distributions on the plane. Let the random variable $\boldsymbol{X} = (X_1, X_2)^T$ follow a bivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and covariance matrix $\boldsymbol{\Sigma}$. The vector $\boldsymbol{U} = \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}$, with $\boldsymbol{U} = (U_1, U_2)^T$, is the projection of $\boldsymbol{X}$ onto the unit circle. Then, the variable Θ is defined by taking $U_1 = \cos\Theta$ and $U_2 = \sin\Theta$, and it is said to follow a projected normal distribution $PN_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (Small, 1996; Mardia and Jupp, 2000). Since the general structure of the covariance matrix is

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & a\sigma_1\sigma_2 \\ a\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

the density function of Θ depends on five parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, a)$, and its explicit form is rather complicated. When $\boldsymbol{\Sigma} = \boldsymbol{I}_2$, with $\boldsymbol{I}_2$ being the identity matrix of order 2, the density function of Θ is unimodal and symmetric. In addition, if $\mu_1 = \mu_2 = 0$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_2$, the distribution is uniform on the circle. This model provides the basis for a widely used regression model for a circular response and real-valued covariates, (Presnell et al., 1998; Scapini et al., 2002).

All the distributions presented so far are symmetric (except for some particular cases of the projected normal distribution with $\boldsymbol{\Sigma} \neq \boldsymbol{I}_2$), but there are also asymmetric circular models, such as the Kato-Jones density (Kato and Jones, 2015). Pewsey et al. (2013, Ch. 4) and Ley and Verdebout (2017, Ch. 2) can be consulted for other distributions. In addition, some multimodal models are considered in literature, such as the generalized von Mises distribution (Gatto and Jammalamadaka, 2007). Furthermore, when considering the projected normal distribution with $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2$, the resulting density is symmetric and with antipodal modes, *i.e.*, it has two equal modes in opposite directions. However, with the objective of having a better parameter interpretation, the general approach for modeling multimodal data is to use finite mixtures of unimodal densities. The mixture density function is constructed as

$$f(\theta) = \sum_{k=1}^{K} p_k f_k(\theta), \quad \theta \in \mathbb{T}, \quad p_k \geqslant 0,$$

with $\sum_k p_k = 1$ and $f_k(\cdot)$ being circular densities. Finite mixtures of von Mises distributions are widely used and have been studied by many authors (Mardia and Sutton, 1978; Spurr, 2015; Bartels, 1984).

Although mixture models are flexible ways of modeling probability density functions, when the underlying density is a very complex function or there are not evidences to assume a specific parametric form for the density, nonparametric methods such as the circular kernel density estimator (see, for example, Oliveira et al., 2013) are usually employed to estimate the true density function. Let $\Theta_1, \ldots, \Theta_n$ be an independent and identically distributed sample drawn from the density $f(\cdot)$. The circular kernel density estimator of $f(\cdot)$ at the point θ_0

is given by

$$\hat{f}(\theta_0) = \frac{1}{n}\sum_{i=1}^{n} K_\kappa(\Theta_i - \theta_0), \tag{1.10}$$

where $K_\kappa(\cdot)$ is a circular kernel with concentration parameter κ. It is assumed that κ is a sequence depending on n, such that κ grows as n tends to infinity. The most widely used circular kernel in the statistical literature is the von Mises density, although the wrapped normal also has some interesting properties. However, as in the real-valued context, the selection of the kernel is not a very relevant issue in practice, while the concentration parameter considerably affects the performance of the estimator. Large values of κ lead to an undersmoothed estimator of the density function, while small values of κ yield oversmoothed estimators.

1.2.3 Regression for circular data

Circular observations might be accompanied by other measurements, which may be either circular or real-valued. It may be of interest to study the dependence between both variables from a regression perspective. Depending on the nature of each variable, three different regression scenarios are possible: circular covariate and real-valued response (circular-linear regression), real-valued covariate and circular response (linear-circular regression) and circular covariate and circular response (circular-circular regression). In this section, the different circular regression contexts are introduced and the representation of circular regression functions is shown. In addition, the main parametric regression models for circular data are presented by means of considering parametric alternatives to the nonparametric methods developed in this thesis. It is worth noting that most of the models described in this section have been used in the literature to analyze some of the data sets which will be employed in this dissertation for illustration or application purposes.

Circular-linear regression

In the case where the explanatory variable Θ has a circular nature and the response variable Y is real-valued, the following parametric model[1] is usually considered:

$$Y = \beta_0 + \beta_1 \cos\Theta + \beta_2 \sin\Theta + \varepsilon, \tag{1.11}$$

where β_0, β_1 and β_2 are parameters that must be estimated and the error term, ε, is independent from Θ and follows a normal distribution with mean 0 and constant variance σ^2. This can be considered as a multiple linear model and the parameters can be estimated via least-squares. Figure 1.2 presents simulated data from this model along with the true regression curve. The left panel shows a planar representation, where the periodicity can be seen by joining the vertical extremes of the plot. The center panel displays a polar representation,

[1] Although only one predictor variable is considered for simplicity, the model can be extended to include more covariates of a circular nature via their sine and cosine components.

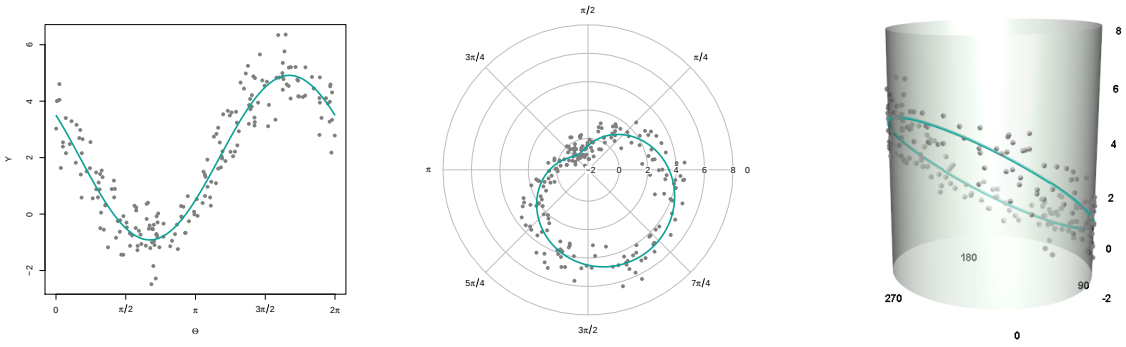


Figure 1.2: Scatter plots of $n = 200$ sample points generated from model (1.11) with parameters $\beta_0 = 2$, $\beta_1 = 1.5$ and $\beta_2 = -2.5$, with the true regression function. Left: planar representation. Center: radial representation. Right: cylindrical representation.

where the magnitude of the response variable is given by the distance from the center of the plot. Lastly, the right panel presents a 3-dimensional representation on the cylinder.

It is important to note that model (1.11) is equivalent to the following cosine model

$$Y = \gamma_0 + \gamma_1 \cos(\Theta - \gamma_2) + \varepsilon, \tag{1.12}$$

taking $\gamma_0 = \beta_0$, $\gamma_1 = \sqrt{\beta_1^2 + \beta_2^2}$ and $\gamma_2 = \text{atan2}(\beta_2, \beta_1)$. Model (1.11) can be extended to include more sine and cosine terms. Its general form for a certain $K \in \mathbb{N}$ is given by

$$Y = \sum_{k=0}^{K} [\beta_k \cos(k\Theta) + \gamma_k \sin(k\Theta)] + \varepsilon. \tag{1.13}$$

Again, the estimation procedure for these models can be carried out by least-squares in a multiple linear regression setting.

Other models can be contemplated in this scenario, for example, Batschelet (1981) proposed a skew-cosine model to account for regression functions exhibiting skewness by including a fourth parameter and another cosine term inside the trigonometric functions in model (1.11). The same author discussed an alternative model capable of describing some departures from sinusoidal oscillations.

Linear-circular regression

Consider now the case where the response variable Φ is circular, which is quite different from the previous setting. If the predictor variable is real-valued, the regression function can be regarded as a curve on the surface of an infinite cylinder, as it can be seen in the right panel of Figure 1.3, although a planar representation of the data is given in the left panel. Note that, in this representation, the top and bottom parts of the plot should be joined. The image displays simulated data from a linear-circular model jointly with the true regression function. The

dependence of Φ on the explanatory variable X can be modeled by ensuring that its mean direction depends on X. Classical linear-circular regression models are based on the von

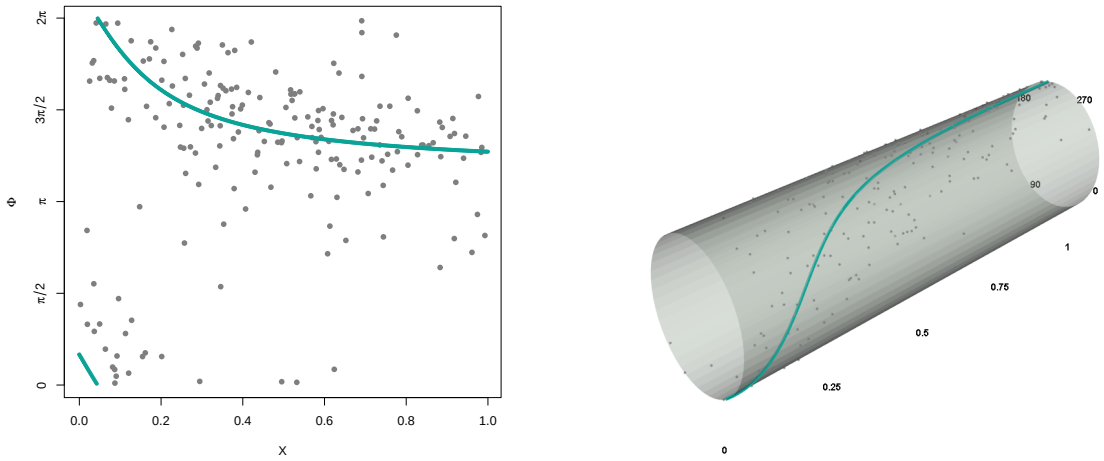


Figure 1.3: Scatter plots of $n = 200$ sample points generated from model (1.14) with the true regression function. Left: planar representation. Right: cylindrical representation.

Mises distribution. Let $\Phi_1, ..., \Phi_n$ be angular observations such that each observation follows a von Mises distribution where the mean direction depends on the predictor, *i.e.,*

$$\Phi_i \sim \text{vM}[\mu(X_i), \kappa], \quad i = 1, \ldots, n,$$

where κ is a constant value. One of the earliest linear-circular regression models was proposed by Gould (1969). This model assumes that the mean direction depends linearly on X:

$$\mu(X_i) = \beta_0 + \beta_1 X_i, \quad i = 1, \ldots, n, \ \beta_0 \in \mathbb{T}, \ \beta_1 \in \mathbb{R},$$

where the parameters β_0 and β_1 are estimated through maximum likelihood. This model is considered to be implausible in many situations and, in addition, it has been pointed out that some problems arise when fitting the model, since the maximum likelihood estimators for the parameters are not unique. Johnson and Wehrly (1978) proposed a regression model which was generalized by Fisher and Lee (1992). The location model of Fisher and Lee assumes that the μ's are related to the covariate by means of a link function $\eta(\cdot)$ so that

$$\mu(X_i) = \beta_0 + \eta(\beta_1 X_i), \quad i = 1, \ldots, n, \ \beta_0 \in \mathbb{T}, \ \beta_1 \in \mathbb{R}, \tag{1.14}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

where β_0 and β_1 are parameters that must be estimated. The link function must be chosen in order to map the real line onto the circle, so $\eta(\cdot)$ should be a monotone function with support $(-\infty, \infty)$ and range $\mathbb{T} = [0, 2\pi)$ or $[-\pi, \pi)$. Since the parameter β_0 is considered as the origin, the link function should verify $\eta(0) = 0$. One possibility is to choose the link function

from a parametric family of links, such as

$$\eta(u) = 2\tan^{-1}[\text{sgn}(u)|u|^{\lambda}],$$

where $\text{sgn}(\cdot)$ is the sign function. Note that for this family of link functions it is necessary that both γ and the observations Φ_i belong to $[-\pi, \pi)$. Although λ could be estimated from the data as in the estimation of Box-Cox transformations, usually $\lambda = 1$ is taken obtaining

$$\eta(u) = 2\tan^{-1}(u).$$

The estimation of the parameters is done through maximum likelihood using an iteratively reweighted least-squares algorithm, although Presnell et al. (1998) pointed out some computational problems that arise from this procedure. These authors presented a more flexible model which does not rely on the von Mises assumption, the Projected Multivariate Linear Model (PMLM). This is a parametric model which assumes that the directions conditioned by the values of the predictor variables follow a projected normal distribution, *i.e.*, the directions are projections onto the unit circle of a bivariate normal distribution (see Section 1.2.2).

Regression models where the parameter dependent on the predictor variable is the concentration are also plausible. Fisher and Lee (1992) generalized a concentration model originally proposed by Johnson and Wehrly (1978), in which the concentration parameters are modeled by

$$\kappa_j = h(X_i), \quad i = 1, \ldots, n,$$

where the function $h(\cdot)$ maps $\mathbb{R}$ onto $[0, \infty)$. The most common option is selecting

$$h(u) = a\exp(\omega u), \quad a \in [0, \infty),\ \omega \in \mathbb{R}. \tag{1.15}$$

Again, the parameters a and ω must be estimated from the data. Fisher and Lee (1992) also added a hybrid model, in which both the location and the concentration parameters depend on the predictors. Such model is a combination of models (1.14) and (1.15), and can be obtained straightforwardly.

Circular-circular regression

In the case where both variables in the regression model are circular, the interpretation of the regression function could be a curve on the surface of a torus, as shown in the right panel of Figure 1.4, where simulated data from a circular-circular model is displayed. Consider the explanatory variable Θ and the response variable Φ. Jammalamadaka and SenGupta (2001) used the fact that the functions $\sin(\cdot)$ and $\cos(\cdot)$ are periodic with period 2π, and they can be expressed in terms of their Fourier series expansions. Thus, the authors proposed fitting the

following general model:

$$\begin{aligned} \cos(\Phi) &= \beta_0^c + \sum_{k=1}^{K} [\beta_k^c \cos(k\Theta) + \omega_k^c \sin(k\Theta)] + \varepsilon^c, \\ \sin(\Phi) &= \beta_0^s + \sum_{k=1}^{K} [\beta_k^s \cos(k\Theta) + \omega_k^s \sin(k\Theta)] + \varepsilon^s, \end{aligned} \tag{1.16}$$

where ε^c and ε^s are, respectively, the errors associated to the cosine and sine components. This model does not assume any specific distribution for the errors. The estimation of the parameters is done by least-squares. When selecting the *degree* K of the polynomial, it is recommended to start with a small value of K and then compute the reduction in the error sum of squares when using a *degree* of $(K + 1)$. If the reduction is large enough, then $(K + 1)$ is considered and the process is repeated until there is not a significantly reduction (see Jammalamadaka and SenGupta, 2001, for details).

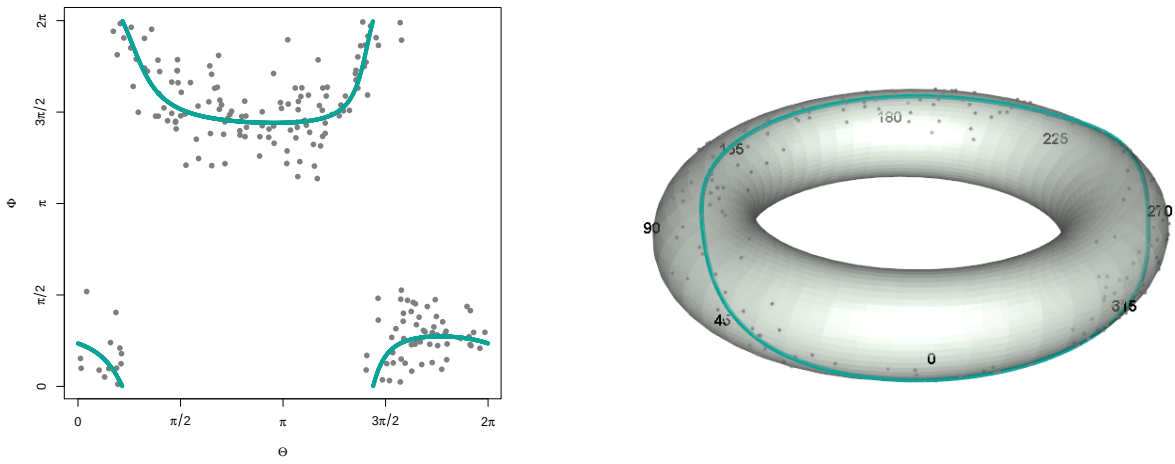


Figure 1.4: Scatter plots of $n = 200$ sample points generated from model (1.16) with the true regression function. Left: Planar representation. Right: Toroidal representation.

1.3 Objectives and manuscript distribution

In Section 1.1, different kernel methods to study the relationship between two or more variables have been presented. However, these techniques are tailored for real-valued variables, and most kernel regression methods have not been studied when, at least, one of the variables is embedded in the unit circumference. Therefore, the main goal of this thesis is to derive procedures which allow to carry out analogous analysis to those presented in Section 1.1, when at least one of the variables presents a circular nature. The objectives and contents of

the manuscript are briefly described in this section.

Chapter 2 reviews kernel regression methods for a circular predictor and a real-valued response or for a circular response, where the regression function is regarded as the conditional expectation of the response. In addition, Chapter 2 introduces nonparametric significance tests and ANCOVA models in those contexts. The new methodology is applied to three different real data sets. The contents of this chapter are collected in Alonso-Pena et al. (2021).

In Chapter 3, a generalization of local likelihood regression, where the explanatory variable presents a circular nature, is presented. Approximations of the bias and variance of the estimators are provided, and a method for selecting the smoothing parameter is proposed. Four different data sets are employed to illustrate the generality of the estimators. This chapter is mainly based on Alonso-Pena et al. (2022b).

Chapter 4 is motivated by a real life problem regarding the modeling of spike counts in macaque's neurons. Employing the previously introduced circular local likelihood estimator, Chapter 4 intends to nonparametrically estimate, jointly, the mean and dispersion functions in double exponential models with a circular predictor. The main contents of this chapter are contained in Alonso-Pena et al. (2022a).

In Chapter 5, the estimation of the conditional modes in the regression context is considered, both when the response variable is real-valued and the predictor is circular and when the response presents a circular nature. An interesting application concerning animal escape data shows the practical applicability of the proposed methods. This chapter is based on Alonso-Pena and Crujeiras (2022).

Finally, the manuscript includes some concluding remarks and discussion in Chapter 6 and four appendices. Appendix A contains the description of the real data sets employed in this thesis. Appendix B provides supplementary simulation results complementing those in the main body of the manuscript and Appendix C gives the proofs of the theoretical results included in the thesis. The `R` package `NPCirc` was updated with the functions implemented for this thesis, and the descriptions of the programmed functions are provided in Appendix D.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 2
Nonparametric inference for circular regression

Parametric regression models for circular variables were presented in Section 1.2.3. These models have been employed in a variety of practical situations, but their applicability is limited due to the restrictive parametric forms. Although some of the models, such as the one in (1.16), can lead to more flexible shapes with the addition of more terms, this method yields an increase in the number of parameters. Hence, this route can imply a loss of interpretability of the parameters, in addition to possible misspecification issues.

An alternative to the previously described models is nonparametric circular regression, introduced by Di Marzio et al. (2009) in the case of real-valued responses and by Di Marzio et al. (2013) in the setting of a circular response. These models allow a flexible estimation of the conditional mean function when the covariate and/or the response are circular variables. However, the pursued data-driven character of the estimators implies that it is not easy to determine when the features of the estimated regression function are actually attributable to the shape of the underlying regression function or are due to the sample noise.

The aim of this chapter is to introduce nonparametric tests to assess simplifying hypothesis in the circular regression context. In Section 2.1, the kernel regression estimators are presented. Section 2.2 is devoted to the introduction of no-effect tests in the context of circular data, while Section 2.3 considers nonparametric analysis of covariance models and introduces tests to determine the equality and parallelism of the regression curves. An extensive simulation study is carried out in Section 2.4, and all the proposals are illustrated with real data in Section 2.5. The contents of this chapter are collected in Alonso-Pena et al. (2021).

Contents

2.1 Nonparametric estimation of the mean regression function 22
2.1.1 Circular covariate and real-valued response 22
2.1.2 Circular response 24
2.2 A test of no-effect 25
2.2.1 Circular covariate and real-valued response 25
2.2.2 Circular response 27
2.3 A nonparametric analysis of covariance model 28
2.3.1 Circular covariate and real-valued response 28

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2.3.2 Circular response 31
2.4 Simulation study 33
2.4.1 Results for the no-effect tests 34
2.4.2 Results for the ANCOVA tests 37
2.5 Real data illustration 39

2.1 Nonparametric estimation of the mean regression function

2.1.1 Circular covariate and real-valued response

Consider a circular predictor variable, Θ, with support on $\mathbb{T} = [0, 2\pi)$, and a real-valued response variable, which will be denoted by Y, with support on $\mathbb{R}$. The relationship between both variables may be described by means of the conditional expectation of Y given the value of Θ, *i.e.*, a *mean* regression model can be formulated as

$$Y = m(\Theta) + \sigma(\Theta)\varepsilon, \quad \text{where } \mathbb{E}(\varepsilon|\Theta = \theta_0) = 0,\ \mathbb{E}(\varepsilon^2|\Theta = \theta_0) = 1, \tag{2.1}$$

which implies $\text{Var}(\varepsilon|\Theta = \theta_0) = 1$. Given a bivariate sample of both variables, $\{(\Theta_i, Y_i)\}_{i=1}^n$, Di Marzio et al. (2009) considered the estimation of the regression function $m(\cdot)$ at a fixed point $\theta_0 \in \mathbb{T}$ by using a Taylor-like expansion. Assuming that $m(\cdot)$ is at least p times differentiable at θ_0, for data points Θ_i in a neighborhood of θ_0, it holds that

$$m(\Theta_i) \approx m(\theta_0) + m'(\theta_0)\sin(\Theta_i - \theta_0) + ... + \frac{m^{(p)}(\theta_0)}{p!}\sin^p(\Theta_i - \theta_0), \tag{2.2}$$

where $m^{(p)}(\cdot)$ denotes the pth derivative of $m(\cdot)$. By considering $\beta_\nu = m^{(\nu)}(\theta_0)/\nu!$ for $\nu = 0, \ldots, p$, the regression function and its derivatives can be estimated via weighted local least-squares, where the weights are given by a circular kernel $K_\kappa(\cdot)$ with concentration parameter κ. Thus, by minimizing the expression

$$\sum_{i=1}^{n}\left[Y_i - \sum_{j=0}^{p}\beta_j \sin^j(\Theta_i - \theta_0)\right]^2 K_\kappa(\Theta_i - \theta_0) \tag{2.3}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

with respect to the vector of local parameters $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^T$, the estimators of the regression function and its derivatives are given by

$$\hat{m}^{(\nu)}(\theta_0) = \nu!\hat{\beta}_\nu, \quad \nu = 0, \ldots, p, \tag{2.4}$$

where ν represents the order of the derivative. Note that Di Marzio et al. (2009) considered only the *local linear* estimator obtained when $p = 1$.

Let $\boldsymbol{Y}$ be the vector of responses $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, $\boldsymbol{W}$ be the diagonal weight matrix $\boldsymbol{W} = \{\text{diag}[K_\kappa(\Theta_i - \theta_0)]\}_{i=1,\ldots,n}$ and

$$\boldsymbol{\Theta} = \begin{pmatrix} 1 & \sin(\Theta_1 - \theta_0) & \ldots & \sin^p(\Theta_1 - \theta_0) \\ 1 & \sin(\Theta_2 - \theta_0) & \ldots & \sin^p(\Theta_2 - \theta_0) \\ \vdots & \vdots & & \vdots \\ 1 & \sin(\Theta_n - \theta_0) & \ldots & \sin^p(\Theta_n - \theta_0) \end{pmatrix}.$$

The sine-polynomial estimator of $\boldsymbol{\beta}$ can be explicitly expressed as

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{\Theta}^T \boldsymbol{W} \boldsymbol{\Theta})^{-1} \boldsymbol{\Theta}^T \boldsymbol{W} \boldsymbol{Y}. \tag{2.5}$$

As in all kernel methods, the selection of the smoothing parameter is of great importance, since it substantially affects the performance of the estimator. However, when employing circular kernels, the role of the smoothing (concentration) parameter is opposite to the role of the bandwidth when using *linear* kernels. Focusing on the estimation of $m(\cdot)$, when the concentration κ is very small, the estimation procedure leads to a global fit of a sine-polynomial of degree p, whereas if κ is very large, the estimation results in the interpolation of the data.

Di Marzio et al. (2009) showed that if $p = 1$ and the employed circular kernel is a von Mises density, the optimal concentration parameter which minimizes the Mean Squared Error (MSE) of the estimator $\hat{m}(\theta_0) = \hat{\beta}_0$ is given by

$$\left(\frac{2\sqrt{\pi} n f(\theta) [m''(\theta)]^2}{\sigma^2(\theta)} \right)^{\frac{2}{5}}. \tag{2.6}$$

However, the optimal concentration parameter depends on unknown quantities. In practice, the only proposals available in the literature to select the concentration parameter are a rule of thumb based on a preliminary parametric estimator in the directional-linear setting (García-Portugués, 2014) and a cross-validation method implemented by Oliveira et al. (2013). The cross-validation technique, which will be used throughout this chapter, selects the parameter κ which minimizes

$$\sum_{i=1}^{n} [Y_i - \hat{m}_{-i}(\Theta_i)]^2, \tag{2.7}$$

where $\hat{m}_{-i}(\cdot)$ denotes the local sine-polynomial estimator of $m(\cdot)$ constructed without the ith observation.

USC DE SANTIAGO DE COMPOSTELA

2.1.2 Circular response

Let Φ be an angular response variable and Δ a general predictor, which can be either real-valued or circular. Di Marzio et al. (2013) defined the regression function of Φ conditioned on the value of Δ as the function $m(\cdot)$ which minimizes the expectation of $d[\Phi, m(\Delta)]$, where $d(\cdot,\cdot)$ is the cosine distance defined in (1.9). For a fixed value δ_0 in the support of Δ, the quantity $\mathbb{E}\{d[\Phi, m(\delta_0)]\}$ is minimized at

$$m(\delta_0) = \text{atan2}[\mathbb{E}(\sin\Phi|\Delta = \delta_0)f(\delta_0), \mathbb{E}(\cos\Phi|\Delta = \delta_0)f(\delta_0)],$$

where $f(\cdot)$ is the density function of Δ. Consequently, the regression function for a circular response $m(\cdot)$ is the conditional circular mean of Φ given the value of Δ which, given a bivariate sample $\{(\Delta_i, \Phi_i)\}_{i=1}^n$, yields the following regression model

$$\Phi_i = [m(\Delta_i) + \varepsilon_i](\text{mod } 2\pi), \tag{2.8}$$

where ε_i are independent and identically distributed (iid) random angles with zero mean direction, finite concentration and are independent from the Δ_is. The approach of Di Marzio et al. (2013) in order to estimate $m(\cdot)$ is the following: given that $\sin\Phi, \cos\Phi \in [-1, 1]$, it is possible to estimate $\mathbb{E}(\sin\Phi|\Delta = \delta_0)$ and $\mathbb{E}(\cos\Phi|\Delta = \delta_0)$ by applying a kernel smoothing method, such as a Nadaraya-Watson or a local linear estimator if the predictor is real-valued ($\Delta = X$), or the estimator described in Section 2.1.1 if the predictor is circular ($\Delta = \Theta$). Therefore, for a fixed δ_0, $m(\delta_0)$ can be estimated as

$$\hat{m}(\delta_0) = \text{atan2}[\hat{g}_1(\delta_0), \hat{g}_2(\delta_0)], \tag{2.9}$$

where $\hat{g}_1(\delta_0)$ and $\hat{g}_2(\delta_0)$ are proportional to the kernel estimators of $\mathbb{E}(\sin\Phi|\Delta = \delta_0)$ and $\mathbb{E}(\cos\Phi|\Delta = \delta_0)$, respectively, which can be written as

$$\hat{g}_1(\delta_0) = \frac{1}{n}\sum \sin(\Phi_i)W(\Delta_i - \delta_0), \quad \hat{g}_2(\delta_0) = \frac{1}{n}\sum \cos(\Phi_i)W(\Delta_i - \delta_0).$$

The weights $W(\cdot)$ will depend on the type of kernel regression estimator. For example, if a Nadaraya-Watson type estimator is used, the weights take the form

$$W(X_i - x_0) = L_h(X_i - x_0) = \frac{1}{h}L\left(\frac{X_i - x_0}{h}\right), \quad \text{if } \Delta = X,$$

$$W(\Theta_i - \theta_0) = K_\kappa(\Theta_i - \theta_0), \quad \text{if } \Delta = \Theta.$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

On the other hand, if a local linear type estimator is chosen, the weights are given by

$$\frac{1}{n}L_h(X_i - x_0)\left[s_2(L) - (X_i - x_0)s_1(L)\right], \quad \text{if } \Delta = X,$$

$$\frac{1}{n}K_\kappa(\Theta_i-\theta_0)\left[\tilde{s}_2(K_\kappa)-\sin(\Theta_i-\theta_0)\tilde{s}_1(K_\kappa)\right], \quad \text{if } \Delta=\Theta, \tag{2.10}$$

where $s_l(L)=\sum_{k=1}^n L_h(X_k-x_0)(X_k-x_0)^l$ and $\tilde{s}_l(K_\kappa)=\sum_{k=1}^n K_\kappa(\Theta_k-\theta_0)\sin^l(\Theta_k-\theta_0)$.

The role of the smoothing parameter depends on the nature of the predictor. On the one hand, when the explanatory variable is real-valued, the bandwidth has the same role as in the local polynomial estimator for real-valued variables (see Section 1.1.1). On the other hand, when the predictor is circular, the role of the concentration parameter is as described in the previous section.

Regarding some extensions of the estimators presented above, in the setting with real-valued predictors, Meilán-Vila et al. (2021) introduced a multiple extension of the estimator in (2.9). In addition, a goodness-of-fit test for multiple regression with circular response was proposed by Meilán-Vila et al. (2022).

2.2 A test of no-effect

As mentioned in Chapter 1, a first question to answer when trying to fit a regression model is to assess if there is a significant effect of the covariate over the response. For that purpose, nonparametric no-effect tests will be proposed for the different regression scenarios involving a circular response and/or covariate.

2.2.1 Circular covariate and real-valued response

Consider the regression model in (2.1). A test to ascertain the effect of the covariate is constructed with the following hypotheses:

$$\begin{aligned} &H_0: Y_i=\mu+\varepsilon_i, \quad \mu\in\mathbb{R},\\ &H_1: Y_i=m(\Theta_i)+\varepsilon_i, \quad m(\Theta_i)\neq\mu \text{ for some } i=1,\dots,n. \end{aligned}$$

First, it will be assumed that the errors follow a normal distribution with mean zero and constant variance σ^2, although this condition will be relaxed later. A test statistic can be constructed by adapting the proposal of Bowman and Azzalini (1997) to the circular context, using the nonparametric estimator presented in Section 2.1.1. Therefore, the residual sums of squares are used to quantify how much the models explain the data under each of the two hypotheses. Then, the test statistic takes the form

$$C_1=\frac{RSS_0-RSS}{RSS},$$

where the residual sums of squares under the null and the alternative are given by

$$RSS_0 = \sum_{i=1}^{n}(Y_i - \hat{\mu})^2, \quad \text{and} \quad RSS = \sum_{i=1}^{n}[Y_i - \hat{m}(\Theta_i)]^2. \tag{2.11}$$

The constant parameter μ is estimated with the sample mean of the responses, while the regression curve under H_1 is estimated with the nonparametric estimator for circular predictors and real-valued responses ($\hat{\beta}_0$ in (2.4)). The nonparametric estimator is a linear form in the data, *i.e.*, $\hat{\boldsymbol{m}} = \boldsymbol{S}\boldsymbol{Y}$, where $\hat{\boldsymbol{m}}$ is the vector with the fitted values, $\boldsymbol{S}$ is the smoothing matrix and $\boldsymbol{Y}$ is the vector containing the responses. Consequently, the residual sums of squares can be expressed in vector-matrix notation

$$RSS_0 = \boldsymbol{Y}^T(\boldsymbol{I}_n - \boldsymbol{L})^T(\boldsymbol{I}_n - \boldsymbol{L})\boldsymbol{Y} \quad \text{and} \quad RSS = \boldsymbol{Y}^T(\boldsymbol{I}_n - \boldsymbol{S})^T(\boldsymbol{I}_n - \boldsymbol{S})\boldsymbol{Y},$$

where $\boldsymbol{L}$ is a $n \times n$ matrix with n^{-1} in all its components and $\boldsymbol{I}_n$ is the identity matrix of order n. Thus, the test statistic can be rewritten as

$$C_1 = \frac{\boldsymbol{Y}^T\boldsymbol{B}\boldsymbol{Y}}{\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}},$$

with $\boldsymbol{A} = (\boldsymbol{I}_n - \boldsymbol{S})^T(\boldsymbol{I}_n - \boldsymbol{S})$ and $\boldsymbol{B} = \boldsymbol{I}_n - \boldsymbol{L} - \boldsymbol{A}$. Now, a p-value for the test is obtained as

$$p = \mathbb{P}\left(\frac{\boldsymbol{Y}^T\boldsymbol{B}\boldsymbol{Y}}{\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}} > Obs\right) = \mathbb{P}[\boldsymbol{Y}^T(\boldsymbol{B} - \boldsymbol{A}\cdot Obs)\boldsymbol{Y} > 0],$$

with Obs being the observed value of the statistic. Although under the null hypothesis $\mathbb{E}(Y_i) = \mu$, it is easy to see that μ disappears in the expression of C_1 due to the differences involved. Then, the p-value calculation is equivalent to

$$p = \mathbb{P}[\boldsymbol{\varepsilon}^T(\boldsymbol{B} - \boldsymbol{A}\cdot Obs)\boldsymbol{\varepsilon} > 0].$$

Now, given that the matrix $\boldsymbol{B} - Obs\cdot\boldsymbol{A}$ is symmetric, it holds that $\boldsymbol{\varepsilon}^T(\boldsymbol{B} - \boldsymbol{A}\cdot Obs)\boldsymbol{\varepsilon}$ is a quadratic form in normal variables of the type $\boldsymbol{z}^T\boldsymbol{C}\boldsymbol{z}$ where $\mathbb{E}(\boldsymbol{z}) = \boldsymbol{0}$ and $\boldsymbol{C}$ is symmetric. Then, the results in Bowman and Azzalini (1997) can be applied, approximating the distribution of C_1 by a more convenient one. With that objective, note that the first three cumulants of $\boldsymbol{\varepsilon}^T(\boldsymbol{B} - \boldsymbol{A}\cdot Obs)\boldsymbol{\varepsilon}$ can be obtained as

$$\iota_s = 2^{s-1}(s-1)!\text{tr}(\boldsymbol{V}\boldsymbol{C})^s, \quad s = 1, 2, 3,$$

where $\text{tr}(\cdot)$ denotes the trace operator and $\boldsymbol{V} = \boldsymbol{Cov}(\boldsymbol{z}, \boldsymbol{z})$. Then, the distribution of $\boldsymbol{\varepsilon}^T(\boldsymbol{B} - \boldsymbol{A}\cdot Obs)\boldsymbol{\varepsilon}$ is approximated by a shifted and scaled χ^2 distribution $(c + a\chi^2_b)$, with parameters calculated as

$$a = |\iota_3|/(4\iota_2), \quad b = (8\iota_2^3)/\iota_3^2, \quad c = \iota_1 - ab,$$

USC DE SANTIAGO DE COMPOSTELA

where a is the scale parameter, c is the location parameter and b the number of degrees of freedom. Thus, the p–value can be computed as $\mathbb{P}\left(\chi_b^2 > -c/a\right)$.

Note that the Gaussian condition about the errors can be relaxed by assuming that the errors have zero mean and constant variance. Then, the calibration of the test can be done by bootstrap, similarly to the other scenarios to be presented next.

2.2.2 Circular response

Consider now the regression model in (2.8). The following hypotheses are used to determine the significance of the predictor variable:

$$\begin{aligned} H_0 &: \Phi_i = [\mu + \varepsilon_i](\text{mod}2\pi), \quad \mu \in \mathbb{T}, \\ H_1 &: \Phi_i = [m(\Delta_i) + \varepsilon_i](\text{mod}2\pi), \quad \exists\, i \mid m(\Delta_i) \neq \mu + 2l\pi \;\forall\, l \in \mathbb{Z}. \end{aligned}$$

It will be assumed that the errors have zero mean and finite and constant concentration. In the linear response case, the test statistic was built by using the quadratic distance to measure the differences between the responses and the estimated curves under each of the hypotheses. In this case, it is not possible to use such distance since it is not appropriate for the circle. Therefore, an adequate circular distance must be employed. The proposal in this work is using the distance (1.9) to construct the test statistic. As a result, the proposed statistic takes the form:

$$C_2 = \frac{\sum_{i=1}^n [1 - \cos(\Phi_i - \hat{\mu})] - \sum_{i=1}^n \{1 - \cos[\Phi_i - \hat{m}(\Delta_i)]\}}{\sum_{i=1}^n \{1 - \cos[\Phi_i - \hat{m}(\Delta_i)]\}},$$

where $\hat{\mu}$ is the sample mean direction of the responses, given by

$$\hat{\mu} = \text{atan2}\left(\sum_{i=1}^n \sin\Phi_i, \sum_{i=1}^n \cos\Phi_i\right)$$

and $\hat{m}(\cdot)$ is the nonparametric estimator for circular responses (2.9).

The distribution of C_2 under H_0 is approximated through bootstrap methods. The resampling strategy is specified hereafter.

(i) Given a smoothing parameter h or κ (depending on the nature of the predictor variable), compute the value of the statistic C_2 for the data, denoted by Obs.

(ii) From the computed values of $\hat{\mu}$, obtain the residuals under the null hypothesis ($\hat{\varepsilon}_i = \Phi_i - \hat{\mu}, i = 1, \ldots, n$) and construct the resampled responses as $\Phi_i^* = [\hat{\mu} + \hat{\varepsilon}_i^*](\text{mod } 2\pi)$, where $\hat{\varepsilon}_i^*$ are obtained from sampling the residuals randomly with replacement.

(iii) With the same smoothing parameter as in (i), compute the value of the test statistic for the bootstrap resample, $C_2^{*(b)}$.

(iv) Repeat (ii) and (iii) B times to obtain $C_2^{*(1)}, ..., C_2^{*(B)}$, and approximate the critical value as $\sum_{b=1}^{B} \mathbf{1}_{\{C_2^{*(b)} \geqslant Obs\}}/B$, where $\mathbf{1}_A$ denotes the indicator function of A.

It should be noticed that, as in any nonparametric test (see Bowman and Azzalini, 1997), the outcome may be influenced by the smoothing parameter. An optimal smoothing parameter in terms of estimation might not be suitable for hypotheses testing, because of the bias present in the estimation of $m(\cdot)$. In practice, it is recommended to run the test over a sequence of smoothing parameters in a reasonable range. In the simulation study carried out in Section 2.4, the performance of the tests is analyzed when employing different smoothing parameters derived from a cross-validation bandwidth/concentration.

2.3 A nonparametric analysis of covariance model

In this section, analysis of covariance (ANCOVA) models for circular regression are introduced, in addition to testing tools for equality and parallelism. First, the setting involving a circular predictor and a real-valued response is considered, while the circular response scenarios are analyzed later. A categorical covariate inducing J groups will be now introduced in the model, each one identified by the subscript $j = 1, \ldots, J$. The number of data points in the jth group will be denoted by n_j.

2.3.1 Circular covariate and real-valued response

An ANCOVA regression model for the circular-linear regression scenario is formulated as

$$Y_{ji} = m_j(\Theta_{ji}) + \varepsilon_{ji}, \quad j = 1, \ldots, J, \ i = 1, \ldots, n,$$

where the ε_{ji} are, first, assumed to follow a $N(0, \sigma^2)$ distribution. In the following, two tests will be proposed, one for equality and one for parallelism.

Test of equality

The equality of the curves is tested through the following hypotheses statement:

$$\begin{aligned} H_0 &: Y_{ji} = m(\Theta_{ji}) + \varepsilon_{ji}, \quad \forall\, j = 1, \ldots, J, \\ H_1 &: Y_{ji} = m_j(\Theta_{ji}) + \varepsilon_{ji}, \quad \exists\, j, k = 1, \ldots, J \mid m_j(\cdot) \neq m_k(\cdot). \end{aligned}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The corresponding test statistic takes the form

$$C_3 = \frac{1}{\hat{\sigma}^2} \sum_{j=1}^{J} \sum_{i=1}^{n_j} [\hat{m}_j(\Theta_{ji}) - \hat{m}(\Theta_{ji})]^2.$$

A key point to obtain the expression of C_3 is the estimation of the variance. The circular nature of the predictor must be considered, given that it plays an important role when computing $\hat{\sigma}^2$. The estimation of the variance is carried out by using periodic pseudoresiduals, which are a modification of the pseudoresiduals defined in Gasser et al. (1986). Let $Y_{j[i]}$, with $i = 1, \ldots, n_j$, denote the value of Y corresponding to $\Theta_{j[i]}$, where $\Theta_{j[i]}$ represents the ith smallest value on the real line of the sample from Θ in group j (given that an origin has been chosen). The new pseudoresiduals are defined as

$$\tilde{\varepsilon}_{j[i]} = \frac{\Theta_{j[i+1]} - \Theta_{j[i]}}{\Theta_{j[i+1]} - \Theta_{j[i-1]}} Y_{j[i-1]} + \frac{\Theta_{j[i]} - \Theta_{j[i-1]}}{\Theta_{j[i+1]} - \Theta_{j[i-1]}} Y_{j[i+1]} - Y_{j[i]},$$

with $j = 1, \ldots, J$, $i = 1, \ldots, n_j$. Here, it holds that $Y_{j[n_j+1]} = Y_{j[1]}$, $Y_{j[n_j+2]} = Y_{j[2]}$ and $Y_{j[0]} = Y_{j[n_j]}$. The periodic pseudoresiduals can then be expressed as $\tilde{\varepsilon}_{j[i]} = a_{j[i]} Y_{j[i-1]} + b_{j[i+1]} Y_{j[i+1]} - Y_{j[i]}$, and thus, the variance in each group and the total variance are estimated, respectively, as

$$\hat{\sigma}_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} \frac{1}{c_{j[i]}^2} \tilde{\varepsilon}_{j[i]}^2 \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n - J} \sum_{j=1}^{J} n_j \hat{\sigma}_j^2,$$

where $c_{j[i]}^2 = a_{j[i]}^2 + b_{j[i]}^2 + 1$, $j = 1, \ldots, J$, $i = 1, \ldots, n_j$. After some calculations, this estimator can be written in matrix-vector notation as $\boldsymbol{Y}^T \boldsymbol{K} \boldsymbol{Y} = \boldsymbol{Y}^T \boldsymbol{P}^T \boldsymbol{P} \boldsymbol{Y}$, where $\boldsymbol{P}$ is a $n \times n$ block matrix.

Test of parallelism

For testing parallel regression curves, the following hypotheses are used:

$$H_0 : Y_{ji} = \gamma_j + m(\Theta_{ji}) + \varepsilon_{ji}, \quad \gamma_1 = 0, \gamma_j \in \mathbb{R}, \forall\, j = 1, \ldots, J,$$
$$H_1 : Y_{ji} = m_j(\Theta_{ji}) + \varepsilon_{ji}, \quad \exists\, j, k = 1, \ldots, J \mid m_j(\cdot) \neq m_k(\cdot) + \gamma \;\forall\, \gamma \in \mathbb{R}.$$

The next statistic is used to test the differences between the models under each one of the two hypotheses:

$$C_4 = \frac{1}{\hat{\sigma}^2} \sum_{j=1}^{J} \sum_{i=1}^{n_j} [\hat{\gamma}_j + \hat{m}(\Theta_{ji}) - \hat{m}_j(\Theta_{ji})]^2.$$

For estimating the shift parameter γ, the model is written in vector-matrix notation:

$$\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{\gamma} + \boldsymbol{m} + \boldsymbol{\varepsilon}, \tag{2.12}$$

where $\boldsymbol{D}$ is a known matrix consisting of 0s and 1s. Given a vector $\boldsymbol{\gamma}$, an estimate of the regression function can be constructed:

$$\hat{\boldsymbol{m}} = \boldsymbol{S}(\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{\gamma}),$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

with $\boldsymbol{S}$ being a smoothing matrix constructed with the circular-linear regression method. Substituting this estimator in equation (2.12) and applying the least-squares method, an estimate of $\boldsymbol{\gamma}$ is derived:

$$\hat{\boldsymbol{\gamma}} = [\boldsymbol{D}^T(\boldsymbol{I}_n - \boldsymbol{S_1})^T(\boldsymbol{I}_n - \boldsymbol{S_1})\boldsymbol{D}]^{-1}\boldsymbol{D}^T(\boldsymbol{I}_n - \boldsymbol{S_1})^T(\boldsymbol{I}_n - \boldsymbol{S_1})\boldsymbol{Y} = \widetilde{\boldsymbol{W}}\boldsymbol{Y},$$

where $\boldsymbol{S_1}$ is a preliminary smoothing matrix. After $\hat{\boldsymbol{\gamma}}$ is obtained, the regression function $m(\cdot)$ evaluated at the sample points is estimated as

$$\hat{\boldsymbol{m}} = \boldsymbol{S}(\boldsymbol{Y} - \boldsymbol{D}\hat{\boldsymbol{\gamma}}).$$

However, for estimating the vector of parameters $\hat{\boldsymbol{\gamma}}$ it is necessary to choose a first smoothing parameter κ_1, independent of the one used to estimate the actual curves, and the selection of this parameter is not trivial. In the real-valued case, Bowman and Azzalini (1997) used $2R/n$, where R is the range of the design points, as this choice restricts the smoothing to approximately eight neighboring observations when the data are equally spaced (if a normal kernel is used).

With the objective of finding an adequate automatic rule in the circular setting, a local smoothing parameter, which showed a good performance in practice, is employed. Let $d_2(\cdot,\cdot)$ be defined as

$$d_2(\Phi,\Theta) = \min\{|\Phi - \Theta|, 2\pi - |\Phi - \Theta|\}, \quad \Phi,\Theta \in \mathbb{T},$$

i.e., $d_2(\cdot,\cdot)$ is the geodesic distance. The new proposal consists in finding a preliminary vector of smoothing parameters, $\boldsymbol{h_1}$, containing one parameter for each observation, in which the parameter associated to observation Θ_{ji}, $h_{1;ji}$, will be the distance to its 8th nearest neighbor (considering distance $d_2(\cdot,\cdot)$). Then, $\boldsymbol{h_1}$ is used to obtain a vector of smoothing parameters valid for the circular case using the results in Gumbel et al. (1953), which show that for large values of κ the von Mises vM(μ,κ) converges in distribution to a $N(\mu, 1/\sqrt{\kappa})$. Thus, if $h_{1;ji}$ is the preliminary smoothing parameter corresponding to Θ_{ji}, the concentration parameter for this observation will be $\kappa_{1;ji} = 1/h_{1;ji}^2$.

Distribution of the statistics

In order to obtain the distributions of C_3 and C_4 under H_0, it must be noted that their numerators can be expressed, respectively, as

$$\boldsymbol{Y}^T(\boldsymbol{S_d} - \boldsymbol{S})^T(\boldsymbol{S_d} - \boldsymbol{S})\boldsymbol{Y}$$

and

$$\boldsymbol{Y}^T[\boldsymbol{D}\widetilde{\boldsymbol{W}} + \boldsymbol{S}(\boldsymbol{I}_n - \boldsymbol{D}\widetilde{\boldsymbol{W}}) - \boldsymbol{S_d}]^T[\boldsymbol{D}\widetilde{\boldsymbol{W}} + \boldsymbol{S}(\boldsymbol{I}_n - \boldsymbol{D}\widetilde{\boldsymbol{W}}) - \boldsymbol{S_d}]\boldsymbol{Y},$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

where $\boldsymbol{S}$ is the smoothing matrix under the equality (or parallelism) assumption and $\boldsymbol{S_d}$ is the block matrix constructed with the smoothing matrices for each group. Then, both statistics can be expressed in the form $\boldsymbol{Y}^T\boldsymbol{Q}\boldsymbol{Y}/\boldsymbol{Y}^T\boldsymbol{G}\boldsymbol{Y}$, where $\boldsymbol{Q}$ is a symmetric matrix and $\boldsymbol{G}$ is obtained straightforwardly from the variance estimator. Now, under the null hypothesis, if the same concentration parameter is used for obtaining the matrices $\boldsymbol{S}$ and $\boldsymbol{S_d}$, by analyzing the bias terms of the circular-linear estimator ($\hat{\beta}_0$ in (2.4)), it can be shown that the distribution of $\boldsymbol{Y}^T\boldsymbol{Q}\boldsymbol{Y}/\boldsymbol{Y}^T\boldsymbol{G}\boldsymbol{Y}$ is almost equivalent to the distribution of $\boldsymbol{\varepsilon}^T\boldsymbol{Q}\boldsymbol{\varepsilon}/\boldsymbol{\varepsilon}^T\boldsymbol{G}\boldsymbol{\varepsilon}$. Then, using the first three cumulants of $\boldsymbol{\varepsilon}^T(\boldsymbol{Q}-\boldsymbol{G}\cdot Obs)\boldsymbol{\varepsilon}$ (where Obs is the observed value of C_3 or C_4), the shifted and scaled χ^2 approximation described in Section 2.2 can be employed.

Note that conditions over the residuals can be relaxed, assuming only zero mean and constant variance. In such case, the distribution of the statistics can be obtained through bootstrap methods, in a similar way as in the following scenarios.

2.3.2 Circular response

An ANCOVA model for a circular response variable can be expressed as

$$\Phi_{ji} = [m_j(\Delta_{ji}) + \varepsilon_{ji}](\text{mod}2\pi), \quad j = 1, \ldots, J,\ i = 1, \ldots, n_j.$$

It will be assumed that the errors ε_{ji} have zero mean and finite and constant concentration κ. The tests of equality and parallelism are presented next.

Test of equality

The hypotheses stated for testing the equality of the curves are

$$H_0 : \Phi_{ji} = [m(\Delta_{ji}) + \varepsilon_{ji}](\text{mod}2\pi), \quad \forall\, j = 1, \ldots, J,$$
$$H_1 : \Phi_{ji} = [m_j(\Delta_{ji}) + \varepsilon_{ji}](\text{mod}2\pi), \quad \exists\, j, k = 1, \ldots, J \mid m_j(\cdot) \neq m_k(\cdot) + 2l\pi\ \forall\, l \in \mathbb{Z}.$$

In pursuance of constructing the test statistic, one needs to measure the differences between the estimated regression functions under each of the hypotheses. Since $\hat{m}(\cdot)$ and $\hat{m}_j(\cdot)$ have circular nature, the distance defined in (1.9) will be adequate to construct the test statistic. However, a second problem is encountered, as a dispersion measure is needed to normalize the statistic. A new approach is to estimate the circular variance (1.8). Consequently, the proposed test statistic takes the form

$$C_5 = \frac{1}{\overline{D}} \sum_{j=1}^{J} \sum_{i=1}^{n_j} \left\{1 - \cos[\hat{m}_j(\Delta_{ji}) - \hat{m}(\Delta_{ji})]\right\},$$

where $\bar{D}$ is an estimator of the circular variance defined as

$$\bar{D} = \frac{1}{n-J} \sum_{j=1}^{J} \sum_{i=1}^{n_j} \{1 - \cos[\Phi_{ji} - \hat{m}_j(\Delta_{ji})]\}.$$

Test of parallelism

When considering a circular response, both for circular or real-valued covariates, the regression curves might be parallel in a way in which the shape of the regression function is the same for all groups except for an angular shift. This behavior can be tested with the following hypotheses statement:

$$H_0 : \Phi_{ji} = [\gamma_j + m(\Delta_{ji}) + \varepsilon_{ji}](\text{mod}2\pi), \quad \gamma_1 = 0,\ \gamma_j \in \mathbb{T},\ \forall\ j = 1, \ldots, J,$$
$$H_1 : \Phi_{ji} = [m_j(\Delta_{ji}) + \varepsilon_{ji}](\text{mod}2\pi), \quad \exists\ j, k = 1, \ldots, J \mid m_j(\cdot) \neq [m_k(\cdot) + \gamma](\text{mod}2\pi).$$

As before, the circular distance (1.9) is used for constructing the test statistic.

$$C_6 = \frac{1}{\bar{D}} \sum_{j=1}^{J} \sum_{i=1}^{n_j} \{1 - \cos[\hat{\gamma}_j + \hat{m}(\Delta_{ji}) - \hat{m}_j(\Delta_{ji})]\}.$$

Again, the shift parameters must be estimated. As a first step, a first smoothing parameter needs to be selected to obtain the preliminary estimator of the global regression function, namely $\hat{m}^1(\cdot)$, to minimize the bias in the preliminary estimation of $m(\cdot)$ and therefore in the estimation of $\gamma_1, ..., \gamma_J$. Although it is recommended to explore several parameters, an automatic rule was derived. When the predictors are linear ($\Delta = X$), the rule consists of using a vector of smoothing parameters in which each of them corresponds to one observation. Each parameter will be the distance to the 8th nearest observation. On the other hand, in the case where the predictor is of a circular nature ($\Delta = \Theta$), the rule is the same as in the test of parallelism for circular-linear regression (Section 2.3.1). Then, the parameters $\gamma_1, \ldots, \gamma_J$ are estimated by solving the minimization problem

$$\begin{aligned} &\arg \min_{\gamma_1,\ldots,\gamma_J} \sum_{j=1}^{J} \sum_{i=1}^{n_j} \{1 - \cos[\Phi_{ji} - \gamma_j - \hat{m}^1(\Delta_{ji})]\} \\ &\text{s.t.} \qquad \gamma_j \in \mathbb{T},\ \forall\ j = 1, \ldots, J. \end{aligned} \tag{2.13}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Proposition 2.1. *For $j = 1, \ldots, J$, let $\bar{C}_j$ and $\bar{S}_j$ denote*

$$\bar{C}_j = \sum_{i=1}^{n_j} \cos[\Phi_{ji} - \hat{m}^1(\Delta_{ji})] \text{ and } \bar{S}_j = \sum_{i=1}^{n_j} \sin[\Phi_{ji} - \hat{m}^1(\Delta_{ji})].$$

If, for all $j = 1, \ldots, J$, it holds $\bar{C}_j \neq 0$ or $\bar{S}_j \neq 0$, then $(\hat{\gamma}_1, \ldots, \hat{\gamma}_J)$, where $\hat{\gamma}_j = \text{atan}2(\bar{S}_j, \bar{C}_j)$,

$j = 1, \ldots, J$, *is a global minimum of minimization problem* (2.13).

The proof of Proposition 2.1 is given in Appendix C.1. Note that the estimators $\hat{\gamma}_1, ..., \hat{\gamma}_J$ obtained will not be unbiased, due to the bias of the preliminary estimator $\hat{m}^1(\cdot)$. However, the bias is smaller as the sample size increases.

Distribution of the statistics

The distribution of C_5 and C_6 under H_0 is obtained with bootstrap methods. The resampling strategy is described next.

(i) Choose a smoothing parameter, for example the one selected by cross-validation, to obtain the estimators $\hat{m}(\cdot)$ and $\hat{m}_1(\cdot), ..., \hat{m}_J(\cdot)$.

(i, for the test of parallelism) Choose also a preliminary smoothing parameter h_1 or κ_1 (depending on the nature of the explanatory variable) and obtain the nonparametric estimator $\hat{m}^1(\cdot)$ and the shift parameter estimator $\hat{\boldsymbol{\gamma}}$.

(ii) Compute the observed value of statistic C_5 or C_6, namely Obs.

(iii) Obtain the residuals under the null hypothesis ($\hat{\varepsilon}_{ji}$) and construct the resampled responses (Φ^*_{ji}) from the bootstrap residuals, $\hat{\varepsilon}^*_{ji}$, obtained from sampling the residuals randomly with replacement.

$$\text{(Equality)} : \hat{\varepsilon}_{ji} = \Phi_{ji} - \hat{m}(\Delta_{ji}) \text{ and } \Phi^*_{ji} = [\hat{m}(\Delta_{ji}) + \hat{\varepsilon}^*_{ji}] (\text{mod } 2\pi).$$
$$\text{(Parallelism)} : \hat{\varepsilon}_{ji} = \Phi_{ji} - \hat{\gamma}_j - \hat{m}(\Delta_{ji}) \text{ and } \Phi^*_{ji} = [\hat{\gamma}_j + \hat{m}(\Delta_{ji}) + \hat{\varepsilon}^*_{ji}] (\text{mod } 2\pi).$$

(iv) Using the smoothing parameter employed in (i) for estimating $\hat{m}(\cdot)$, compute the value of the test statistic for the bootstrap resample, $C_5^{*(b)}$ or $C_6^{*(b)}$.

(iv, for the test of parallelism) Use also the preliminary smoothing parameter obtained in (i).

(v) Repeat (iii) and (iv) B times to approximate the critical value as $\sum_{b=1}^{B} \mathbf{1}_{\{C_5^{*(b)} \geqslant Obs\}}/B$ or $\sum_{b=1}^{B} \mathbf{1}_{\{C_6^{*(b)} \geqslant Obs\}}/B$.

2.4 Simulation study

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The finite sample performance of the tests, both in terms of calibration and power, is explored by simulation. The no-effect tests, for the different regression scenarios, are investigated first. Tests for equality and parallelism are also analyzed for all the scenarios. In all settings, it was chosen that the circular kernels were taken as von Mises densities, while the normal density

was selected as the linear kernel. In addition, in the scenarios where the nonparametric estimator for a real-valued response ($\hat{\beta}_0$ in (2.4)) is used, the sine-polynomial with $p = 1$ was considered. For the kernel estimators for a circular response, the local linear weights (2.10) were employed.

2.4.1 Results for the no-effect tests

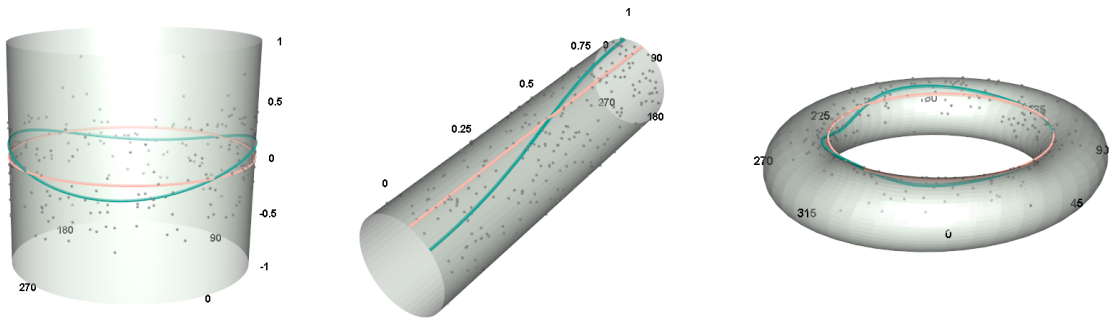
For the no-effect tests, data from different regression models (depending on the scenario) are simulated, with sample size $n \in \{50, 100, 250, 400\}$. The regression models for the different scenarios are the following:

- *Circular-linear:* $Y = \beta \sin\Theta \cos\Theta + \varepsilon, \quad \beta = 0, 0.25, 0.5.$
- *Linear-circular:* $\Phi = [3\pi/8 + \beta \cos(3X) + \varepsilon] (\text{mod } 2\pi)$, $\beta = 0, 0.5, 1.$
- *Circular-circular:* $\Phi = [3\pi/4 + \beta \sin(2\Theta + 2\sin(\Theta + \pi/2)) + \varepsilon] (\text{mod } 2\pi)$, $\beta = 0, 0.35, 0.5.$

Circular predictor variables were drawn from circular uniform distributions while real-valued predictors were drawn from a $U(0, 1)$ distribution. When using the first value of β in each model, data are simulated under the null hypothesis of no-effect of the predictor over the response. With the other two values of β, the alternative hypothesis holds. For the real-valued response case, the errors are drawn from a normal distribution with zero mean and standard deviation 0.25, which enables calibration by a shifted and scaled χ^2 distribution. Rescaled exponential errors with rate parameter 5 are also used, and in this case calibration is done through a bootstrap procedure. When the response is circular, the errors are drawn from von Mises distributions with zero mean direction. Although different values of the error concentration were analyzed, only the results for $\kappa = 2$ (for the model with real-valued predictors) and $\kappa = 4$ (for the model with circular covariate) are shown here. Realizations of the simulated data for the second value of β are represented in Figure 2.1, where the true regression function was also depicted in green, and the curve indicating the no-effect hypothesis was added in salmon color.

While the calibration of the tests remains the same regardless of the concentration, the power is affected, as higher concentrations lead to higher percentages of rejection. For calculating the percentage of rejections, the number of Monte Carlo samples is 500. For the bootstrap procedure, the number of bootstrap replicates is 500.

As mentioned in Section 2.2, the outcome of the tests may be seriously influenced by the smoothing parameter. Here, the performance of the tests is studied when the smoothing parameter is selected by cross-validation (cv) and when other parameters which undersmooth or oversmooth the regression estimators are used. Although different values of the smoothing parameters were employed, when the covariate is circular, only the results obtained with $cv/8$ and $4cv$ (as the parameters which respectively oversmooth and undersmooth the estimated



(a) Circular-linear *(b) Linear-circular* *(c) Circular-circular*

Figure 2.1: Scatter plots of $n = 250$ sample points generated from the simulated models for the no-effect test model, with the true regression function (green) and the curve corresponding to H_0 (salmon). The considered parameters are $\beta = 0.25$ (circular-linear), $\beta = 0.5$ (linear-circular) and $\beta = 0.35$ (circular-circular).

curve) are presented. In the linear-circular case, since the kernel used is linear, the parameter $4cv$ is considered for an oversmoothed estimator and $cv/4$ for an undersmoothed curve. The choice of these values is motivated by its performance in practice, since they are the values that presented a correct calibration of the tests jointly with the highest power, even when varying the error variance or error concentration.

Percentages of rejection of the tests for a significance level of $\alpha = 0.05$ are displayed in Table 2.1 for different sample sizes. Further simulation results for significance levels $\alpha = 0.10$ and $\alpha = 0.01$ are provided in Tables B.1 and B.2 of Appendix B.1. In what follows, it will be referred to the results for $\alpha = 0.05$. Focusing on the calibration of the tests, the smoothing parameters obtained by cross-validation do not provide an appropriate calibration of the test under the null hypothesis, given that percentages of rejection obtained with this parameter are around 10% in all cases for the first value of β. However, when using the other values of the bandwidth, percentages of rejection are close to the significance level $\alpha = 0.05$, being just slightly conservative when considering $4cv$ for $n = 50$ in the circular-linear context.

On the other hand, the performance of the tests under the alternative is shown when the second and third values of β are considered. In such cases, percentages of rejection tend to one as the sample size increases. The best performance is obtained, in general, when considering cv as the smoothing parameter. Focusing on the two cases where the test seems to be well calibrated, the best performance is obtained when using an undersmoothed estimator. In that case, in the studied scenarios, the percentage of rejections is above 0.2, when $n = 50$, 0.5, when $n = 100$ and above 0.98, when $n = 250$.

With the objective of comparing the two calibration alternatives in the circular-linear test, in Table B.5 of Appendix B.1, the shifted and scaled χ^2 test is also employed with errors

Table 2.1: Percentages of rejection (for $\alpha = 0.05$) obtained with the no-effect tests using different smoothing parameters. Results for the first value of β show empirical size, whereas results for the other values of β show empirical power.

Circular-linear regression. χ^2 calibration. Normal errors

	$4cv$			cv			$cv/8$		
n	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$
50	0.016	0.222	0.806	0.080	0.440	0.952	0.044	0.166	0.606
100	0.040	0.564	0.998	0.078	0.736	1	0.052	0.286	0.976
250	0.062	0.984	1	0.094	0.998	1	0.052	0.844	1
400	0.064	1	1	0.106	1	1	0.060	0.988	1

Circular-linear regression. Bootstrap calibration. Exponential errors

	$4cv$			cv			$cv/8$		
n	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$
50	0.040	0.396	0.922	0.090	0.648	0.994	0.058	0.294	0.814
100	0.034	0.820	1	0.094	0.916	1	0.046	0.566	0.998
250	0.054	1	1	0.094	1	1	0.052	0.980	1
400	0.058	1	1	0.096	1	1	0.042	1	1

Linear-circular regression. Von Mises errors

	$cv/4$			cv			$4cv$		
n	$\beta=0$	$\beta=0.5$	$\beta=1$	$\beta=0$	$\beta=0.5$	$\beta=1$	$\beta=0$	$\beta=0.5$	$\beta=1$
50	0.042	0.410	0.980	0.070	0.494	0.772	0.056	0.474	0.756
100	0.066	0.734	1	0.100	0.808	0.974	0.068	0.802	0.974
250	0.044	0.978	1	0.066	0.990	1	0.046	0.990	1
400	0.044	1	1	0.082	1	1	0.054	1	1

Circular-circular regression. Von Mises errors

	$4cv$			cv			$cv/8$		
n	$\beta=0$	$\beta=0.35$	$\beta=0.5$	$\beta=0$	$\beta=0.35$	$\beta=0.5$	$\beta=0$	$\beta=0.35$	$\beta=0.5$
50	0.048	0.244	0.462	0.118	0.436	0.712	0.068	0.298	0.480
100	0.038	0.598	0.928	0.088	0.768	0.980	0.038	0.488	0.834
250	0.048	0.986	1	0.088	0.998	1	0.044	0.934	1
400	0.058	1	1	0.102	1	1	0.058	0.998	1

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

generated by the exponential distribution and the bootstrap calibration of the test is used with the normal errors. In general, similar results are obtained with both calibration methods. The χ^2 test seems to be well calibrated (for $4cv$ and $cv/8$), even when errors are generated from the exponential distribution. Regarding the power, a slightly better behavior is observed, in general, with the bootstrap calibrated test. When the objective is having a more efficient test (in computational terms), if the errors are normally distributed, since both tests provide very similar percentages of rejections, the χ^2 is the recommended calibration alternative.

2.4.2 Results for the ANCOVA tests

The performance of the equality and parallelism tests will be illustrated in this section. For that purpose, data will be simulated from different models depending on the regression scenario, where each model has two groups with sample sizes

$$(n_1, n_2) \in \{(50, 50), (50, 100), (100, 100), (100, 250), (250, 250)\}.$$

The regression models are the following:

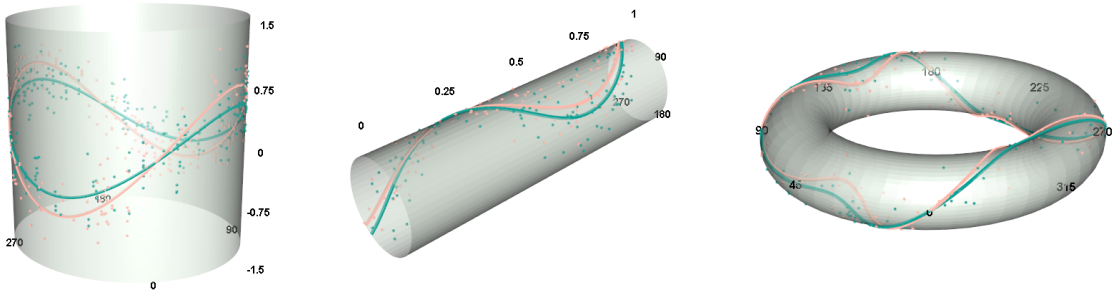
- *Circular-linear:*
  Group 1: $Y = \cos\Theta \sin\Theta + \varepsilon$,
  Group 2: $Y = \beta \cos\Theta \sin\Theta + \varepsilon, \quad \beta = 1, 1.5, 1.75$

- *Linear-circular:*
  Group 1: $\Phi = [2\sin(4X - 1) + \varepsilon] (\text{mod } 2\pi)$,
  Group 2: $\Phi = [\beta \sin(4X - 1) + \varepsilon] (\text{mod } 2\pi), \quad \beta = 2, 1.75, 1.5.$

- *Circular-circular:*
  Group 1: $\Phi = [2\sin(2\Theta) + \varepsilon] (\text{mod } 2\pi)$,
  Group 2: $\Phi = [\beta \sin(2\Theta) + \varepsilon] (\text{mod } 2\pi), \quad \beta = 2, 2.5, 3.$

Circular predictor variables were drawn from circular uniform distributions while real-valued predictors were drawn from a $U(0, 1)$ distribution. For the test of parallelism the same models are used, but a shift is added to the responses in the second group. The value of the shift parameter is 0.2 in the circular-linear case and $\pi/8$ in the tests for circular responses. As before, when the first value of β is used the data are drawn from the null hypothesis and the alternative is considered if any of the other two values of β is used. In the circular-linear regression case, the χ^2 calibration is applied to the simulated data with normally distributed errors, with zero mean and standard deviation 0.25. Rescaled exponential errors with rate parameter 5 are also used, calibrating the distribution of the tests with the bootstrap procedure. For the tests with circular responses, the errors are simulated from a von Mises distribution with mean zero and concentration parameter $\kappa = 6$ for the test with a real-valued covariate and $\kappa = 4$ in the circular-circular case. The number of samples, as well as the number of bootstrap replicates, is fixed to 500. Figure 2.2 shows realizations of the data ($n_1 = n_2 = 100$)

drawn from the simulated models, under the hypothesis of different regression curves (second value of β in each model), along with the true regression curves for each group.



(a) Circular-linear *(b) Linear-circular* *(c) Circular-circular*

Figure 2.2: Scatter plots of sample points generated from the simulated models with parameters $\beta = 1.5$ (circular-linear), $\beta = 1.75$ (linear-circular) and $\beta = 2.5$ (circular-circular) and with the true regression functions. The sample size is 100 for each group.

Percentages of rejection for a significance level of $\alpha = 0.05$ are shown in Table 2.2 for different samples sizes, although results for $\alpha = 0.10$ and $\alpha = 0.01$ can be found, respectively, in Tables B.3 and B.4 of Appendix B.1. In this case, the applied smoothing parameter was the one obtained by cross-validation (cv), but the performance of the tests with bandwidths that either undersmooth or oversmooth the estimated regression curves (see Tables B.7 and B.8 of Appendix B.1) was also explored.

Regarding the calibration of the tests, percentages of rejection lie around 5% in all scenarios when $\alpha = 0.05$ and the cv smoothing parameter are considered. Referring to the power of the tests, when the data are drawn from the alternative hypothesis, percentages of rejection are closer to 1 as the sample size increases. However, the bootstrap calibration of the circular-linear test applied to exponential errors obtains low percentages of rejection (between 10% and 20%) for $n_1 = 50$. In all the other studied scenarios, the percentage of rejections is above 0.3, when considering the case $n_1 = n_2 = 50$, above 0.6, when $n_1 = n_2 = 100$, and above 0.97, when $n_1 = n_2 = 250$.

In Table B.6 of Appendix B.1, it can be observed that the aforementioned behavior of the bootstrap parallelism test for circular-linear regression is also obtained when generating errors from the normal distribution (the percentage of rejections for $n_1 = 50$ under H_1 is around 10%). Thus, it seems that a worse behavior is obtained when employing the bootstrap calibration instead of the χ^2 test and the errors follow a normal distribution. Regarding the calibration of the χ^2 test when errors follow the exponential distribution, it was obtained that percentages of rejection under H_0 are slightly higher than $\alpha = 0.05$ (around 7% or 8%). Therefore, due to the anticonservative behavior, it is recommended that the calibration by the shifted and scaled χ^2 distribution is only used when the normality assumption holds.

As for the percentages of rejection obtained with other values of the smoothing parameters, the results in Appendix B.1 (Tables B.7 and B.8) show that with undersmoothed estimated regression curves, the percentages of rejection lie around the nominal level $\alpha = 0.05$ under H_0, although in that case the power of the tests is lower than when using the smoothing parameter selected by cross-validation. On the other hand, when oversmoothing, the tests are not well calibrated under H_0, obtaining very large percentages of rejection (even surpassing 30% of rejections in some cases).

2.5 Real data illustration

To show the applicability of the methods proposed in this chapter, this section contains an illustration of the no-effect, equality and parallelism tests with three different data sets, one corresponding to the mechanical field and two related to biological disciplines.

Flywheel data set

Consider the flywheels data set, described in Appendix A.1, consisting on the angle of imbalance (circular variable) of 60 flywheel devices, the magnitude of correction needed for balance (real-valued variable) and a categorical variable indicating the metal employed for the production. Although the sample size for each group is small ($n_j = 15$, with $j = 1, 2, 3, 4$), this example is just meant to illustrate the previously proposed techniques.

A single nonparametric regression model can be constructed, without considering the different groups, as in the left panel of Figure 2.3 (continuous line), where the regression function was estimated with the circular-linear nonparametric estimator with $p = 1$ (see Section 2.1) using all the data. The dashed line represents the average of the responses, corresponding to the estimated model under the hypothesis of no-effect of the covariate. The nonparametric estimation of the regression function changes for the different values of the predictor variable, but it could be possible that the responses did not depend on the angle of imbalance and that the features of the curve were due to sample noise. To ascertain this, the no-effect test for circular predictors (presented in Section 2.2) is applied to the data. The Kolmogorov-Smirnoff test was used to test the normality of the residuals, obtaining that the normality assumption was not rejected. Therefore, the χ^2 approximation is used, but similar results are obtained with the bootstrap version of the test. A range of smoothing parameters between 0 and 15 was considered, obtaining a p-value for each bandwidth. The results, showed in the top panel of Figure 2.4, were lower than 0.05 for all the concentration parameters considered, concluding, for this significance level, that the angle of imbalance is significant.

USC DE COMPOSTELA

However, since the metal used in the molding is different, it could be possible to have different regression curves for the groups. The right panel in Figure 2.3 shows the data represented with different colors for each of the groups, with their corresponding estimated

Table 2.2: Percentages of rejection (for $\alpha = 0.05$) obtained with the ANCOVA tests using the smoothing parameters obtained by cross-validation. Results for the first value of β show empirical size, whereas results for the other values of β show empirical power.

Circular-linear regression. χ^2 calibration. Normal errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
50	50	0.054	0.522	0.916	0.054	0.592	0.928
50	100	0.040	0.718	0.978	0.046	0.732	0.978
100	100	0.072	0.902	1	0.054	0.926	1
100	250	0.050	0.984	1	0.056	0.986	1
250	250	0.044	1	1	0.054	1	1
Circular-linear regression. Bootstrap calibration. Exponential errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
50	50	0.040	0.744	0.970	0.076	0.120	0.216
50	100	0.056	0.844	0.994	0.054	0.132	0.204
100	100	0.046	0.980	1	0.054	0.570	0.780
100	250	0.058	1	1	0.062	1	1
250	250	0.052	1	1	0.058	1	1
Linear-circular regression. Von Mises errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 2$	$\beta = 1.75$	$\beta = 1.5$	$\beta = 2$	$\beta = 1.75$	$\beta = 1.5$
50	50	0.052	0.304	0.898	0.044	0.382	0.938
50	100	0.066	0.450	0.976	0.062	0.520	0.988
100	100	0.050	0.650	0.998	0.058	0.668	1
100	250	0.046	0.784	1	0.048	0.830	1
250	250	0.046	0.976	1	0.056	0.970	1
Circular-circular regression. Von Mises errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 2$	$\beta = 2.5$	$\beta = 3$	$\beta = 2$	$\beta = 2.5$	$\beta = 3$
50	50	0.060	0.408	0.966	0.056	0.384	0.968
50	100	0.062	0.498	0.994	0.062	0.562	0.996
100	100	0.052	0.834	1	0.048	0.838	1
100	250	0.064	0.942	1	0.072	0.950	1
250	250	0.064	1	1	0.034	0.998	1

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

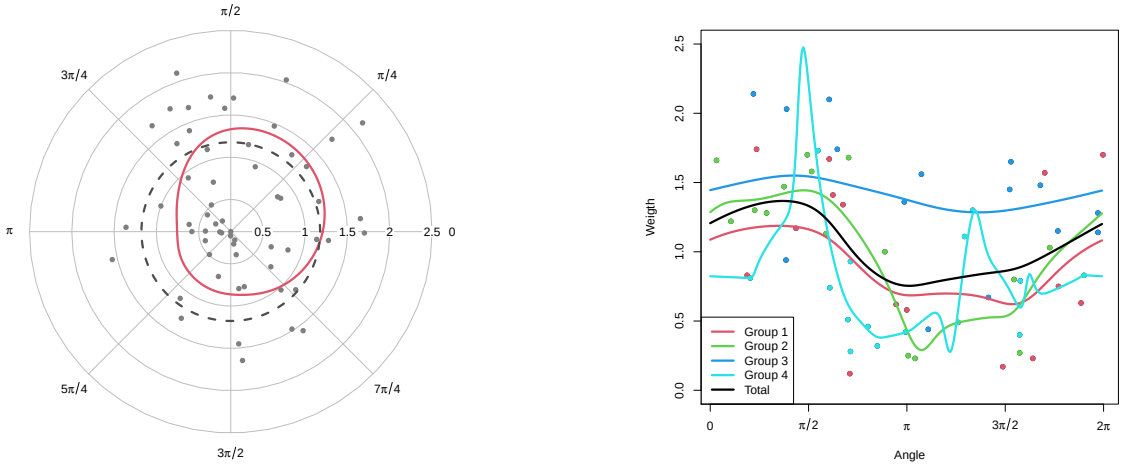


Figure 2.3: Scatter plots of the angle of imbalance (in radians) against the balancing weight (in inch-ounces). Left: polar representation with estimated regression curve (continuous line) and estimated regression curve under the no-effect hypothesis (dashed line). Right: planar representation with estimated regression curve for the whole sample and for each group (as indicated in the legend).

regression curves. The regression function for all the data was represented in black. The smoothing parameter was selected by cross-validation, using only the data belonging to each group.

The test of equality is first applied to the data, to ascertain if all the regression curves are the same. Again, normality was not rejected for the residuals, so the χ^2 calibration was employed. The obtained value of the statistic is 20.96, while the p-value is 0.0263, lower than the nominal level $\alpha = 0.05$. Thus, for this significance level, the hypothesis of equal regression curves is rejected. This result is obtained using the concentration parameter selected by cross-validation for all the data (2.85). For a better application of the test, a sequence of concentration values is used, obtaining, as shown in Figure 2.4 (top panel), that the equality assumption is not rejected for concentration values approximately larger than 5, although given the sample size, large smoothing parameters are quite unrealistic in practice. Then, it can be concluded that there is evidence for saying that the four regression curves are not equal for a significance level of 0.05.

Once the equality hypothesis is rejected, it could be checked if the regression curves are parallel. The parallelism test is applied with the smoothing parameter selected by cross-validation (2.85), and the obtained value for the test statistic is 5.44, while the p-value is 0.4695, much greater than $\alpha = 0.05$. Thus, there is no evidence for rejecting the null hypothesis of parallel regression curves. To avoid compromising results because of the selection of the smoothing parameter, the test is applied using a range of smoothing values. The trace of the test shows that the null hypothesis is not rejected for $\alpha = 0.05$ for any of the smooth-

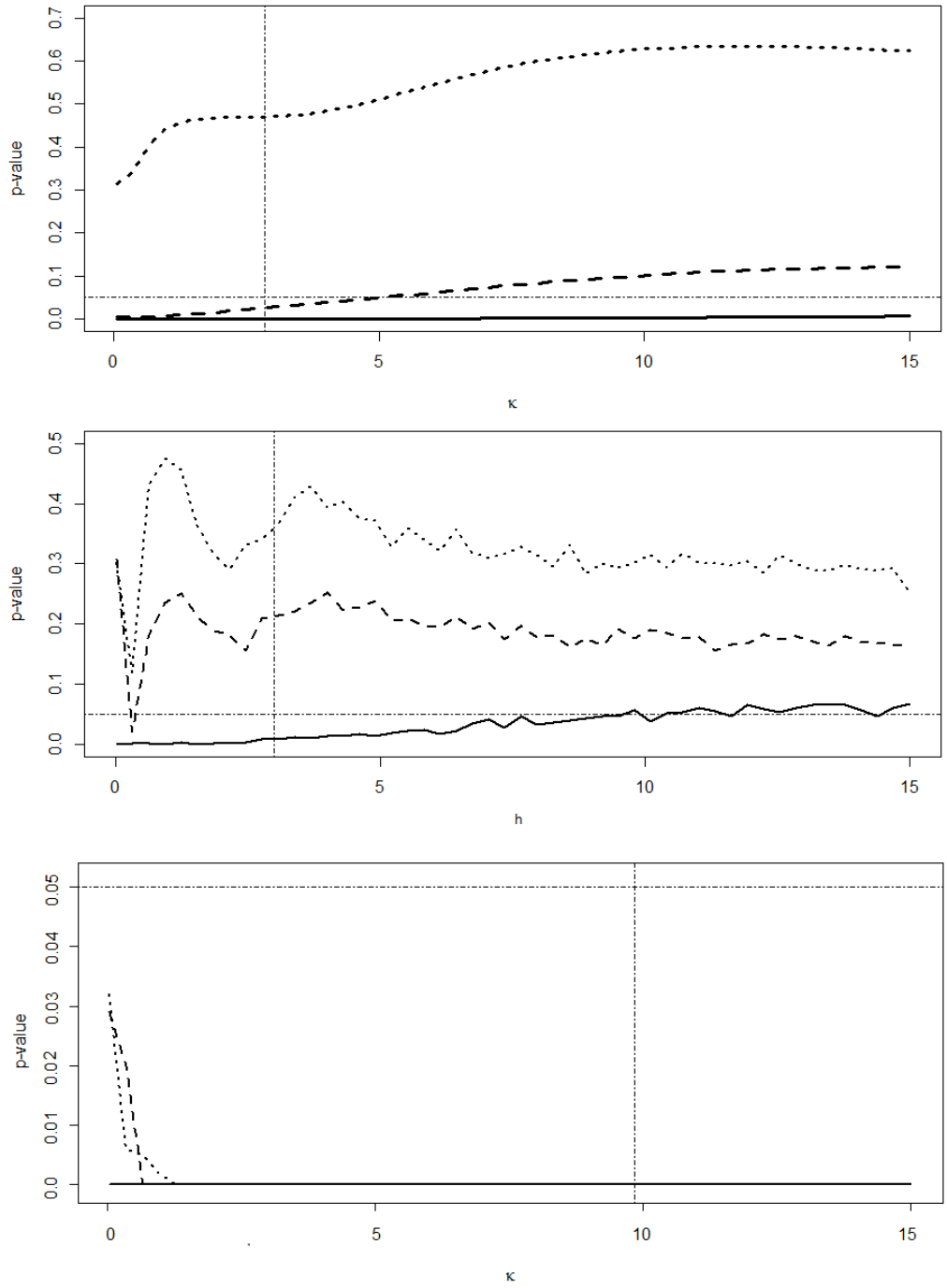


Figure 2.4: Traces of the no-effect tests (continuous line) equality tests (dashed line) and parallelism tests (dotted line) for the flywheels data (top), sand hoppers data (middle) and crickets data (bottom). Horizontal dashed-dotted line represents the significance level $\alpha = 0.05$ *and vertical dashed-dotted line indicates the smoothing parameter selected by cross-validation in each scenario.*

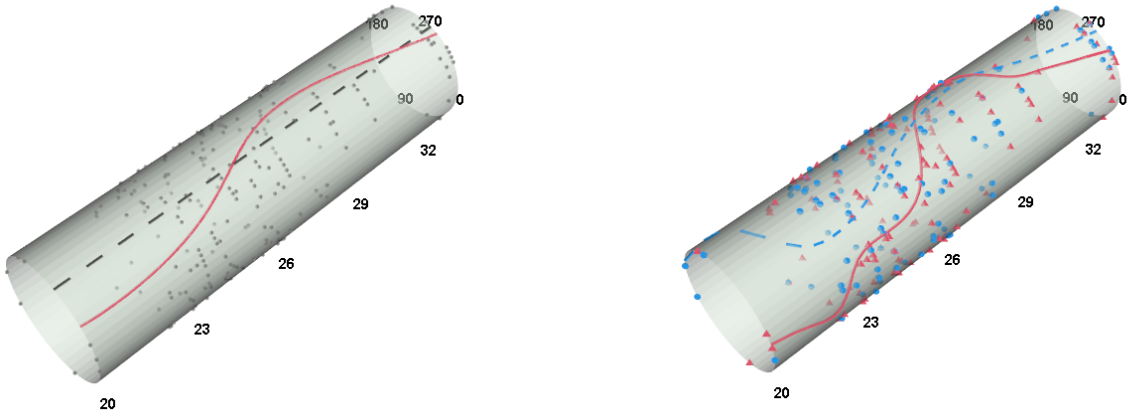


Figure 2.5: Representations on the cylinder of the sand hoppers data. Left: estimated regression curve (continuous line) and estimated regression curve under the no-effect hypothesis (dashed line) for the whole sample. Right: scatter plot and estimated regression curves for each group; screened group in red; unscreened group in blue.

ing parameters considered (κ lying between .05 and 15), as it can be seen in Figure 2.4 (top panel).

Sand hoppers data set

In what follows, the aim is to apply the nonparametric significance test proposed in Section 2.2.2 and the ANCOVA tests introduced in Section 2.3.2 to the sand hoppers data set described in Appendix A.2. A regression model in which the response variable is the direction of movement and the predictor variable is the temperature will be considered. For the ANCOVA models, the type of arena will be the categorical variable, which determines two groups: the unscreened and the screened sand hoppers. In this study, only the male animals and the observations which took place in October will be considered. The total number of observations considered for the analysis is 261, with 125 belonging to the unscreened group and 136 in the screened group.

The left panel in Figure 2.5 shows a representation on the cylinder of the escape direction against temperature, with the estimated regression curve obtained with the cross-validation method for selecting the bandwidth. The no-effect curve, *i.e.* a curve representing the global mean direction of the responses, is also represented. The first goal here is to ascertain if the temperature actually has an effect on the responses, for which the no-effect test for linear-circular regression is used. The test was applied using 1000 bootstrap replicates and over a sequence of smoothing parameters between 0.01 and 50. The p-value was smaller than $\alpha = 0.05$ for all the smoothing parameters $h < 9$, being the bandwidth value obtained by

cross-validation equal to 2.98. Therefore, there are evidences to state that the temperature affects the preferred direction of the sand hoppers.

Once it is known that the direction of movement is significantly influenced by the temperature, the question relies on whether the regression functions for the screened and the unscreened animals are the same. The right panel in Figure 2.5 shows a representation of the data distinguishing the screened (red) and the unscreened (blue) groups, with the estimated regression functions. The smoothing parameter was selected by cross-validation in each group.

The plot suggests that the behavior of the screened animals could be different from the behavior of the unscreened sand hoppers. This issue can be assessed by using the nonparametric test of equality for linear-circular regression. The test was applied to the data with the smoothing parameter selected by cross-validation (2.98) and using 1000 bootstrap replicates, obtaining a critical value of 0.234. Then, there are not evidences, for a significance value $\alpha = 0.05$, to conclude that the two regression curves are different.

As mentioned before, the conclusion is that there are no evidences to reject that both curves are equal. For illustrative purposes, it can also be seen that there are no evidences against the hypothesis of parallelism by applying the proposed test, in which a p-value of 0.357 is obtained when using the smoothing parameter selected by cross-validation and 1000 bootstrap replicates. As it was already stated, it is recommended to run the test over a sequence of smoothing parameters obtaining the trace of the tests, which are shown in the middle panel of Figure 2.4 for a sequence of 50 bandwidths ranging from 0.01 to 15. The corresponding p-values for the tests of equality and parallelism were higher than $\alpha = 0.05$ for all the parameters considered (except for one in the equality test).

Crickets data set

Lastly, the no-effect, equality and parallelism tests are applied to the cricket's data set described Appendix A.3. Note that one of the aims in Sato et al. (2019) was to determine the existence of differences in the regression curves for the running and jumping groups. However, the authors used a parametric model (similar to the one in (1.16)) to estimate the regression curves, and the differences between the curves were analyzed just graphically.

In order to illustrate the methods described in this chapter, a regression model where the response is the escape direction (circular variable) and the predictor is the stimulus direction (circular variable) will be considered first. The data are represented on the torus in the left panel of Figure 2.6, where the estimated regression function when using the kernel estimator presented in Section 2.1.2 was also represented (continuous line). The dashed curve represents the global circular mean of the responses, which corresponds to the model under the hypothesis of no-effect. By comparing the curves graphically, it seems that there is an effect of the stimulus direction in the cricket's escape responses. To ascertain that this is not an artifact of the sample noise, the no–effect test is applied to the cricket's data set. A sequence of concentration parameters ranging between 0.05 and 15 and 1000 bootstrap replications are

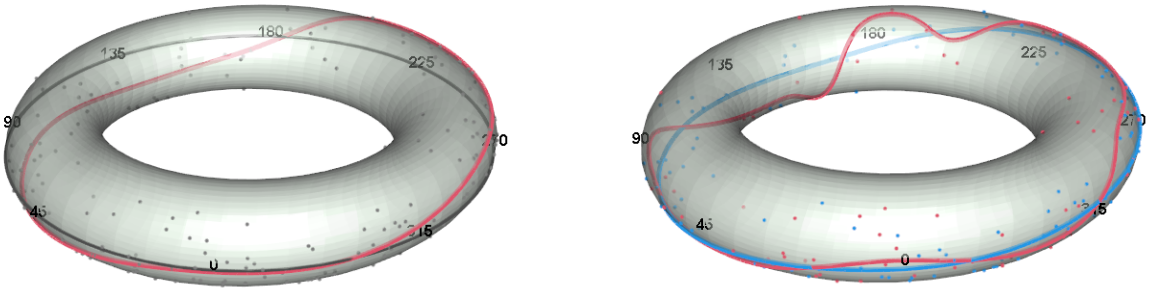


Figure 2.6: Representations on the torus of the crickets data. Left: estimated regression curve (continuous line) and estimated regression curve under the no-effect hypothesis (dashed line) for the whole sample. Right: scatter plot and estimated regression curves for each group; running responses in red; jumping responses in blue.

selected. The p-value was considerably smaller than $\alpha = 0.05$ for all the considered smoothing parameters (see continuous line in the bottom panel of Figure 2.4), concluding that there are evidences to state that the angle of stimulus actually influences the animal's direction of escape.

Next, the effect of the categorical variable determining the type of escaping response (running or jumping) is considered. In the right panel of Figure 2.6, the data corresponding to the running group is represented in red color, while the jumping responses are depicted in blue. The kernel estimator for a circular predictor and a circular response was employed to estimate the relationship between the angle of stimulus and the direction of escape in each group, as shown in Figure 2.6. Although it seems that the regression curve is similar for both groups when the direction of stimulus is around 0 or 240 degrees, the graphical representation of the curves suggests that the two regression functions might be different in a global sense. To statistically ascertain this issue, the equality test is employed, using 1000 bootstrap replications of the data. When the concentration parameter is selected by cross-validation (9.84), the obtained critical value is much lower than the nominal level $\alpha = 0.05$. When computing the trace of the test (see bottom panel of Figure 2.4), it is observed that this happens for all the considered values of the concentration parameter, concluding that there are evidences to reject the hypothesis of equality of the curves for this value of α. Lastly, the trace of the parallelism test is fairly similar to the trace of the equality test, therefore it is also concluded that, for a nominal level of $\alpha = 0.05$, there are evidences to state that the curves are not parallel either.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 3

A general framework for circular regression

In the previous chapter, inference tools for nonparametric regression with circular variables were introduced, understanding the regression function as the conditional mean of the response variable, given the covariate. However, a more general setting can be considered by estimating a characteristic of the conditional density function, once the latter has been selected. The estimation of these general regression curves can be carried out through the maximization of the local kernel weighted log-likelihood, an idea which was introduced in the real-valued context by Staniswalis (1989) and Fan et al. (1995), and also studied by Fan et al. (1998).

This chapter presents a general methodology to nonparametrically estimate curves of interest, in the context of regression involving a circular predictor, by maximizing the local circular kernel weighted log-likelihood. This approach allows to estimate regression curves with a circular covariate and a general response, which can be discrete or continuous. This method englobes as particular cases the proposal of Di Marzio et al. (2009), exposed in Section 2.1, when using a normal likelihood and the kernel logistic method of Di Marzio et al. (2018) if the likelihood is set to a Bernoulli distribution. Additionally, nonparametric Poisson or gamma regression can be also performed via circular local likelihood.

The general local maximum likelihood estimation procedure, for circular covariates, is presented in Section 3.1. Section 3.2 shows how to compute both the bias and variance of the estimators. The selection of the smoothing parameter is discussed in Section 3.3, while the empirical performance of the estimators is studied via simulations in Section 3.4 for several models, including continuous and discrete responses. Illustrations with real data sets are shown in Section 3.5. Finally, extensions of the method to include more covariates with different nature are discussed in Section 3.6. The contents of this chapter are collected in Alonso-Pena et al. (2022b).

Contents

3.1 Local likelihood estimation for circular regression 48

3.2 Bias and variance of the estimator 51

3.2.1 Bias of the estimator 51

3.2.2 Variance of the estimator 52

3.3 Selection of the smoothing parameter 53

3.3.1 Selection of the pilot concentration: least-squares case 55

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3.3.2 Selection of the pilot concentration: general case 56
3.4 Simulation study 58
3.5 Real data illustration 66
3.6 Extensions to partially linear and additive models 71

3.1 Local likelihood estimation for circular regression

Let Θ be a continuous circular variable defined on $\mathbb{T} = [0, 2\pi)$ and Y a random variable which can be either a discrete or a real-valued continuous variable. Given a random bivariate sample $\{(\Theta_i, Y_i)\}_{i=1}^n$, the goal of this section is estimating a generic unknown function $g(\cdot)$, which may represent, for example, the conditional mean regression function of Y given $\Theta = \theta$ or a transformed conditional mean function.

For a given $\theta_0 \in \mathbb{T}$, the function of interest, $g(\theta_0)$, is approximated by employing the same Taylor-like expansion than in (2.2). For data points Θ_i in a neighborhood of θ_0,

$$g(\Theta_i) \approx g(\theta_0) + g'(\theta_0)\sin(\Theta_i - \theta_0) + ... + \frac{g^{(p)}(\theta_0)}{p!}\sin^p(\Theta_i - \theta_0), \tag{3.1}$$

where $g^{(p)}(\cdot)$ denotes the pth derivative of $g(\cdot)$. This approximation can be expressed in vector notation as

$$g(\Theta_i) \approx \mathbf{\Theta}_i^T \boldsymbol{\beta}, \tag{3.2}$$

where $\mathbf{\Theta}_i = [1, \sin(\Theta_i - \theta_0), \ldots, \sin^p(\Theta_i - \theta_0)]^T$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$, with $\beta_\nu = g^{(\nu)}(\theta_0)/\nu!$, for $\nu = 0, 1, ..., p$. For each observation (Θ_i, Y_i), where Θ_i belongs to a neighborhood of θ_0, the contribution of this observation to the log-likelihood will be $\ell[g(\Theta_i), Y_i]$ weighted by $K_\kappa(\Theta_i - \theta)$, where $K_\kappa(\cdot)$ is a circular kernel function with concentration parameter κ (which acts as the smoothing parameter), increasing as $n \to \infty$. In this chapter, it will be assumed that $K_\kappa(\cdot)$ belongs to a specific class of kernels satisfying the condition

$$K_\kappa(\cdot) = c_\kappa K\{\kappa[1 - \cos(\cdot)]\}, \tag{3.3}$$

where $K : [0, \infty) \to \mathbb{R}$ with

$$\int_0^\infty r^{\frac{j-1}{2}} K^l(r)dr < \infty \quad j \in \mathbb{N} \text{ and } l = 1, 2, 4. \tag{3.4}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The factor c_κ is a normalization constant given by

$$c_\kappa^{-1} = \int_0^{2\pi} K[\kappa(1 - \cos\theta)]d\theta = \kappa^{-1/2}\lambda_\kappa(K), \tag{3.5}$$

where $\lambda_\kappa(K) = 2\int_0^{2\kappa} r^{-\frac{1}{2}}\left(2-\frac{r}{\kappa}\right)^{-\frac{1}{2}} K(r)dr$. Recall that κ is a sequence depending on n and, for this class of circular kernels, it holds that $\kappa \to \infty$ as $n \to \infty$. Thus, for a large κ it holds $c_\kappa^{-1} \sim \kappa^{-1/2}\lambda(K)$, with $\lambda(K) = 2^{\frac{1}{2}}\int_0^{\infty} r^{-\frac{1}{2}}K(r)dr$. Condition (3.3) is usually assumed in the directional setting (Hall et al., 1987; Bai et al., 1988; García-Portugués et al., 2013). The von Mises kernel is an example of a kernel satisfying (3.3). In this case, the normalization constant is given by $c_\kappa = \exp(\kappa)/[2\pi I_0(\kappa)]$ and $K(r) = \exp(-r)$.

Based on (3.2), the quantity $\ell[g(\Theta_i), Y_i]$ can be approximated by $\ell(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}, Y_i)$. Consequently, the local circular kernel weighted log-likelihood can be defined as

$$\mathcal{L}_p(\boldsymbol{\beta}; \kappa, \theta_0) = \sum_{i=1}^{n} \ell(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}, Y_i)K_\kappa(\Theta_i - \theta_0), \tag{3.6}$$

where the subscript p denotes the degree of the trigonometric polynomial used for the approximation in (3.1). By maximizing the local log-likelihood in (3.6) with respect to $\boldsymbol{\beta}$, the estimations of the local parameters, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, ..., \hat{\beta}_p)^T$ are obtained. Then, the estimators of the target function $g(\cdot)$ and its derivatives, at the point θ_0, are given by

$$\hat{g}^{(\nu)}(\theta_0) = \nu!\hat{\beta}_\nu, \quad \nu = 0, ..., p,$$

where ν represents the order of the derivative. In practice, an adequate choice of the order of the sine-polynomial to estimate $g(\cdot)$ is $p = 1$, leading to a local linear type estimator. Note that the maximization of (3.6) may not have an explicit solution in some cases, in which numerical methods must be employed in order to obtain the estimators.

The methodology proposed in this section is a general approach which allows to obtain local sine-polynomial estimators for a broad class of regression contexts involving a circular covariate. Apart from including the two particular cases already studied in the literature (normal and Bernoulli), it allows to estimate the transformed regression function in a wider collection of scenarios, for example, when having a Poisson or gamma likelihood.

Some particular cases

Some interesting particular cases of the circular local likelihood estimator are presented hereafter. First, consider the scenario where $g(\cdot)$ is the regression function $m(\cdot)$ in model (2.1). If the errors are normally distributed, the conditional distribution of the response is $[Y|\Theta = \theta_0] \sim N[g(\theta_0), \sigma^2(\theta_0)]$. Consequently, the local circular kernel weighted log-likelihood for a fixed $\theta_0 \in \mathbb{T}$, namely $\mathcal{L}_p(\boldsymbol{\beta}; \kappa, \theta_0)$, is given by

$$-\log\left[\sigma(\theta)\sqrt{2\pi}\right]\sum_{i=1}^{n} K_\kappa(\Theta_i - \theta_0) - \frac{1}{2\sigma^2(\theta)}\sum_{i=1}^{n}\left[Y_i - \sum_{j=0}^{p}\beta_j \sin^j(\Theta_i - \theta_0)\right]^2 K_\kappa(\Theta_i - \theta_0).$$

Maximizing the previous expression with respect to $\boldsymbol{\beta}$ is equivalent to minimizing

$$\sum_{i=1}^{n}\left[Y_i - \sum_{j=0}^{p}\beta_j \sin^j(\Theta_i - \theta_0)\right]^2 K_\kappa(\Theta_i - \theta_0),$$

which corresponds to the local polynomial least-squares minimization problem in (2.3), studied by Di Marzio et al. (2009) and Oliveira et al. (2013).

Now, consider a binary variable of interest, Y, with a Bernoulli conditional distribution, *i.e.*, $[Y|\Theta = \theta_0] \sim \text{Bernoulli}[p(\theta_0)]$. Consider the target function

$$g(\theta_0) = \text{logit}[p(\theta_0)] = \log\left[\frac{p(\theta_0)}{1 - p(\theta_0)}\right]. \tag{3.7}$$

Then, the contribution to the local log-likelihood is

$$\ell[g(\theta_0), y] = y g(\theta_0) - \log\left\{1 + \exp[g(\theta_0)]\right\}$$

and, thus, the local circular kernel log-likelihood is given by

$$\mathcal{L}_p(\boldsymbol{\beta}; \kappa, \theta_0) = \sum_{i=1}^{n}\left\{Y_i \boldsymbol{\Theta}_i^T \boldsymbol{\beta} - \log\left[1 + \exp(\boldsymbol{\Theta}_i^T \boldsymbol{\beta})\right]\right\} K_\kappa(\Theta_i - \theta_0),$$

leading to a nonparametric logistic regression model with a circular covariate, studied by Di Marzio et al. (2018) in the context of nonparametric classification.

Lastly, another interesting case arises when Y is a count variable, following a Poisson distribution where the mean parameter depends on the value of Θ. The considered target function $g(\cdot)$ will be the logarithm of the mean function, $g(\theta) = \log[\mathbb{E}(Y|\Theta = \theta_0)]$. Therefore, the conditional mean function will be given by $\mathbb{E}[Y|\Theta = \theta_0] = \exp[g(\theta_0)]$. Then, the local log-likelihood is

$$\mathcal{L}_p(\boldsymbol{\beta}; \kappa, \theta_0) = \sum_{i=1}^{n}\left[Y_i \boldsymbol{\Theta}_i^T \boldsymbol{\beta} - \exp\left(\boldsymbol{\Theta}_i^T \boldsymbol{\beta}\right) - \log(Y_i!)\right] K_\kappa(\Theta_i - \theta_0).$$

USC

Since the last term does not depend on $\boldsymbol{\beta}$, the maximization of the previous expression is equivalent to the maximization of

$$\sum_{i=1}^{n}\left\{Y_i \sum_{j=0}^{p}\beta_j \sin^j(\Theta_i - \theta_0) - \exp\left[\sum_{j=0}^{p}\beta_j \sin^j(\Theta_i - \theta_0)\right]\right\} K_\kappa(\Theta_i - \theta_0).$$

3.2 Bias and variance of the estimator

For many inferential tasks it is important to compute the bias and variance of the estimator $\hat{\boldsymbol{\beta}}$, obtained after maximizing (3.6). Estimating these quantities will also be useful in order to select a smoothing parameter. In this section, the approach of Fan et al. (1998) is followed to give approximations of the bias and variance of the estimators presented in Section 3.1.

3.2.1 Bias of the estimator

The bias of $\hat{\boldsymbol{\beta}}$ comes from the approximation of the target function by the sine-polynomial in (3.1). Consequently, the bias can be approximated by computing the difference of two maximum local likelihood fits with different accuracies, *i.e.*, with distinct values of the degree of the sine-polynomial. Denote the error approximation at Θ_i, resulting from (3.1), by

$$\epsilon(\Theta_i) = g(\Theta_i) - \sum_{\nu=0}^{p} \frac{g^{(\nu)}(\theta_0)}{\nu!} \sin^{\nu}(\Theta_i - \theta_0).$$

Assume the existence of the $(p+a+1)$th derivative of the function $g(\cdot)$ at the point θ_0 for some $a \in \mathbb{N}$. Then, the error term can be approximated by a further sine-polynomial expansion,

$$\epsilon(\Theta_i) \approx \beta_{p+1} \sin^{p+1}(\Theta_i - \theta_0) + ... + \beta_{p+a} \sin^{p+a}(\Theta_i - \theta_0) = \epsilon_i. \tag{3.8}$$

The choice of a will affect how well the bias is estimated, but a large value of a will lead to a big computational effort. A good choice in practice, which gives a satisfactory performance when estimating the bias while reducing the computational time, is $a = 2$. Suppose, for now, that the quantities ϵ_i, $i = 1, ..., n$, are known, which is not true in practice. With this assumption, the local log-likelihood could be approximated in a more precise way as

$$\mathcal{L}_p^*(\boldsymbol{\beta}; \kappa, \theta_0) = \sum_{i=1}^{n} \ell(\boldsymbol{\Theta}_i^T \boldsymbol{\beta} + \epsilon_i, Y_i) K_\kappa(\Theta_i - \theta_0). \tag{3.9}$$

Denote by $\hat{\boldsymbol{\beta}}^*$ the maximizer of $\mathcal{L}_p^*(\boldsymbol{\beta}; \kappa, \theta_0)$. The bias of $\hat{\boldsymbol{\beta}}$ can be estimated by $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^*$. However, it would not be necessary to compute $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^*$, as it will be seen below. Let

$$\mathcal{L}_p^{*'}(\boldsymbol{\beta}; \kappa, \theta_0) = \frac{\partial}{\partial \boldsymbol{\beta}} \mathcal{L}_p^*(\boldsymbol{\beta}; \kappa, \theta_0), \quad \text{and} \quad \mathcal{L}_p^{*''}(\boldsymbol{\beta}; \kappa, \theta_0) = \frac{\partial^2}{\partial \boldsymbol{\beta}^2} \mathcal{L}_p^*(\boldsymbol{\beta}; \kappa, \theta_0)$$

be, respectively, the gradient vector and the Hessian matrix of $\mathcal{L}_p^*(\boldsymbol{\beta}; \kappa, \theta_0)$. Since $\hat{\boldsymbol{\beta}}^*$ is the maximizer of $\mathcal{L}_p^*(\boldsymbol{\beta}; \kappa, \theta_0)$, it yields $\mathcal{L}_p^{*'}(\hat{\boldsymbol{\beta}}^*; \kappa, \theta_0) = 0$ and, hence, by using a Taylor expansion, it holds that

$$0 = \mathcal{L}_p^{*'}(\hat{\boldsymbol{\beta}}^*; \kappa, \theta_0) \approx \mathcal{L}_p^{*'}(\hat{\boldsymbol{\beta}}; \kappa, \theta_0) + \mathcal{L}_p^{*''}(\hat{\boldsymbol{\beta}}; \kappa, \theta_0)(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}}).$$

Consequently, the approximated bias vector of $\hat{\boldsymbol{\beta}} = (\beta_0, \ldots, \beta_p)^T$, namely $\hat{\boldsymbol{b}}_p(\theta_0; \kappa)$, is obtained as

$$\hat{\boldsymbol{b}}_p(\theta_0; \kappa) = \left[\mathcal{L}_p^{*\prime\prime}(\hat{\boldsymbol{\beta}}; \kappa, \theta_0)\right]^{-1} \mathcal{L}_p^{*\prime}(\hat{\boldsymbol{\beta}}; \kappa, \theta_0). \tag{3.10}$$

Note that the bias in (3.10) cannot be computed in practice, since it depends on the unknown quantities $\epsilon_1, \ldots, \epsilon_n$. The process to obtain a data-driven approximation of the bias is as follows. First, a pilot concentration parameter, namely κ^*, is selected. This pilot concentration is used to fit a sine-polynomial of degree $(p + a)$, obtaining estimates $\hat{\boldsymbol{\beta}}^{(p+a)} = (\hat{\beta}_0, \ldots, \hat{\beta}_{p+a})^T$. Second, these quantities are substituted into (3.8) and, thus, the estimators of $\epsilon_1, \ldots, \epsilon_n$, denoted by $\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n$, are obtained. Plugging $\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n$ into (3.9) yields the estimated bias vector from (3.10), namely $\hat{\boldsymbol{b}}_p^e(\theta_0; \kappa)$. Then, recalling that $\beta_\nu = g^{(\nu)}(\theta)/\nu!$, the estimated bias of $\hat{g}^{(\nu)}(\theta_0)$ is given by

$$\hat{B}_{p,\nu}(\theta_0; \kappa) = \nu! \boldsymbol{e}_{\nu+1}^T \hat{\boldsymbol{b}}_p^e(\theta_0; \kappa), \tag{3.11}$$

where $\boldsymbol{e}_{\nu+1}$ denotes the vector with all entries equal to zero except the one in the $(\nu + 1)$th position. The choice of the pilot concentration, κ^*, will be discussed in Section 3.3.

3.2.2 Variance of the estimator

Now, the computation of the variance of the estimated vector of local parameters will be described. Since estimator $\hat{\boldsymbol{\beta}}$ is the maximizer of the circular local likelihood, it holds $\mathcal{L}_p'(\hat{\boldsymbol{\beta}}; \kappa, \theta_0) = 0$, and a Taylor expansion gives

$$0 = \mathcal{L}_p'(\hat{\boldsymbol{\beta}}; \kappa, \theta_0) \approx \mathcal{L}_p'(\boldsymbol{\beta}; \kappa, \theta_0) + \mathcal{L}_p''(\boldsymbol{\beta}; \kappa, \theta_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Therefore,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \approx -\left[\mathcal{L}_p''(\boldsymbol{\beta}; \kappa, \theta_0)\right]^{-1} \mathcal{L}_p'(\boldsymbol{\beta}; \kappa, \theta_0).$$

Now, notice that

$$\begin{aligned}
&\mathrm{Var}(\hat{\boldsymbol{\beta}}|\Theta_1, \ldots, \Theta_n) \\
&= \mathrm{Var}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\Theta_1, \ldots, \Theta_n) \approx \mathrm{Var}\left\{-\left[\mathcal{L}_p''(\boldsymbol{\beta}; \kappa, \theta_0)\right]^{-1} \mathcal{L}_p'(\boldsymbol{\beta}; \kappa, \theta_0)|\Theta_1, \ldots, \Theta_n\right\} \\
&\approx \left[\mathcal{L}_p''(\boldsymbol{\beta}; \kappa, \theta_0)\right]^{-1} \mathrm{Var}\left[\mathcal{L}_p'(\boldsymbol{\beta}; \kappa, \theta_0)|\Theta_1, \ldots, \Theta_n\right] \left[\mathcal{L}_p''(\boldsymbol{\beta}; \kappa, \theta_0)\right]^{-1}.
\end{aligned}$$

The matrix $\mathcal{L}_p''(\boldsymbol{\beta}; \kappa, \theta_0)$ can be estimated by $\mathcal{L}_p''(\hat{\boldsymbol{\beta}}; \kappa, \theta_0)$, but $\mathrm{Var}\left[\mathcal{L}_p'(\boldsymbol{\beta}; \kappa, \theta_0)|\Theta_1, \ldots, \Theta_n\right]$ is still unknown and it is necessary to estimate it. From the definition of the circular local log-likelihood in (3.6), it holds

$$\mathcal{L}_p'(\boldsymbol{\beta}; \kappa, \theta_0) = \sum_{i=1}^{n} \ell'(\boldsymbol{\Theta}_i^T \boldsymbol{\beta}, Y_i) \boldsymbol{\Theta}_i K_\kappa(\Theta_i - \theta_0)$$

and, consequently,

$$\text{Var}\left[\mathcal{L}_p'(\boldsymbol{\beta};\kappa,\theta_0)|\Theta_1,\ldots,\Theta_n\right] = \sum_{i=1}^{n} \text{Var}\left[\ell'(\boldsymbol{\Theta}_i^T\boldsymbol{\beta},Y_i)|\Theta_1,\ldots,\Theta_n\right]\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T K_\kappa^2(\Theta_i-\theta_0). \tag{3.12}$$

Now, because of (3.2), the expression in (3.12) can be approximated by

$$\text{Var}\left[\ell'(g(\theta_0),Y)|\Theta=\theta_0\right]\sum_{i=1}^{n}\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T K_\kappa^2(\Theta_i-\theta_0) = \text{Var}\left\{\ell'[g(\theta_0),Y]|\Theta=\theta_0\right\}\boldsymbol{\Gamma}_n,$$

where $\boldsymbol{\Gamma}_n$ is a $(p+1)\times(p+1)$ matrix with (i,j)th element given by $\gamma_{n,i+j-2}$ and

$$\gamma_{n,j} = \sum_{i=1}^{n}\sin^j(\Theta_i-\theta_0)K_\kappa^2(\Theta_i-\theta_0).$$

Then, it holds

$$\begin{aligned}&\text{Var}(\hat{\boldsymbol{\beta}}|\Theta_1,\ldots,\Theta_n)\\ &\approx \text{Var}\left\{\ell'[g(\theta_0),Y]|\Theta=\theta_0\right\}\left[\mathcal{L}_p''(\hat{\boldsymbol{\beta}};\kappa,\theta_0)\right]^{-1}\boldsymbol{\Gamma}_n\left[\mathcal{L}_p''(\hat{\boldsymbol{\beta}};\kappa,\theta_0)\right]^{-1} = \Xi_p(\theta_0;\kappa).\end{aligned}$$

For the estimation of $\text{Var}\left\{\ell'[g(\theta_0),Y]|\Theta=\theta_0\right\}$, two cases are distinguished:

A. $\text{Var}\left\{\ell'[g(\theta_0),Y]|\Theta=\theta_0\right\} = v[g(\theta_0)]$ for some known function $v(\cdot)$, as it happens for the Bernoulli or Poisson likelihoods. In this case it is estimated as $v[\hat{g}(\theta_0)]$.

B. When the form in A is not available, a pilot estimator $\hat{\boldsymbol{\beta}}^{(p+a)}$ obtained by fitting a local polynomial of degree $p+a$, with a pilot concentration κ^*, is used as in the bias calculations. Then, $\text{Var}\left\{\ell'[g(\theta_0),Y]|\Theta=\theta_0\right\}$ is estimated by

$$\frac{\sum_{i=1}^{n}[\ell'(\tilde{\boldsymbol{\Theta}}_i^T\hat{\boldsymbol{\beta}}^{(p+a)},Y_i)]^2K_{\kappa^*}(\Theta_i-\theta_0)}{\sum_{i=1}^{n}K_{\kappa^*}(\Theta_i-\theta_0)},$$

with $\tilde{\boldsymbol{\Theta}}_i^T = [1,\sin(\Theta_i-\theta_0),\ldots,\sin^{p+a}(\Theta_i-\theta_0)]$.

The expression of the variance of $\hat{g}^{(\nu)}(\theta_0)$ constructed with a sine-polynomial of degree p is given by

$$\hat{V}_{p,\nu}(\theta_0;\kappa) = \nu!^2\boldsymbol{e}_{\nu+1}^T\Xi_p(\theta_0;\kappa)\boldsymbol{e}_{\nu+1}. \tag{3.13}$$

3.3 Selection of the smoothing parameter

As in least-squares kernel regression for circular predictors (see Section 2.1.1), the selection of the smoothing parameter κ will be crucial, since a large value of κ will lead to a small bias but a large variance, and these roles reverse if the value of κ is small. In order to automatically

select a smoothing parameter for the estimation of $g^{(\nu)}(\theta_0)$, it is possible to minimize an estimation of the integrated version of the Mean Squared Error (MSE):

$$\hat{\kappa}_{p,\nu} = \arg\min_{\kappa>0} \int_0^{2\pi} \widehat{\text{MSE}}_{p,\nu}(\alpha;\kappa)d\alpha,$$

where $\widehat{\text{MSE}}_{p,\nu}(\theta_0;\kappa)$ denotes an estimator of the MSE of $g^{(\nu)}(\theta_0)$ constructed with the concentration parameter κ. This quantity can be computed by approximating the bias and variance of the estimator as described in Sections 3.2.1 and 3.2.2, respectively, obtaining the bias and variance estimates given by (3.11) and (3.13). Then, the estimated MSE is given by

$$\widehat{\text{MSE}}_{p,\nu}(\theta_0;\kappa) = \hat{B}^2_{p,\nu}(\theta_0;\kappa) + \hat{V}_{p,\nu}(\theta_0;\kappa).$$

Note that in order to estimate the MSE, it is necessary to first select a pilot smoothing parameter, κ^*, and fit locally a sine-polynomial of degree $p+a$. Hence, this smoothing parameter selection method will be referred to as the refined rule, since first it is necessary to select a preliminary smoothing parameter.

As mentioned in Section 2.1, in the particular case of the normal likelihood (least-squares case), Di Marzio et al. (2009) derived an expression for the optimal smoothing parameter minimizing the asymptotic MSE of the estimator in the case $p = 1$ and $\nu = 0$, and specifying the von Mises density as the kernel. The expression of the optimal concentration parameter is given in equation (2.6). In the directional setting, where it is assumed that the predictor lies on a hypersphere of arbitrary dimension, García-Portugués (2014) derived an optimal expression for the concentration minimizing the MSE (assuming that the kernel satisfies (3.3)), which in the particular case of data on the circumference and a von Mises kernel is equivalent to the parameter in equation (2.6).

Note that, as it was also mentioned in Section 2.1, the optimal expression of the concentration parameter cannot be used in practice, and the only proposals available are the rule of thumb given in García-Portugués (2014) (in the more general directional-linear setting) and the cross-validation method in equation (2.7). Thus, the previously described method for selecting the smoothing parameter by minimizing the estimated bias and variance constitutes a novel approach, not only in the general local likelihood case, but also in the context of kernel least-squares regression for a circular predictor.

In what follows, the selection of the pilot concentration parameter will be discussed. Although the role of the smoothing parameter is reversed when employing circular kernels, a similar approach to that of Fan et al. (1998) could be used to select the pilot concentration. The main idea is to come up with a criterion which, when minimized, leads to an approximated optimal smoothing parameter. In Section 3.3.1, the problem of obtaining a pilot concentration in the least-squares case is studied. The general case is discussed in Section 3.3.2.

3.3.1 Selection of the pilot concentration: least-squares case

In this section, the least-squares scenario exposed in Section 2.1.1 is considered. Let the relationship between the variables Θ and Y be modeled as in (2.1), with $m(\cdot) = g(\cdot)$. Consider the matrices $\boldsymbol{\Theta}$ and $\boldsymbol{W}$ introduced in Section 2.1.1. Following Fan and Gijbels (1995), the Circular Residual Squares Criterion (CRSC) is defined as

$$\text{CRSC}(\theta_0;\kappa) = \hat{\sigma}^2(\theta_0)\left(1+\frac{p+1}{N}\right),$$

where

$$\hat{\sigma}^2(\theta_0) = \frac{\sum_{i=1}^n (Y_i-\hat{Y}_i)^2 K_\kappa(\Theta_i-\theta_0)}{\text{tr}(\boldsymbol{W}) - \text{tr}\left[(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^T\boldsymbol{W}^2\boldsymbol{\Theta}\right]}, \tag{3.14}$$

and N^{-1} being the first diagonal element of the matrix $(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^T\boldsymbol{W}^2\boldsymbol{\Theta}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1} = \boldsymbol{S}_n^{-1}\boldsymbol{\Gamma}_n\boldsymbol{S}_n^{-1}$. The quantities $\hat{Y}_i$, with $i=1,\ldots,n$, denote the fitted values obtained after fitting a pth order sine-polynomial locally. When κ is very small, the bias of the estimator will be large and so will be $\hat{\sigma}^2(\theta_0)$, obtaining a large $\text{CRSC}(\theta_0;\kappa)$. On the contrary, if κ is too large, the variance will be also large and, hence, N^{-1} will be big as well, resulting in a large $\text{CRSC}(\theta_0;\kappa)$.

Proposition 3.1 gives the expectation of the CRSC quantity. The following notation will be used:

$$b_j^*(K) = \begin{cases} 0 & \text{if } j \text{ is odd,} \\ \dfrac{\int_0^\infty r^{\frac{j-1}{2}}K(r)dr}{\int_0^\infty r^{-\frac{1}{2}}K(r)dr} & \text{if } j \text{ is even,} \end{cases} \quad d_j^*(K) = \begin{cases} 0 & \text{if } j \text{ is odd} \\ \dfrac{\int_0^\infty r^{\frac{j-1}{2}}K^2(r)dr}{\left(\int_0^\infty r^{-\frac{1}{2}}K(r)dr\right)^2} & \text{if } j \text{ is even.} \end{cases}$$

Further, let $\boldsymbol{B}$ and $\boldsymbol{D}$ be the $(p+1)\times(p+1)$ matrices having, respectively, the (i,j)th element given by $b_{i+j-2}^*(K)$ and $d_{i+j-2}^*(K)$. The vector $\left[b_{p+1}^*(K),\ldots,b_{2p+1}^*(K)\right]^T$ will be denoted by $\boldsymbol{c}_p$.

Proposition 3.1. *Assume that the circular kernel satisfies (3.3) and (3.4) and that $f(\cdot)>0$, the density function of Θ, is continuously differentiable. Additionally, assume that $Var(Y|\Theta=\theta)=\sigma^2(\theta)>0$ exists and is continuous at $\theta=\theta_0$. If $\kappa\to\infty$ and $n\kappa^{-\frac{1}{2}}\to\infty$, then the expression of $\mathbb{E}[CRSC(\theta_0;\kappa)|\Theta_1,\ldots,\Theta_n]$ is given by*

$$\sigma^2(\theta_0) + C_p\beta_{p+1}^2 2^{p+1}\kappa^{-(p+1)} + (p+1)\frac{\sigma^2(\theta_0)a_0\kappa^{\frac{1}{2}}}{2^{\frac{1}{2}}nf(\theta_0)} + o_P\left(\frac{1}{\kappa^{p+1}}+\frac{\kappa^{\frac{1}{2}}}{n}\right),$$

where $C_p = b_{2p+2}^*(K) - \boldsymbol{c}_p^T\boldsymbol{B}\boldsymbol{c}_p$, the value a_0 is the first diagonal element of the matrix $\boldsymbol{B}^{-1}\boldsymbol{D}\boldsymbol{B}^{-1}$ and $\beta_{p+1} = g^{(p+1)}(\theta_0)/(p+1)!$.

The proof can be found in Appendix C.2. Given the previous result, it follows that the parameter minimizing $\mathbb{E}[\text{CRSC}(\theta_0;\kappa)|\Theta_1,\ldots,\Theta_n]$ with respect to κ is approximately equal

to

$$\kappa_0(\theta_0) = \left[\frac{2^{\frac{2p+5}{2}} C_p \beta_{p+1}^2 n f(\theta_0)}{a_0 \sigma^2(\theta_0)} \right]^{\frac{2}{2p+3}}. \tag{3.15}$$

It can be seen that the value of $\kappa_0(\theta_0)$ shares some similarities with the optimal κ minimizing the asymptotic mean squared error of $\hat{g}^{(\nu)}(\theta_0)$ which, for odd $(p-\nu)$, is given by

$$\kappa_{\text{opt},p,\nu}(\theta_0) = \left[\frac{(p+1-\nu) n f(\theta_0) \beta_{p+1}^2 (\boldsymbol{e}_{\nu+1}^T \boldsymbol{B}^{-1} \boldsymbol{c}_p)^2 2^{\frac{2p+5}{2}}}{(1+2\nu) a_\nu \sigma^2(\theta_0)} \right]^{\frac{2}{2p+3}}, \tag{3.16}$$

where a_ν is the $(\nu+1)$th diagonal element of the matrix $\boldsymbol{B}^{-1}\boldsymbol{D}\boldsymbol{B}^{-1}$. The derivation of (3.16) is presented in Appendix C.2.

Remark 3.1. Equation (3.16) is a generalization of the optimal parameter obtained by Di Marzio et al. (2009), who studied the case $p = 1$ and $\nu = 0$. In this particular case, it also coincides with the optimal parameter obtained by García-Portugués (2014) in the directional setting.

Comparing equations (3.15) and (3.16), it is easy to see that

$$\kappa_{\text{opt},p,\nu}(\theta_0) = \xi_{p,\nu}(K) \kappa_0(\theta_0),$$

where $\xi_{p,\nu}(K)$ only depends on p, ν and the function $K(\cdot)$, and is given by

$$\xi_{p,\nu}(K) = \left[\frac{(p+1-\nu) a_0 \left(\boldsymbol{e}_{\nu+1}^T \boldsymbol{B}^{-1} \boldsymbol{c}_p \right)^2}{(1+2\nu) a_\nu C_p} \right]^{\frac{2}{2p+3}}.$$

Consequently, the approach for selecting a global pilot concentration parameter is as follows. First, the value of the concentration minimizing the integrated CRSC is obtained:

$$\hat{\kappa}_p = \arg\min_{\kappa > 0} \int_0^{2\pi} \text{CRSC}(\alpha; \kappa) d\alpha.$$

Afterwards, the pilot concentration $\hat{\kappa}_{p,\nu}^{\text{CRSC}}$ is selected as

$$\hat{\kappa}_{p,\nu}^{\text{CRSC}} = \xi_{p,\nu}(K) \hat{\kappa}_p.$$

3.3.2 Selection of the pilot concentration: general case

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

In the general circular local likelihood problem, two alternative routes can be followed to select the pilot concentration parameter. First, if the target function $g(\cdot)$ is a transformed mean function, *i.e.*, $g(\cdot) = \eta[\mu(\cdot)]$ with $\mu(\theta_0) = \mathbb{E}(Y|\Theta = \theta_0)$, the CRSC criterion can still be used, but substituting $\hat{Y}_i = \eta^{-1}(\boldsymbol{\Theta}_i^T \hat{\boldsymbol{\beta}})$ in (3.14).

Another possibility is to use an extended version of the CRSC, namely ECRSC, regarding the local likelihood problem as an iterative local least-squares problem, as in Fan et al. (1998). In what follows, some details about this extended criterion are given.

Consider the Fisher scoring method for updating the vector of estimated parameters $\hat{\boldsymbol{\beta}} \in \mathbb{R}^{p+1}$, in which, for a current value $\boldsymbol{\beta}_c$, it is updated as

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_c - \left\{\mathbb{E}[\mathcal{L}_p''(\boldsymbol{\beta}_c; \kappa, \theta_0)|\Theta_1, \ldots, \Theta_n]\right\}^{-1} \mathcal{L}_p'(\boldsymbol{\beta}_c; \kappa, \theta_0). \tag{3.17}$$

Now,

$$\begin{aligned}\mathbb{E}[\mathcal{L}_p''(\boldsymbol{\beta}_c; \kappa, \theta_0)|\Theta_1, \ldots, \Theta_n] &= \sum_{i=1}^{n} \mathbb{E}[\ell''(\boldsymbol{\Theta}_i^T \boldsymbol{\beta}_c, Y_i)]\boldsymbol{\Theta}_i \boldsymbol{\Theta}_i^T K_\kappa(\Theta_i - \theta_0) \\ &\approx \mathbb{E}\{\ell''[g(\theta_0), Y]|\Theta = \theta_0\} \sum_{i=1}^{n} \boldsymbol{\Theta}_i \boldsymbol{\Theta}_i^T K_\kappa(\Theta_i - \theta_0),\end{aligned}$$

where it was assumed that the expectation $\mathbb{E}\{\ell''[g(\cdot), Y]\}$ is continuous. Plugging this expression into equation (3.17) implies that the updating rule with the approximated expectation is

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \boldsymbol{\beta}_c - \left(\mathbb{E}\{\ell''[g(\theta_0), Y]|\Theta = \theta_0\} \sum_{i=1}^{n} \boldsymbol{\Theta}_i \boldsymbol{\Theta}_i^T K_\kappa(\Theta_i - \theta_0)\right)^{-1} \sum_{i=1}^{n} \boldsymbol{\Theta}_i \ell'(\boldsymbol{\Theta}_i^T \boldsymbol{\beta}_c, Y_i) K_\kappa(\Theta_i - \theta_0) \\ &= \left[\sum_{i=1}^{n} \boldsymbol{\Theta}_i \boldsymbol{\Theta}_i^T K_\kappa(\Theta_i - \theta_0)\right]^{-1} \sum_{i=1}^{n} Z_i \boldsymbol{\Theta}_i K_\kappa(\Theta_i - \theta_0),\end{aligned}$$

where

$$Z_i = \boldsymbol{\Theta}_i^T \boldsymbol{\beta}_c - \frac{\ell'(\boldsymbol{\Theta}_i^T \boldsymbol{\beta}_c, Y_i)}{\mathbb{E}\{\ell''[g(\theta_0), Y]|\Theta = \theta_0\}}$$

and the conditional expectation in the expression of Z_i can be computed with the value of $\boldsymbol{\beta}_c$. Consequently, it can be seen that at the end of the iteration process, the estimator $\hat{\boldsymbol{\beta}}$ is obtained by regressing Z_i over Θ_i using the local sine-polynomial of order p. Hence, the ECRSC criterion is defined as

$$\text{ECRSC}(\theta_0; \kappa) = \hat{\sigma}_*^2(\theta_0) \left(1 + \frac{p+1}{N}\right),$$

where $\hat{\sigma}_*^2(\theta_0)$ is computed as in equation (3.14) but using the variable Z_i. More details on the justification of this can be found in Fan et al. (1998). The concentration selector based on the ECRSC criterion can be obtained by first computing

$$\hat{\kappa}_p^* = \arg\min_{\kappa > 0} \int_0^{2\pi} \text{ECRSC}(\alpha; \kappa) d\alpha \tag{3.18}$$

and then evaluating

$$\hat{\kappa}_{p,\nu}^{\text{ECRSC}} = \xi_{p,\nu}(K)\hat{\kappa}_p^*.$$

In the case where $g(\cdot) = \eta[\mu(\cdot)]$, the ECRSC criterion will be approximately the same as the CRSC criterion, but including weights $\{\eta'[\mu(\alpha)]\}^{-2}$ in equation (3.18).

3.4 Simulation study

The aim of this section is to study the empirical performance of the circular local likelihood estimator presented in Section 3.1, as well as the behavior of the concentration selection methods introduced in Section 3.3. Different regression models with responses drawn from normal, Bernoulli, Poisson and gamma conditional distributions are explored. The considered models are described in Table 3.1 and represented in Figure 3.1.

For each model, 500 replications of the data are simulated, and the target function $g(\cdot)$ is estimated with the local sine-polynomial estimator ($p = 1$, $\nu = 0$). Sample sizes are fixed to $n = 100, 250, 500$ and the predictor variable, Θ, is drawn from a circular uniform distribution in all models. The concentration parameter was selected by the refined rule in Section 3.3, where the pilot estimator was constructed with a local sine-polynomial of degree 3 and the pilot concentration parameter was selected by the CRSC criterion (in the normal case) and the ECRSC criterion (in the other cases). For comparison purposes, the estimators obtained by selecting the smoothing parameter directly with the CRSC/ECRSC criterion and with a cross-validation method are also computed.

The quality of the estimators was obtained by estimating the Integrated Squared Error (ISE) as

$$\frac{\int_0^{2\pi}[\hat{g}_{(b)}(\theta) - g(\theta)]^2 d\theta}{\int_0^{2\pi} g(\theta)^2 d\theta}, \tag{3.19}$$

where $\hat{g}_{(b)}(\theta)$ represents the estimator of $g(\theta)$ for the bth replication of the data and the integrals are approximated numerically by Simpson's rule. Averages of these errors across the 500 replications are reported.

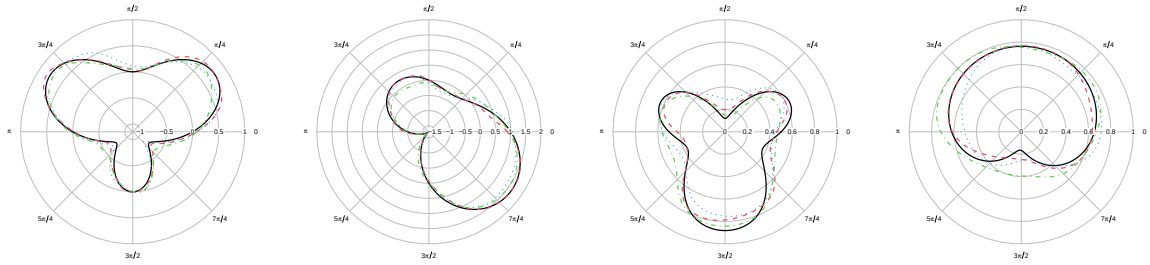
For the Bernoulli, Poisson and gamma likelihoods, the estimator involves an iterative solution, and for very large concentrations the estimator may not even exist. These situations are avoided by only considering values of the concentration parameter for which the estimators exist.

Normal likelihood. As already noted in Section 3.1, the normal likelihood case was studied by Di Marzio et al. (2009) and Oliveira et al. (2013), and a cross-validation method was employed in the latter to select the concentration parameter. The performance of the refined rule to select the smoothing parameter exposed in Section 3.3 is analyzed next.

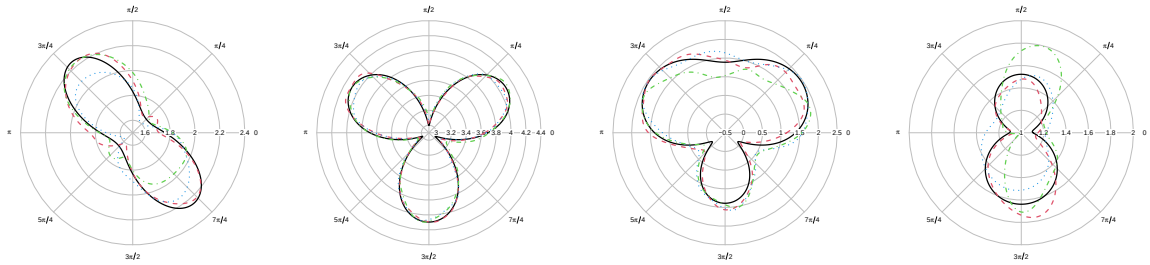
First, the focus is kept on the concentration parameters selected by each method. Figure 3.2 shows kernel density estimators of the selected smoothing parameters when using the

Table 3.1: Description of the simulated models.

Model	$(Y \mid \Theta = \theta)$	Model elements
Normal (N)	$\mathrm{N}\left(\mu(\theta), \sigma^2\right)$ $g(\theta) = \mu(\theta)$	N1: $g(\theta) = \sin(2\theta)\cos(\theta)$; N2: $g(\theta) = 1.75\cos(\theta - \pi)\sin(\theta) + \cos(\theta)$; $\sigma = 0.35$, $\sigma = 0.5$
Bernoulli (B)	Bernoulli $(p(\theta))$ $g(\theta) = \mathrm{logit}[p(\theta)]$	B1: $g(\theta) = 2\sin(\theta)\cos(2\theta)$ B2: $g(\theta) = \log\{1.6 + 1.5\sin(\theta) + 0.1\exp[\cos(\theta)]\}$
Poisson (P)	Poisson $(\mu(\theta))$ $g(\theta) = \log[\mu(\theta)]$	P1: $\mu(\theta) = 5 + \exp[1.5\sin(2\theta - 3)]$ P2: $\mu(\theta) = 40 + 20\sin(2\theta)$
Gamma (G)	Gamma $(\alpha, \beta(\theta))$ $\mu(\theta) = \alpha\beta^{-2}(\theta)$ $g(\theta) = \log[\mu(\theta)]$	G1: $\mu(\theta) = 4 + 4\sin(2\theta)\cos(\theta)$; G2: $\mu(\theta) = 5 + 2\cos[3 + 1.5\sin(\theta)]$; $\alpha = 2$ $\alpha = 1$



(a) N1 *(b) N2* *(c) B1* *(d) B2*



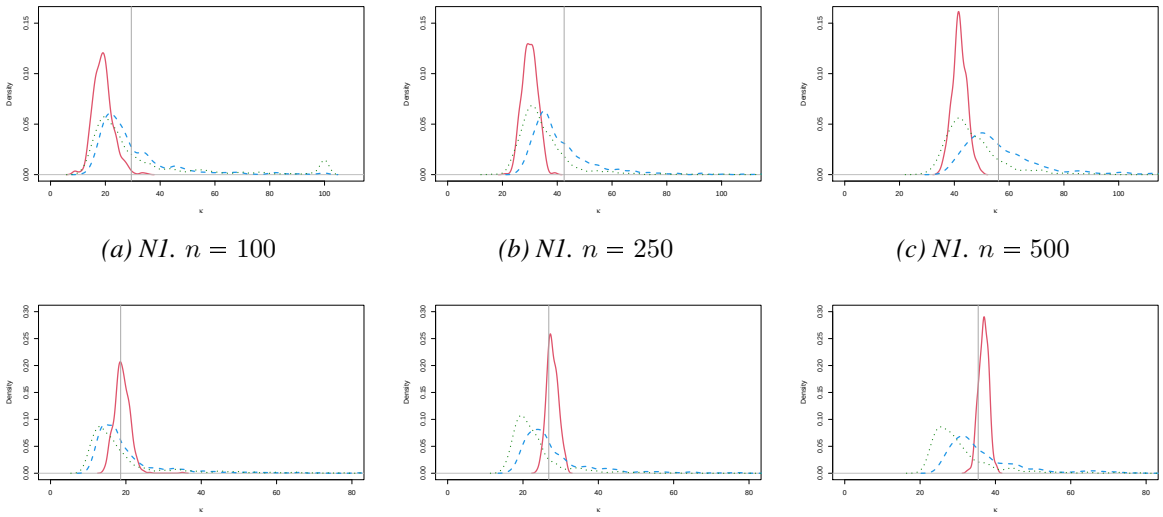
(e) P1 *(f) P2* *(g) G1* *(h) G2*

Figure 3.1: Radial representations of the true mean functions (continuous line) in all models, with representative estimates when $n = 250$ (5th ISE percentile with dashed line, 50th ISE percentile with dotted line and 95th ISE percentile with dotted-dashed line). The smoothing parameters were selected by the refined rule.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

refined rule, the CRSC criterion and the cross-validation method in models N1 and N2, for the different sample sizes. The optimal concentration parameters given by equation (3.16) were also computed and represented as a vertical line. It is observed that, for model N1, the parameters selected by the refined rule are usually smaller than those obtained by the CRSC

rule or the cross-validation criterion, and also than the optimal concentration minimizing the MISE of the estimator. For model N2, the parameters obtained by the refined rule are usually larger than the ones obtained by the other two methods, and are fairly close to the optimal concentration. However, in all scenarios, the distribution of the parameters selected by the refined rule is highly peaked and symmetric, while being skewed for the other two methods. Note that sometimes the CRSC and cross-validation selectors lead to high concentrations (occasionally selecting the maximum possible value) and this behavior is not observed with the refined rule. As expected, the selected parameters are generally larger when increasing the sample size.

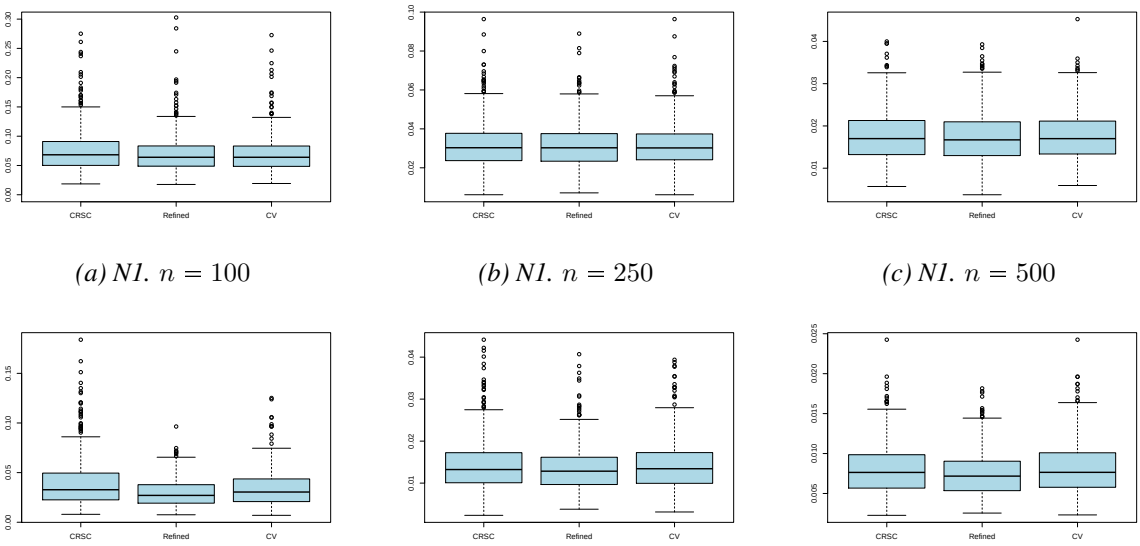


(a) N1. $n = 100$ *(b) N1.* $n = 250$ *(c) N1.* $n = 500$

(d) N2. $n = 100$ *(e) N2.* $n = 250$ *(f) N2.* $n = 500$

Figure 3.2: Kernel density estimators of the obtained values of κ *for models N1 (top row) and N2 (bottom row) with the refined rule (red, continuous line), CRSC (green, dotted line) and cross-validation (blue, dashed line). The sample sizes are* $n = 100$, $n = 250$ *and* $n = 500$. *Grey vertical line represents the optimal concentration parameters. Note that different scales are shown in each panel.*

Regarding the performance of each criterion in terms of the approximated ISE, computed as in (3.19), Figure 3.3 shows boxplots of the approximated ISE for models N1 and N2, the different sample sizes and for each concentration selection method. It can be observed that, for model N1, the distribution of the approximated ISE is similar for the three methods, while for model N2 it seems that the values of the approximated ISE are moderately lower for the refined rule. By means of summarizing the results, Table 3.2 shows the average approximated ISE for both models and all methods. For model N1, the average approximated ISE is very similar for the three smoothing parameter selection methods, although it is slightly lower for the refined rule. For model N2, even though the average approximated ISE is of the same order for the three methods, the refined rule always gives the best results.



(a) N1. $n = 100$ *(b) N1.* $n = 250$ *(c) N1.* $n = 500$

(d) N2. $n = 100$ *(e) N2.* $n = 250$ *(f) N2.* $n = 500$

Figure 3.3: Boxplots of the approximated ISE for models N1 (top row) and N2 (bottom row) with the CRSC, refined rule and cross-validation. The sample sizes are $n = 100$, $n = 250$ *and* $n = 500$. *Note that different scales are shown in each panel.*

In order to graphically assess the quality of the estimators when selecting the concentration parameter with the refined rule, Figure 3.1 shows the true regression functions of models N1 and N2, along with three representatives of the estimators corresponding to the 5th, 50th and 95th percentiles of the approximated ISE, where the sample size was fixed to $n = 250$.

Bernoulli likelihood. The case where the likelihood is a Bernoulli distribution corresponds to the local logistic regression method for circular covariates studied by Di Marzio et al. (2018). The aim here is to estimate the logit function in (3.7). In this case, following Di Marzio et al. (2018), the cross-validation criterion is based on the maximization of

$$\sum_{i=1}^{n} \left(Y_j \hat{g}_{-i}(\Theta_i) - \log\{1 + \exp[\hat{g}_{-i}(\Theta_i)]\}\right), \tag{3.20}$$

with $\hat{g}_{-i}(\cdot)$ the local sine-polynomial estimator of $g(\cdot)$ constructed without the ith observation.

Regarding the concentration parameters selected by each method, Figure 3.4 shows kernel density estimators of the obtained smoothing parameters. In this case, there is not a closed form available for the optimal concentration. For model B1, the behavior of the cross-validation method and the refined rule is quite similar when $n = 100$, while the distribution of the parameters selected by the ECRSC is less concentrated, often selecting the maximum

Table 3.2: Monte Carlo averages of the approximated ISE obtained with the CRSC/ECRSC rule, the refined rule and cross-validation for all models.

		Model 1			Model 2		
	n	(E)CRSC	Refined	CV	(E)CRSC	Refined	CV
N	100	$7.58 \cdot 10^{-2}$	$6.98 \cdot 10^{-2}$	$7.06 \cdot 10^{-2}$	$3.94 \cdot 10^{-2}$	$2.96 \cdot 10^{-2}$	$3.34 \cdot 10^{-2}$
	250	$3.18 \cdot 10^{-2}$	$3.13 \cdot 10^{-2}$	$3.16 \cdot 10^{-2}$	$1.44 \cdot 10^{-2}$	$1.34 \cdot 10^{-2}$	$1.44 \cdot 10^{-2}$
	500	$1.76 \cdot 10^{-2}$	$1.73 \cdot 10^{-2}$	$1.77 \cdot 10^{-2}$	$8.06 \cdot 10^{-3}$	$7.54 \cdot 10^{-3}$	$8.18 \cdot 10^{-3}$
B	100	$8.78 \cdot 10^{-2}$	$7.31 \cdot 10^{-2}$	$6.95 \cdot 10^{-2}$	$4.49 \cdot 10^{-2}$	$3.50 \cdot 10^{-2}$	$3.25 \cdot 10^{-2}$
	250	$3.18 \cdot 10^{-2}$	$2.97 \cdot 10^{-2}$	$2.91 \cdot 10^{-2}$	$1.58 \cdot 10^{-2}$	$1.32 \cdot 10^{-2}$	$1.50 \cdot 10^{-2}$
	500	$1.77 \cdot 10^{-2}$	$1.69 \cdot 10^{-2}$	$1.66 \cdot 10^{-2}$	$8.08 \cdot 10^{-3}$	$7.32 \cdot 10^{-3}$	$8.28 \cdot 10^{-3}$
P	100	$4.04 \cdot 10^{-3}$	$3.46 \cdot 10^{-3}$	$4.09 \cdot 10^{-3}$	$4.69 \cdot 10^{-4}$	$4.66 \cdot 10^{-4}$	$4.81 \cdot 10^{-4}$
	250	$1.78 \cdot 10^{-3}$	$1.67 \cdot 10^{-3}$	$1.87 \cdot 10^{-3}$	$2.02 \cdot 10^{-4}$	$2.00 \cdot 10^{-4}$	$2.05 \cdot 10^{-4}$
	500	$9.76 \cdot 10^{-4}$	$9.32 \cdot 10^{-4}$	$1.02 \cdot 10^{-3}$	$1.10 \cdot 10^{-4}$	$1.09 \cdot 10^{-4}$	$1.11 \cdot 10^{-4}$
G	100	$7.35 \cdot 10^{-2}$	$6.13 \cdot 10^{-2}$	$8.19 \cdot 10^{-2}$	$1.99 \cdot 10^{-2}$	$1.50 \cdot 10^{-2}$	$1.79 \cdot 10^{-2}$
	250	$3.54 \cdot 10^{-2}$	$2.75 \cdot 10^{-2}$	$3.67 \cdot 10^{-2}$	$9.22 \cdot 10^{-3}$	$6.53 \cdot 10^{-3}$	$7.88 \cdot 10^{-3}$
	500	$2.01 \cdot 10^{-2}$	$1.53 \cdot 10^{-2}$	$1.95 \cdot 10^{-2}$	$5.01 \cdot 10^{-3}$	$3.72 \cdot 10^{-3}$	$4.38 \cdot 10^{-3}$

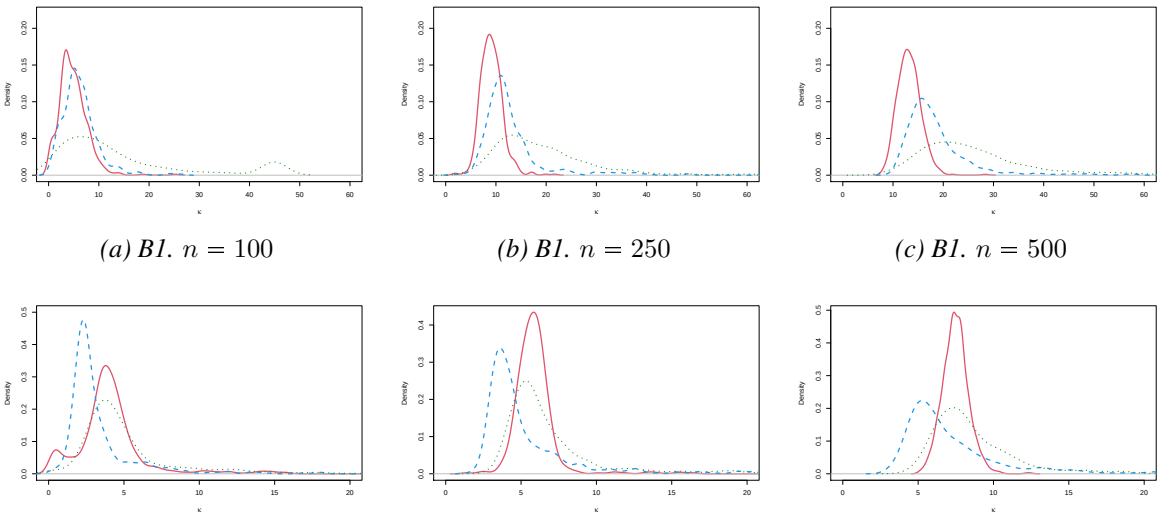
possible value. Surprisingly, when the sample size increases, the estimated distribution of the selected parameters is less peaked, specially for the cross-validation method. For model B2, the parameters selected by the ECRSC and refined rule are centered around the same values, while the concentrations obtained by cross-validation are usually smaller. While the latter method produces less peaked densities when increasing the sample size, the density of the parameters selected by the refined rule is more peaked for larger n.

Next, the performance of the three methods (in terms of the approximated ISE) is explored, for which boxplots are represented in Figure 3.5. It is observed that cross-validation outperforms the other two methods for model B1, although the behavior of all methods becomes more similar when increasing the sample size. On the contrary, for model B2, the refined rule obtains smaller values of the approximated ISE when $n = 250$ and $n = 500$. In all cases, the largest values of the approximated ISE are obtained with the ECRSC method. The same conclusions are derived when observing the average approximated ISE in Table 3.2.

Focusing now on the refined rule, Figure 3.1 shows radial representations of the true models and three estimators obtained after simulating data from models B1 and B2 when $n = 250$. The estimators correspond to the 5th, 50th and 95th percentiles of the approximated ISE.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

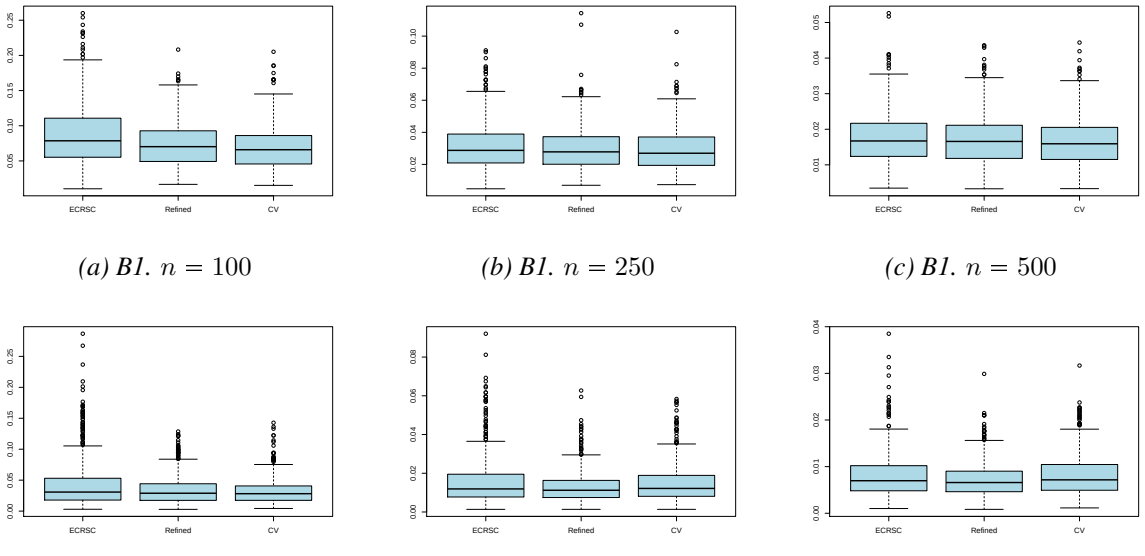
Poisson likelihood. Now, Poisson data is considered, and the aim is to estimate the transformed mean function given by the logarithm, *i.e.*, $g(\theta) = \log[\mu(\theta)]$. Similarly to (3.20), the



(a) B1. $n = 100$ *(b) B1.* $n = 250$ *(c) B1.* $n = 500$

(d) B2. $n = 100$ *(e) B2.* $n = 250$ *(f) B2.* $n = 500$

Figure 3.4: Kernel density estimators of the obtained values of κ *for models B1 (top row) and B2 (bottom row) with the refined rule (red, continuous line), ECRSC (green, dotted line) and cross-validation (blue, dashed line). The sample sizes are* $n = 100$*,* $n = 250$ *and* $n = 500$*. Note that different scales are shown in each panel.*



(a) B1. $n = 100$ *(b) B1.* $n = 250$ *(c) B1.* $n = 500$

(d) B2. $n = 100$ *(e) B2.* $n = 250$ *(f) B2.* $n = 500$

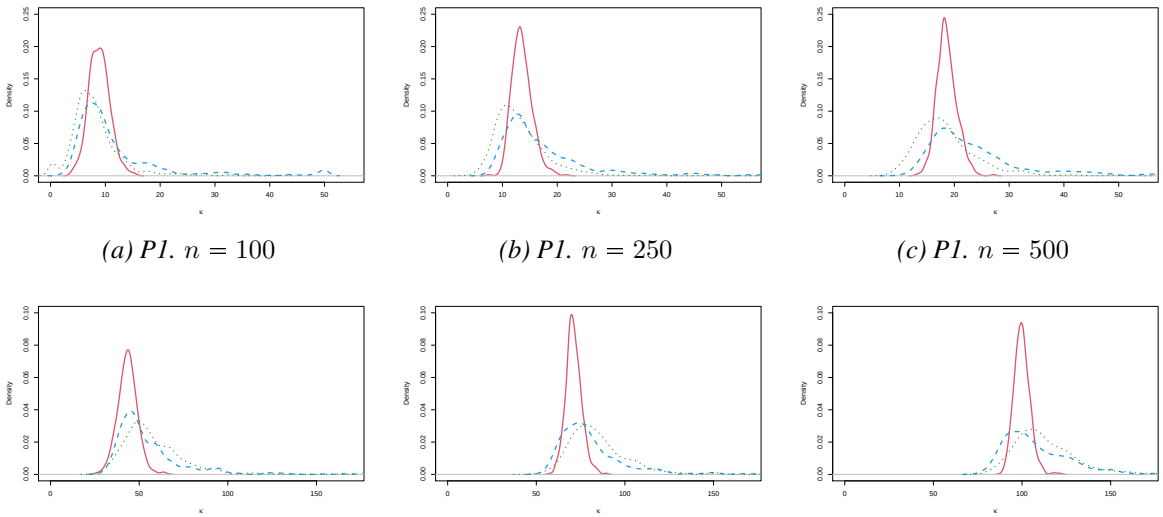
Figure 3.5: Boxplots of the approximated ISE for models B1 (top row) and B2 (bottom row) with the ECRSC, refined rule and cross-validation. The sample sizes are $n = 100$*,* $n = 250$ *and* $n = 500$*. Note that different scales are shown in each panel.*

cross-validation method selects κ by maximizing

$$\sum_{i=1}^{n}\{Y_j\hat{g}_{-i}(\Theta_i) - \exp[\hat{g}_{-i}(\Theta_i)]\},$$

where $\hat{g}_{-i}(\cdot)$ denotes the local sine-polynomial estimator of $g(\cdot)$ constructed without the ith observation.

The kernel density estimators of the selected κ with each selection method are shown in Figure 3.6. The estimated distribution of the obtained parameters is similar to the ones attained in the normal likelihood case, since the estimated density obtained for the refined rule is generally symmetric and highly peaked, while the other two are skewed, selecting sometimes concentration parameters which are overly large.

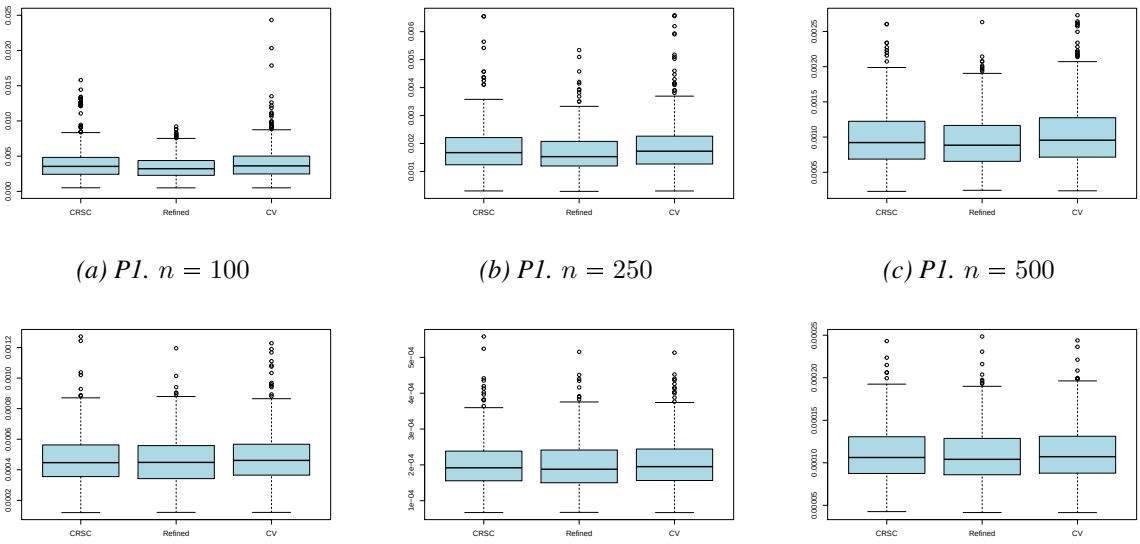


(a) P1. $n = 100$ *(b) P1.* $n = 250$ *(c) P1.* $n = 500$

(d) P2. $n = 100$ *(e) P2.* $n = 250$ *(f) P2.* $n = 500$

Figure 3.6: Kernel density estimators of the obtained values of κ *for models P1 (top row) and P2 (bottom row) with the refined rule (red, continuous line), ECRSC (green, dotted line) and cross-validation (blue, dashed line). The sample sizes are* $n = 100$, $n = 250$ *and* $n = 500$. *Note that different scales are shown in each panel.*

Boxplots of the approximated ISE obtained with each method are displayed in Figure 3.7. For model P1, it can be observed that the approximated ISEs obtained with the refined rule are slightly smaller than the ones attained with the ECRSC and the cross-validation method. On the other hand, for model P2, it seems that the three methods achieve a similar performance in terms of ISE. The average approximated ISE for each method can be found in Table 3.2, and it is observed that it is of the same order for the three methods, although always slightly lower for the refined rule. As expected, the average approximated ISE becomes lower as the sample size increases.



(a) P1. $n = 100$ *(b) P1.* $n = 250$ *(c) P1.* $n = 500$

(d) P2. $n = 100$ *(e) P2.* $n = 250$ *(f) P2.* $n = 500$

Figure 3.7: Boxplots of the approximated ISE for models P1 (top row) and P2 (bottom row) with the ECRSC, refined rule and cross-validation. The sample sizes are $n = 100$*,* $n = 250$ *and* $n = 500$*. Note that different scales are shown in each panel.*

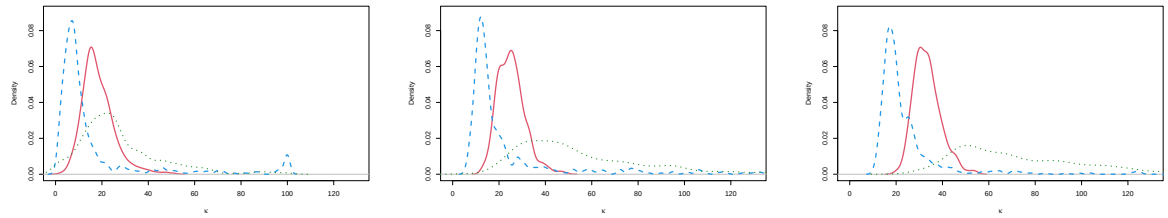
Focusing on the behavior of the refined rule, representative estimates of the target functions are given in Figure 3.1, where estimators corresponding to the 5th, 50th and 95th percentiles of the approximated ISE (when $n = 250$) are depicted. The target function is more difficult to estimate in model P1, where the mean function takes low values. For model P2, however, the three representative estimates are very close to the true target function.

Gamma likelihood. Consider now the case with a gamma likelihood. The target function is the logarithm of the mean function, *i.e.*, $g(\theta) = \log[\mu(\theta)]$. In this case, the log-likelihood depends on the unknown shape parameter. Thus, instead of maximizing the cross-validation function based on the local log-likelihood, κ is selected by minimizing

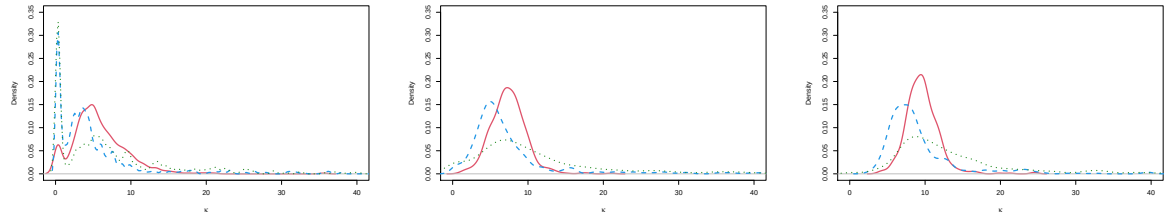
$$\sum_{i=1}^{n} \{Y_i - \exp[\hat{g}_{-i}(\Theta_i)]\}^2.$$

The kernel density estimators of the concentration parameters selected with each method are shown in Figure 3.8. For model G1, the shape of the estimated density for the cross-validation method is more peaked than in the previous scenarios, although it is still quite asymmetric, selecting sometimes values of the smoothing parameter which are too large. On the other hand, the values of the concentration obtained by the ECRSC rule are highly variable, and even more when increasing the sample size. On the contrary, values computed

with the refined rule are reasonably concentrated, and are usually larger than those selected by the cross-validation method. Regarding results for model G2, the selection of κ seems more complicated when $n = 100$, with the ECRSC rule and the cross-validation method frequently selecting values of the concentration very close to zero. Although the refined rule also leads to very small values of the concentration sometimes, this behavior is not shown as often as with the other two methods. As usual, both the cross-validation method and the ECRSC rule sometimes select exceedingly large concentration parameters.



(a) G1. $n = 100$ *(b) G1.* $n = 250$ *(c) G1.* $n = 500$



(d) G2. $n = 100$ *(e) G2.* $n = 250$ *(f) G2.* $n = 500$

Figure 3.8: Kernel density estimators of the obtained values of κ *for models G1 (top row) and G2 (bottom row) with the refined rule (red, continuous line), ECRSC (green, dotted line) and cross-validation (blue, dashed line). The sample sizes are* $n = 100$, $n = 250$ *and* $n = 500$. *Note that different scales are shown in each panel.*

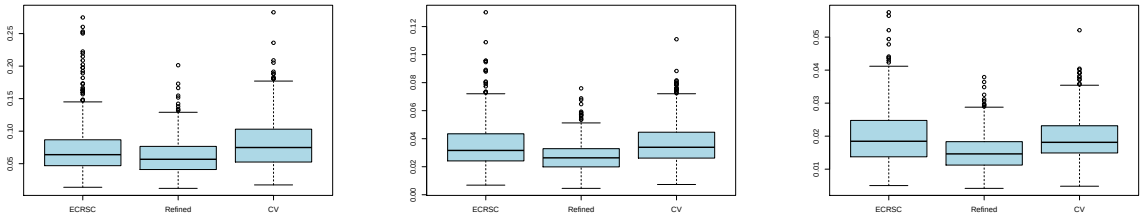
Boxplots of the approximated ISE for each method are represented in Figure 3.9. For both models and all sample sizes, it seems that values of the approximated ISE are usually smaller when employing the refined rule. This is reaffirmed when acknowledging the averaged ISE, shown in Table 3.2, where it can be seen that the refined rule achieves the lowest average approximated ISE in all cases.

Finally, Figure 3.1 represents the true target functions for models G1 and G2, along with the estimators constituting the 5th, 50th and 95th percentiles of the approximated ISE (when $n = 250$ and the concentration was selected by the refined rule).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3.5 Real data illustration

In this section, the practical use of the circular local likelihood estimator and the refined concentration selection rule is illustrated with different real data sets. In all examples, the



(a) G1. $n = 100$ *(b) G1.* $n = 250$ *(c) G1.* $n = 500$

(d) G2. $n = 100$ *(e) G2.* $n = 250$ *(f) G2.* $n = 500$

Figure 3.9: Boxplots of the approximated ISE for models G1 (top row) and G2 (bottom row) with the ECRSC, refined rule and cross-validation. The sample sizes are $n = 100$, $n = 250$ *and* $n = 500$. *Note that different scales are shown in each panel.*

target function is estimated by fitting a local sine-polynomial of degree $p = 1$, where the concentration parameter is obtained with the refined rule selector. The bias and variance are estimated with a sine-polynomial of degree 3, and the pilot smoothing parameter is selected by the CRSC/ECRSC criterion. Employing the asymptotic normal distribution of the local likelihood estimator, pointwise confidence bands for the estimated functions, with a confidence level of 95%, are also computed. Although the bias and variance of the estimators for each value of the predictor variable can be obtained as in Section 3.2, they can change abruptly due to the estimation of higher order derivatives, which may be unstable. Hence, for the pointwise confidence bands, kernel weighted averages of the estimated bias and variance are used:

$$\hat{B}^k_{p,\nu}(\theta_0;\kappa) = \int_0^{2\pi} \hat{B}_{p,\nu}(\theta;\kappa)K_\kappa(\theta-\theta_0)d\theta \quad \text{and} \quad \hat{V}^k_{p,\nu}(\theta_0;\kappa) = \int_0^{2\pi} \hat{V}_{p,\nu}(\theta;\kappa)K_\kappa(\theta-\theta_0)d\theta.$$

Fan et al. (1995) showed, in the context of generalized linear models, that the local likelihood estimator is asymptotically normally distributed. Employing the normal asymptotic distribution, pointwise confidence bands for the estimated functions with, approximately, $(1-\alpha)$ coverage probability can be constructed as

$$\hat{g}(\theta_0) - \hat{B}^k_{p,\nu}(\theta_0;\kappa) \pm \Phi(1-\alpha/2)\hat{V}^k_{p,\nu}(\theta_0;\kappa)^{1/2}, \tag{3.21}$$

where $\Phi(\cdot)$ denotes the quantile of a standard normal distribution.

Human motor resonance data: normal response. An example of a normal response in a regression model with a circular covariate is the human motor resonance data set obtained by Puglisi et al. (2017), which is described in Appendix A.4. With the purpose of illustrating the applicability of circular local likelihood estimator, the relationship between the angular position of the hand (circular predictor variable) and the H-reflex amplitude (real-valued response variable) will be explored. Note that, in this case, the local likelihood estimator is equivalent to the least-squares estimator for a circular predictor (see Section 2.1.1). The data set, which is composed of $n = 70$ observations, is depicted in Figure 3.10 with the estimated regression function. Note that the H-reflex amplitude increases when the angular position ranges from $\pi/2$ to $5\pi/4$ and decreases between $5\pi/4$ and 2π. The 95% pointwise confidence band based on (3.21) is also represented.

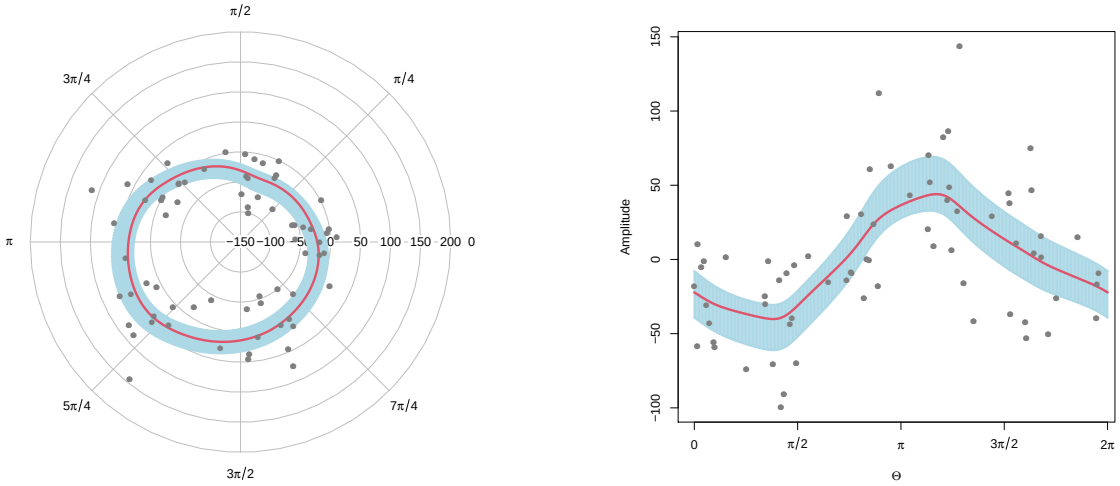


Figure 3.10: Radial and planar representations of the human resonance data with the non-parametric estimator of the regression function and pointwise 95% confidence bands.

Sand hoppers data: Bernoulli response. Next, the circular local likelihood estimator for a Bernoulli response is illustrated with a subset of the sand hoppers data set, described in Appendix A.2. Here, the aim is to estimate the probability of the animals belonging to each of the two species, in accordance to the direction in which they escape. Out of the 1828 observations available in the data set, only the adult animals were considered, given that for juvenile sand hoppers it was not possible to determine the specie they belonged to. The final sample size was $n = 1644$, where 867 animals belonged to the *Talitrus saltator* specie and 777 to the *Talorchestia brito* specie. The logit function in (3.7) was estimated and the estimated probability of belonging to the *Talorchestia brito* specie was represented in Figure 3.11. The

estimated probability is around 0.5 for most of the escape directions, but there is a reasonable reduction of the probability of belonging to the *Talorchestia brito* specie for escape directions around π. The pointwise 95% confidence band suggests that this reduction in the probability is not just an artifact due to the randomness of the data.

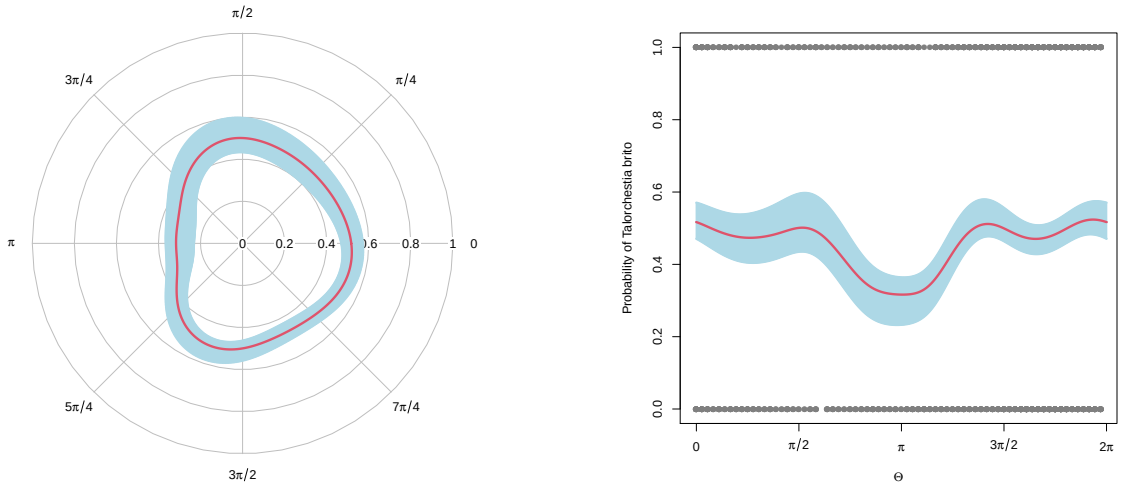


Figure 3.11: Radial and planar representations of the sandhopper data with the estimated probability of belonging to the Talorchestia brito specie and pointwise 95% confidence bands.

Spike count data: Poisson model. The estimator in the case of a Poisson likelihood is now illustrated with the spike count data set, described in Appendix A.5. The data consists of the total number of spikes per trial (response variable representing counts) and the stimulus direction (circular predictor variable) in a V5/MT cell and the sample size is $n = 68$.

Radial and planar representations of the data set are shown in Figure 3.12. To get more insight into the relationship between the variables, the local circular kernel weighted likelihood function was maximized to obtain the estimate of the transformed mean function $g(\theta) = \log[\mu(\theta)]$. The estimated mean function, $\hat{\mu}(\theta) = \exp[\hat{g}(\theta)]$, is represented in Figure 3.12. According to the estimator, the mean count of spikes is larger when the stimulus direction is approximately $\pi/2$, and a lower expected number of spikes is estimated when the stimulus direction is opposite from that, approximately at $3\pi/2$.

Note that this example was used just to illustrate the use of the circular local likelihood estimator when the response variable follows a conditional Poisson distribution. However, Figure 3.12 shows some evidences that the Poisson distribution might not be completely adequate to model this data set, given that, for the Poisson distribution, the mean and the variance are the same, and it seems that the data set could present some over-dispersion with respect to the Poisson. A different conditional distribution will be employed in Chapter 4 to account for this issue.

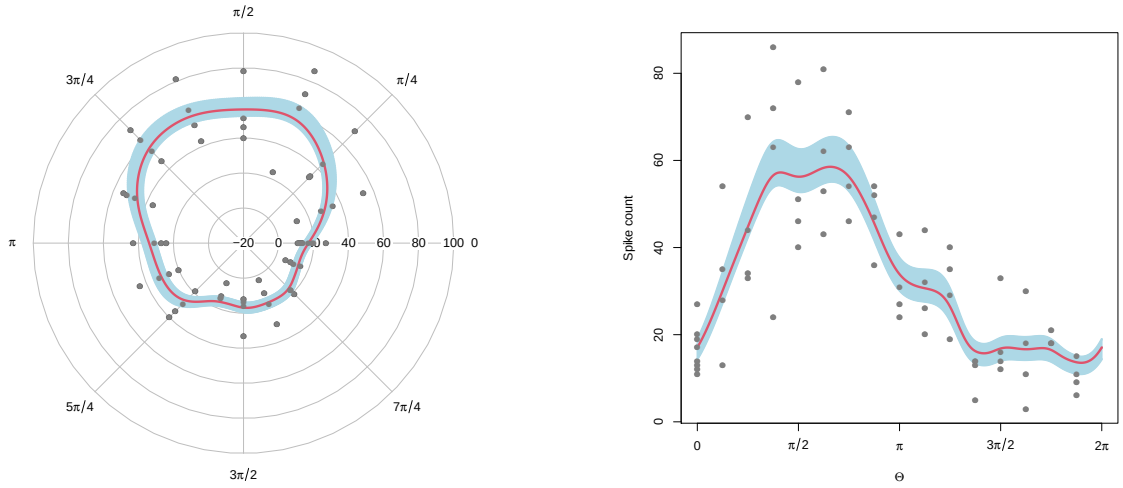


Figure 3.12: Radial and planar representations of the spike count data with the nonparametric estimator of the mean function and pointwise 95% confidence bands.

Pm10 particles data: gamma model. The last example is the pm10 particle data set, described in Appendix A.6. In this case, the response variable is the concentration of pm10 particles, which presents an asymmetric conditional distribution, and the predictor variable is the wind direction. In order to avoid time dependence issues, a subsample of the data was employed, with observations taken every six hours. In this section, only the wind direction was considered as the predictor, but an extension accounting for the effect of the wind speed is discussed in Section 3.6. The aim is to study how the pm10 particles change with the direction of the wind. One must note that there is a pulp factory approximately 2.5 km south-west from the meteorological station and, hence, it would be of interest to determine if the concentration of pm10 particles is actually higher when the wind blows from the factory's direction. For this aim, observations in which the wind speed was lower than 1 km/hour were removed, since wind speeds between 0 and 1 km/hour are considered as periods of calm according to the Beaufort scale. The final sample size is $n = 1156$.

Left panel of Figure 3.13 represents the pm10 measurements against the wind direction. By assuming a gamma likelihood, the logarithm of the conditional mean function was estimated with the estimator presented in Section 3.1. The estimator of the mean is represented in the left panel of Figure 3.13, where the approximate direction of the wind that blows from the pulp factory is represented with a star. Since the values of the response variable distort the representation of the mean function, the center panel of Figure 3.13 shows a zoomed planar representation of the data and the estimated mean, with the 95% pointwise confidence band. It can be observed that the mean concentration of pm10 particles changes with the direction of the wind and it seems that the concentration is higher when the wind blows from the South-West/West direction.

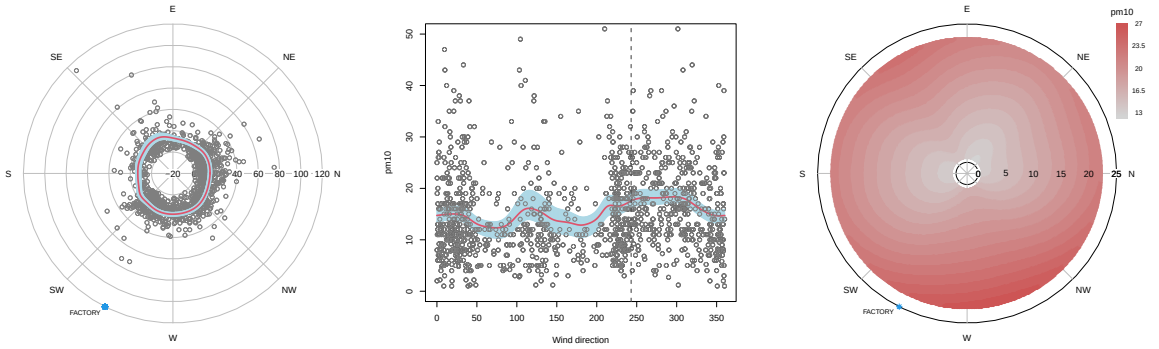


Figure 3.13: Polar representation of the pm10 concentration against wind direction, with the estimated regression curve and 95% pointwise confidence band (left). Planar zoomed representation of the pm10 concentration against wind direction, with the estimated regression curve and 95% pointwise confidence band (center). Estimation of the mean pm10 concentration as a function of the wind direction and wind speed (right). The star (left and right) and vertical dashed line (center) indicate the direction of the factory.

3.6 Extensions to partially linear and additive models

Throughout this chapter, the estimation of conditional characteristics when there is just a single circular predictor was considered. However, in many practical situations, several covariates (of different nature) may influence a response variable. Consider, for example, the pm10 particles data set described in Appendix A.6. In order to study the concentration of pm10 particles, it seems reasonable to consider not only the wind direction as the covariate, but also the wind speed. Furthermore, these covariates could enter the model in a nonparametric form (estimated, for example, by maximizing the local log-likelihood as in Fan et al. (1998)) or in a parametric (possibly linear) way.

Contemplate a more general model with circular and real-valued covariates, with some of them entering the model parametrically and others having a nonparametric effect on the response. For the parametric part, assume that real-valued covariates enter the model linearly, while circular covariates can also enter the model linearly by means of their sine and cosine components. Let $g(\cdot)$ be the target function to estimate, which depends on real-valued covariates $X_1, \ldots, X_k, X_{k+1}, \ldots, X_r$ and circular covariates $\Theta_1, \ldots, \Theta_j, \Theta_{j+1}, \ldots, \Theta_s$. Assume that $X_1, \ldots, X_k$ have a linear effect on the response, $\Theta_1, \ldots, \Theta_j$ also have a linear effect on the response through their sine and cosine components and $X_{k+1}, \ldots, X_r, \Theta_{j+1}, \ldots, \Theta_s$ have an unknown (*i.e.*, nonparametric) effect on the response. Then, the target function $g(\cdot)$

can be modeled as

$$
\begin{aligned}
&g(X_1,\ldots,X_k,X_{k+1},\ldots,X_r,\Theta_1,\ldots,\Theta_j,\Theta_{j+1},\ldots,\Theta_s)\\
&=\alpha_0+\alpha_1X_1+\ldots+\alpha_kX_k+\gamma_{11}\cos(\Theta_1)+\gamma_{12}\sin(\Theta_1)+\ldots+\gamma_{j1}\cos(\Theta_j)+\gamma_{j2}\sin(\Theta_j)\\
&+g^L_{k+1}(X_{k+1})+\ldots+g^L_r(X_r)+g^C_{j+1}(\Theta_{j+1})+\ldots+g^C_s(\Theta_s),
\end{aligned}
$$

where $\alpha_0,\ldots,\alpha_k,\gamma_{11},\gamma_{12},\ldots,\gamma_{j1},\gamma_{j2}$ are the parameters corresponding to the parametric part and $g^L_{k+1}(\cdot),\ldots,g^L_r(\cdot),g^C_{j+1}(\cdot),\ldots,g^C_s(\cdot)$ represent unknown smooth functions, which are respectively defined on $\mathbb{R}$ or $\mathbb{T}$.

Model estimation can be carried out through a backfitting algorithm: first, the global parameters are estimated by maximizing the log-likelihood. By including the estimated global parameters on the model, one of the nonparametric functions is estimated by maximizing the local log-likelihood (with the approach of Fan et al. (1998) if the covariate is real-valued or with the method of Section 3.1 if the covariate is circular). Now, the new estimated function is included in the model and the next nonparametric function is estimated as before, until one has estimated all smooth functions. The previous steps are repeated until convergence. This approach entails the selection of one bandwidth parameter for each real-valued covariate entering the model nonparametrically as well as one concentration parameter for each circular covariate with a nonparametric effect.

For the pm10 particles data set, a gamma likelihood and a partially linear model for the logarithm of the mean function were assumed:

$$g(X,\Theta)=\log[\mu(X,\Theta)]=\alpha_0+\alpha_1X+g^C(\Theta),$$

where X denotes the wind speed, Θ denotes the wind direction and $g^C(\cdot)$ is a smooth function accounting for the effect of the wind direction. The estimated average pm10 concentration is represented in the right panel of Figure 3.13, where the angle represents the wind direction, the distance to the center of the circle represents the wind speed and the color indicates the estimated mean concentration of pm10 particles (as indicated in the legend). As it can be seen, larger values of the wind speed lead to an increase in the concentration of pm10 particles. The largest values of the concentration are obtained for wind directions coming from the South-West/West direction, where the pulp factory is located.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Chapter 4

Joint estimation for over/under-dispersed regression models

The circular local likelihood methodology exposed in the previous chapter constitutes a general approach for flexibly modeling relationships between a circular predictor and a general real-valued (discrete or continuous) response. Nevertheless, this method implies the selection of the conditional distribution of the response which, in practice, is usually set as a distribution belonging to the exponential family, as it was done in all the real data illustrations of Section 3.5.

Although the exponential family allows to model a great variety of situations, it assumes a specific relationship between the mean and the variance of the conditional distribution, which may not always be adequate in practice. As a motivational example, consider the spike count data set, described in Appendix A.5, and used to illustrate the circular local likelihood approach in Section 3.5. In this example, there is a need to study the relationship between the number of spikes produced in a MT/V5 cell of a macaque monkey and the direction of the stimulus provoking the reactions in the cell. Data of this type often arise in the neurological field, and the common approach to analyze it is to assume that the data follow a conditional Poisson distribution, where the mean parameter changes with the covariate through a parametric model.

Usually, the logarithm of the conditional mean is assumed to follow a simple parametric form, similar to the cosine model in (1.12) (see Georgopoulos et al., 1982; Cronin et al., 2010). Some authors have pointed out that such simple curves may be quite restrictive regarding the shape of the regression function. However, in order to model more complex relationships between the variables, one can include more cosine and sine terms, similarly to the regression model in (1.13). Kaufman et al. (2005) pointed out different issues regarding these parametric models, and proposed a nonparametric alternative, in which the estimation is approached by using a periodic extension of Bayesian adaptive regression splines. In the same line, the circular local likelihood method of Chapter 3 allows a kernel-based estimation.

However, although the mean spike count has been the target issue to study in neurological works, it has been pointed out that the variability of spike counts has also a relevant and constructive role in the study of neuronal responses (Stein et al., 2005; Ermentrout et al., 2008). Lombardo et al. (2018) raised the question of how the neural variability itself depends on the directional stimulus, and elaborated on the relevance of this hypothesis. Yet, when assuming a Poisson model, the mean and variance functions are the same and, thus, several

authors indicate that the Poisson distribution is not completely adequate for this kind of data, since the data may be over- or under-dispersed (or both, depending on the direction of the stimulus). Consequently, other alternatives such as the negative binomial, which allows for over-dispersion, have been used in this setting. Nevertheless, in order to account for more complex situations where both over- and under-dispersion are present, specific dispersion models may be used. Stevenson (2016) made use of the Conway-Maxwell-Poisson distribution (a generalization of the Poisson distribution with an additional parameter controlling the dispersion) to analyze the directional tuning of spike counts.

This chapter explores the use of an alternative dispersion model: the double exponential family (Efron, 1986), which includes a dispersion parameter with respect to the exponential model. The double exponential family has been used to study the estimation of both the mean and the dispersion in the context of Generalized Linear Models (GLMs) or Generalized Additive Models (GAMs) with euclidean variables (Gijbels et al., 2010; Gijbels and Prosdocimi, 2011, 2012). The aim of this chapter is to nonparametrically estimate, jointly, location and dispersion conditional functions by maximizing the local log-likelihood of each curve employing circular kernels and assuming a double exponential family. Hence, the approach presented in this chapter suitably accounts for the circular nature of the predictor and allows, in a flexible way, to model under/over-dispersion patterns. Note that this method is not only applicable to count data but can also deal with a wide range of cases such as proportions, normal data or continuous but non-normal data. For instance, in the simple case of a double normal likelihood, the method provides an approach to nonparametrically estimate both the conditional mean and variance functions which, up to the authors' knowledge, has not been done yet in the context of circular data.

The organization of the chapter is as follows: Section 4.1 introduces methods which allow to model over- or under-dispersion with respect to the Poisson distribution. Section 4.2 presents the joint estimation of the mean and dispersion functions in a double exponential model by maximizing the circular local likelihood. The practical performance of the estimators is explored via simulations in Section 4.3 and the analysis of the spike count data set is given in Section 4.4. The contents of this chapter are collected in Alonso-Pena et al. (2022a).

Contents

4.1 Modeling the dispersion 75
4.1.1 The Conway-Maxwell-Poisson model 75
4.1.2 The double exponential family 76
4.1.3 Double exponential regression model for circular covariate 77
4.2 Joint estimation via circular local likelihood 78
4.3 Simulation study 81
4.3.1 Simulation scenarios 81
4.3.2 Simulation results 82

4.3.3 Joint visualization of mean and dispersion functions 84
4.4 Analysis of the number of spikes in a macaque's neuron 85

4.1 Modeling the dispersion

Quite often in practice, real data show a behavior deviating from the exponential family, in the sense that the variance and the mean do not exhibit the specific relationship determined by this family along the covariate range. In fact, regarding the spike count motivation, there is considerable evidence that spike counts in MT/V5 neurons can be more or less variable than a Poisson, and this variability may depend on the stimulus direction (Butts and Goldman, 2006; Stevenson, 2016; Lombardo et al., 2018). Additionally, as previously mentioned, the estimation of the variability of the spike counts may be an essential task regarding research of the neuronal responses and, hence, the appropriate modeling of the dispersion becomes necessary. In some situations, the negative binomial distribution is used for over-dispersed count data, but a more general approach is to consider proper dispersion models. In this section, some dispersion models which can be applied to the directional tuning of spike counts will be shown. In Section 4.1.1, the Conway-Maxwell-Poisson regression method, employed by Stevenson (2016) to estimate both the mean and the dispersion functions, is presented. Section 4.1.2 exposes another alternative to model dispersion: the double exponential family (Efron, 1986), which contains the double Poisson as a particular case, and can be used to model over- and under-dispersion. Finally, Section 4.1.3 gives the formulation of the double exponential regression model, already used to fit over-dispersed or under-dispersed data in the context of GLM or GAM models with euclidean variables.

4.1.1 The Conway-Maxwell-Poisson model

Stevenson (2016) modeled the number of spikes conditional to the direction of tuning by assuming that the conditional distribution of the number of spikes was a Conway-Maxwell-Poisson (Conway and Maxwell, 1962; Shmueli et al., 2005), also known as COM-Poisson. If a random variable Y follows a COM-Poisson distribution with parameters λ and ζ, its mass probability function is given by

$$\mathbb{P}(Y = y) = \frac{1}{(y!)^{\zeta}} \frac{\lambda^y}{Z(\lambda, \zeta)},$$

where $Z(\lambda, \zeta) = \sum_{t=0}^{\infty} \lambda^t/(t!)^{\zeta}$ and $\zeta \geqslant 0$. The parameter ζ controls the dispersion of the distribution: if $\zeta < 1$, the COM-Poisson is over-dispersed with respect to the Poisson; $\zeta > 1$ implies under-dispersion and, for $\zeta = 1$, the COM-Poisson yields the Poisson distribution with parameter λ. Regarding the use of the COM-Poisson model in the regression context with real-valued variables, Sellers and Shmueli (2010) let the parameter λ depend on a pre-

dictor variable through a multiple linear model, considering constant dispersion. In addition, a mixture model accounting for group-level dispersion was proposed by Sellers and Shmueli (2013), and the modeling of the dispersion as a function of the covariates and the extension to GAM models were studied by Chatla and Shmueli (2018).

Regarding circular covariates, when modeling the directional tuning of neural spikes, Stevenson (2016) let both parameters λ and ζ depend on the tuning direction, Θ, by modeling a transformation of the functions $\lambda(\theta)$ and $\zeta(\theta)$ with a linear combination of sine and cosine terms:

$$\begin{aligned} g(\theta) =& \eta[\lambda(\theta)] = \sum_{j=0}^{k} \left[\alpha_j \cos(j\theta) + \delta_j \sin(j\theta)\right], \\ \xi(\theta) =& h[\zeta(\theta)] = \sum_{j=0}^{\tilde{k}} \left[\tilde{\alpha}_j \cos(j\theta) + \tilde{\delta}_j \sin(j\theta)\right], \end{aligned} \tag{4.1}$$

where $\eta(\cdot)$ and $h(\cdot)$ are link functions. Regardless of the apparent simplicity of the COM-Poisson regression model, the estimation of the conditional mean and dispersion functions presents some important drawbacks. First, if $Y|\Theta = \theta \sim$ COM-Poisson$[\lambda(\theta), \zeta(\theta)]$, it holds

$$\mathbb{E}(Y|\Theta = \theta) = \lambda(\theta)\frac{\partial}{\partial\lambda(\theta)} Z[\lambda(\theta), \zeta(\theta)], \quad \text{Var}(Y|\Theta = \theta) = \lambda(\theta)\frac{\partial}{\partial\lambda(\theta)}\mathbb{E}(Y|\Theta = \theta).$$

Consequently, the conditional mean and conditional variance functions depend on the derivatives of the normalizing constant $Z[\lambda(\theta), \zeta(\theta)]$, which is an infinite sum with no closed form. Some efforts have been done in order to approximate the normalizing constant. An asymptotic approximation of $Z[\lambda(\theta), \zeta(\theta)]$ by Shmueli et al. (2005) yields

$$\mathbb{E}(Y|\Theta = \theta) = \lambda(\theta)\frac{\partial}{\partial\lambda(\theta)} Z[\lambda(\theta), \zeta(\theta)] \approx \lambda(\theta)^{1/\zeta(\theta)} - \frac{\zeta(\theta) - 1}{2\zeta(\theta)}, \tag{4.2}$$

$$\text{and} \quad \text{Var}(Y|\Theta = \theta) \approx \frac{\lambda(\theta)^{1/\zeta(\theta)}}{\zeta(\theta)}. \tag{4.3}$$

Thus, a second drawback is the dependence of the mean function on both $\lambda(\theta)$ and $\zeta(\theta)$, and the loss of interpretation of the function $\lambda(\theta)$ with respect to the Poisson model. In addition, the approximation in (4.2) is quite accurate for $\zeta(\theta) \leqslant 1$ or large $\lambda(\theta)$ (see Shmueli et al., 2005), but might not work well for situations with under-dispersion and a small $\lambda(\theta)$. Although some more accurate approximations have been recently derived (see Gaunt et al., 2019), they may be still not very precise for some values of the parameters and the complicated expression of the approximation makes the estimation procedure more complex.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.1.2 The double exponential family

Given the difficulties and interpretation issues presented by the COM-Poisson model, the aim of this section is to model the dispersion in a simpler, but yet effective, way by employing the

double exponential family of densities. The Poisson distribution belongs to the exponential family, which also encompasses a large class of distributions, such as normal or binomial models. In an exponential family, the conditional density of a random variable Y is given by

$$f_e(y; z, \varphi) = \exp\left[\frac{yz - b(z)}{\varphi} + c(y, \varphi)\right], \tag{4.4}$$

where z is the natural location parameter, φ is the scale parameter and the functions $b(\cdot)$ and $c(\cdot, \cdot)$ are known for each specific model. In the particular case of the Poisson distribution with mean λ, it holds $z = \log\lambda$, $\varphi = 1$, $b(\cdot) = \exp(\cdot)$ and $c(\cdot, 1) = -\log(\cdot!)$.

For the general exponential family $\mathbb{E}(Y) = b'(z)$ and $\text{Var}(Y) = \varphi b''(z)$. Thus, this family assumes a specific relationship between the mean and the variance of the variable of interest. However, as stated before, real data often show over-dispersion or under-dispersion. In order to generalize the exponential family to include an additional dispersion parameter, Efron (1986) introduced the double exponential family of densities, which was also studied by Gelfand and Dalal (1990) and Dey et al. (1997), among others. Let $\hat{z}$ be the choice of the natural location parameter z in (4.4) corresponding to the saturated model maximizing $f_e(y; z, \varphi)$. If a random variable Y is double exponential distributed, its density function is given by

$$f(y; z, \gamma, \varphi) = \tilde{c}(z, \gamma)\gamma^{-\frac{1}{2}}[f_e(y; z, \varphi)]^{\frac{1}{\gamma}} f_e(y; \hat{z}, \varphi), \tag{4.5}$$

where $\tilde{c}(z, \gamma)$ is a normalizing constant. The additional parameter γ controls the dispersion: if $\gamma > 1$, there will be over-dispersion with respect to the exponential family and, if $\gamma < 1$, under-dispersion is obtained. As shown by Efron (1986), the expression $\tilde{c}(z, \gamma)$ can be first-order approximated by 1. Lee and Nelder (2000) showed a study on the use of this approximation, which is reasonably accurate. The unnormalized form of the density will be used, as it allows to have a simple expression for inference purposes. The double exponential model approximately yields $\mathbb{E}(Y) = b'(z)$ and $\text{Var}(Y) = \gamma\varphi b''(z)$. As opposed to the COM-Poisson model, the natural location parameter keeps the same interpretability as in the exponential family, since the expected value in both families is approximately the same.

The particular case of the Poisson distribution with mean λ yields that, approximately, $\mathbb{E}(Y) = \lambda$ and $\text{Var}(Y) = \gamma\lambda$. Consequently, the dispersion parameter in the double Poisson model is interpreted as the ratio between the variance and the mean.

4.1.3 Double exponential regression model for circular covariate

The double exponential family has been used for regression purposes with euclidean variables, for the joint estimation of the mean and dispersion functions (see, for example Dey et al., 1997). Now, some details on the use of the double exponential family for regression with a circular covariate are given. Let (Θ, Y) be a pair of random variables where Y is the variable of interest and Θ is a circular predictor variable. The interest lies in modeling both the transformed conditional mean and the conditional dispersion of Y given Θ. It will

be assumed that the variable Y given $\Theta = \theta$ follows a double exponential distribution, with natural location parameter $z(\theta)$ and dispersion parameter $\gamma(\theta)$.

Following Gijbels et al. (2010), in order to obtain a more convenient expression for the conditional density of $Y|\Theta = \theta$, it holds that the conditional version of (4.5) can be approximated as

$$\tilde{f}[y; z(\theta), \gamma(\theta), \varphi] = \gamma(\theta)^{-\frac{1}{2}} \left(\exp\left\{ \frac{1}{2} D[y, z(\theta)] \right\} \right)^{\frac{-1}{\gamma(\theta)}} f_e(y; \hat{z}, \varphi), \tag{4.6}$$

where $D(\cdot, \cdot)$ is the deviance for the exponential family (4.4):

$$D[y, z(\theta)] = 2 \left(\frac{y\hat{z} - b(\hat{z})}{\varphi} - \frac{yz(\theta) - b[z(\theta)]}{\varphi} \right). \tag{4.7}$$

Note that, in equation (4.6), the unnormalized version of the density was employed, using the approximation $\tilde{c}[z(\theta)\gamma(\theta)] \approx 1, \ \forall\, \theta \in \mathbb{T}$. Now, a simple parametric regression model for the mean and the dispersion can be formulated as

$$\begin{aligned} g(\theta) =& \eta[\mu(\theta)] = \eta\{b'[z(\theta)]\} = \alpha_{\mu,0} + \alpha_{\mu,1} \cos(\theta - \alpha_{\mu,2}), \\ \psi(\theta) =& h^{-1}[\gamma(\theta)] = \alpha_{\gamma,0} + \alpha_{\gamma,1} \cos(\theta - \alpha_{\gamma,2}), \end{aligned} \tag{4.8}$$

where $\eta(\cdot)$ is a link function and $h(\cdot)$ is a twice differentiable positive function satisfying $h^{-1}(\cdot)$ exists and $h'(\cdot) \neq 0$. When $\eta(\cdot)$ is chosen as the canonical link, which will be used throughout this chapter, it holds $g(\theta) = \eta[\mu(\theta)] = z(\theta)$. As exposed before, parametric models such as the one in (4.8) may not be flexible enough to model complex relationships between the variables. Thus, in Section 4.2, a nonparametric approach for the joint estimation of $g(\cdot)$ and $\psi(\cdot)$ will be considered.

4.2 Joint estimation via circular local likelihood

Double exponential nonparametric models have been studied in the euclidean context, estimating the mean and dispersion functions by making use of P-splines, constituting a generalization of GLM or GAM models (Gijbels et al., 2010; Gijbels and Prosdocimi, 2011, 2012). This approach has also been extended for robust regression (Croux et al., 2012) and variable selection (Antoniadis et al., 2016). The objective of this section is to jointly estimate the mean and dispersion functions in a nonparametric way, but using the circular local likelihood approach described in Chapter 3.

For the local kernel weighted likelihood approach, the sine-polynomial expansion in (3.1) is applied to both $g(\cdot)$ and $\psi(\cdot)$. Assuming that $g(\cdot)$ and $\psi(\cdot)$ are p times continously differ-

entiable, for a fixed $\theta_0 \in \mathbb{T}$ and for data points in a neighborhood of θ_0, it holds that

$$g(\Theta_i) \approx g(\theta_0) + g'(\theta_0)\sin(\Theta_i - \theta_0) + ... + \frac{g^{(p)}(\theta_0)}{p!}\sin^p(\Theta_i - \theta_0) \approx \boldsymbol{\Theta}_i^T \boldsymbol{\beta}_\mu, \tag{4.9}$$

$$\psi(\Theta_i) \approx \psi(\theta_0) + \psi'(\theta_0)\sin(\Theta_i - \theta_0) + ... + \frac{\psi^{(p)}(\theta_0)}{p!}\sin^p(\Theta_i - \theta_0) \approx \boldsymbol{\Theta}_i^T \boldsymbol{\beta}_\gamma, \tag{4.10}$$

where, as denoted in previous chapters, $\boldsymbol{\beta}_\mu = (\beta_{\mu,0}, \ldots, \beta_{\mu,p})^T$, $\boldsymbol{\beta}_\gamma = (\beta_{\gamma,0}, \ldots, \beta_{\gamma,p})^T$ and $\boldsymbol{\Theta}_i = [1, \sin(\Theta_i - \theta_0), \ldots, \sin^p(\Theta_i - \theta_0)]^T$. Therefore, employing the approximated density in (4.6) yields that the circular local log-likelihood, for a given $\gamma(\cdot)$, can be written as

$$\mathcal{L}_p(\boldsymbol{\beta}_\mu; \kappa, \theta_0) = -\frac{1}{2}\sum_{i=1}^{n}\left[\frac{1}{\gamma(\Theta_i)} D(Y_i, \boldsymbol{\Theta}_i^T \boldsymbol{\beta}_\mu)\right] K_\kappa(\Theta_i - \theta_0), \tag{4.11}$$

with $K_\kappa(\cdot)$ being a circular kernel and κ a concentration parameter controlling the smoothing of the estimator of $g(\cdot)$. On the other hand, for a given $z(\cdot)$, the circular local log-likelihood is

$$\mathcal{L}_p(\boldsymbol{\beta}_\gamma; \nu, \theta_0) = -\frac{1}{2}\sum_{i=1}^{n}\left\{\log[h(\boldsymbol{\Theta}_i^T \boldsymbol{\beta}_\gamma)] + \frac{1}{h(\boldsymbol{\Theta}_i^T \boldsymbol{\beta}_\gamma)} D[Y_i, z(\Theta_i)]\right\} K_\nu(\Theta_i - \theta_0), \tag{4.12}$$

with $K_\nu(\cdot)$ being a circular kernel and ν being the concentration parameter controlling the smoothing of the estimator of $\psi(\cdot)$.

Maximization is carried out through a two-step iterative method, such as Newton-Raphson or Fisher scoring algorithms. Starting with initial values $\boldsymbol{\beta}_\mu^{(0)}$ and $\boldsymbol{\beta}_\gamma^{(0)}$, the maximization over $\boldsymbol{\beta}_\mu$ is first performed and, afterwards, the maximization over $\boldsymbol{\beta}_\gamma$ (with different degrees of smoothing) is carried out. By taking into account the deviance expression in (4.7), for the first step, (4.11) can be written as

$$\mathcal{L}_p(\boldsymbol{\beta}_\mu; \kappa, \theta_0) = \sum_{i=1}^{n}\left\{\frac{1}{\varphi\gamma(\Theta_i)}[Y_i \boldsymbol{\Theta}_i^T \boldsymbol{\beta}_\mu - b(\boldsymbol{\Theta}_i^T \boldsymbol{\beta}_\mu)]\right\} K_\kappa(\Theta_i - \theta_0).$$

The gradient vector is

$$\boldsymbol{v}_\mu = \frac{\partial}{\partial \boldsymbol{\beta}_\mu}\mathcal{L}_p(\boldsymbol{\beta}_\mu; \kappa, \theta_0) = \sum_{i=1}^{n}\left\{\boldsymbol{\Theta}_i \frac{1}{\varphi\gamma(\Theta_i)}[Y_i - b'(\boldsymbol{\Theta}_i^T \boldsymbol{\beta}_\mu)]\right\} K_\kappa(\Theta_i - \theta_0)$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

and the Hessian matrix is given by

$$\boldsymbol{H}_\mu = \frac{\partial^2}{\partial \boldsymbol{\beta}_\mu \partial \boldsymbol{\beta}_\mu^T}\mathcal{L}_p(\boldsymbol{\beta}_\mu; \kappa, \theta_0) = -\sum_{i=1}^{n}\left[\boldsymbol{\Theta}_i \boldsymbol{\Theta}_i^T \frac{1}{\varphi\gamma(\Theta_i)} b''(\boldsymbol{\Theta}_i^T \boldsymbol{\beta}_\mu)\right] K_\kappa(\Theta_i - \theta_0).$$

Now, the Fisher scoring matrix is the conditional expectation of the observed information

matrix

$$\boldsymbol{F}_\mu = -\mathbb{E}(\boldsymbol{H}_\mu|\Theta_1,\ldots,\Theta_n) = -\boldsymbol{H}_\mu$$

and, hence, the Newton-Raphson algorithm and the Fisher scoring method are equivalent with the updating rule $\boldsymbol{\beta}_\mu^{(l+1)} = \boldsymbol{\beta}_\mu^{(l)} + \boldsymbol{F}_\mu^{-1}\boldsymbol{v}_\mu$. The matrix $\boldsymbol{F}_\mu$ can be expressed as $\boldsymbol{F}_\mu = \boldsymbol{\Theta}^T\boldsymbol{W}_\mu\boldsymbol{\Theta}$ where

$$\boldsymbol{\Theta} = \begin{pmatrix} 1 & \sin(\Theta_1-\theta_0) & \ldots & \sin^p(\Theta_1-\theta_0) \\ \vdots & & & \\ 1 & \sin(\Theta_n-\theta_0) & \ldots & \sin^p(\Theta_n-\theta_0) \end{pmatrix},$$

and with $\boldsymbol{W}_\mu$ being the diagonal matrix with elements $b''(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\mu)[\varphi\gamma(\Theta_i)]^{-1}K_\kappa(\Theta_i-\theta_0)$.

For the second step, (4.12) is maximized over $\boldsymbol{\beta}_\gamma$. The gradient vector is given by

$$\boldsymbol{v}_\gamma = \frac{\partial}{\partial\boldsymbol{\beta}_\gamma}\mathcal{L}_p(\boldsymbol{\beta}_\gamma;\nu,\theta_0) = -\frac{1}{2}\sum_{i=1}^n \boldsymbol{\Theta}_i\left\{\frac{h'(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)}{h(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)} - \frac{h'(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)}{h^2(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)}D[Y_i,z(\Theta_i)]\right\}K_\nu(\Theta_i-\theta_0).$$

Now, the expression for the Hessian matrix is

$$\begin{aligned}\boldsymbol{H}_\gamma = \frac{\partial^2}{\partial\boldsymbol{\beta}_\gamma\partial\boldsymbol{\beta}_\gamma^T}\mathcal{L}_p(\boldsymbol{\beta}_\gamma;\nu,\theta_0) = -\frac{1}{2}\sum_{i=1}^n \boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T\Bigg\{&\frac{h''(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)h(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma) - [h'(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)]^2}{h^2(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)} \\ &- \frac{h''(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)h^2(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma) - 2[h'(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)]^2h(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)}{h^4(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)}D[Y_i,z(\Theta_i)]\Bigg\}K_\nu(\Theta_i-\theta_0).\end{aligned}$$

Unlike in the maximization rule for the first step, this time the random responses appear in the expression of the Hessian matrix and, consequently, the Newton-Raphson algorithm and the Fisher scoring method are not equivalent. For the Newton-Raphson algorithm, the updating rule is $\beta_\gamma^{(l+1)} = \beta_\gamma^{(l)} - \boldsymbol{H}_\gamma^{-1}\boldsymbol{v}_\gamma$, where the Hessian matrix can be expressed as $\boldsymbol{H}_\gamma = \boldsymbol{\Theta}^T\boldsymbol{W}_\gamma\boldsymbol{\Theta}$ where $\boldsymbol{W}_\gamma$ is a diagonal matrix with elements

$$\begin{aligned}-\frac{1}{2}\Bigg\{&\frac{h''(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)h(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma) - [h'(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)]^2}{h^2(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)} \\ &- \frac{h''(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)h^2(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma) - 2[h'(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)]^2h(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)}{h^4(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)}D[Y_i,z(\Theta_i)]\Bigg\}K_\nu(\Theta_i-\theta_0).\end{aligned}$$

On the other hand, the Fisher scoring matrix is given by

$$\boldsymbol{F}_\gamma = -\mathbb{E}(\boldsymbol{H}_\gamma|\Theta_1,\ldots,\Theta_n) = \boldsymbol{\Theta}^T\tilde{\boldsymbol{W}}_\gamma\boldsymbol{\Theta},$$

where $\tilde{\boldsymbol{W}}_\gamma$ is a diagonal matrix with elements

$$\frac{1}{2}\left[\frac{h'(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)}{h(\boldsymbol{\Theta}_i^T\boldsymbol{\beta}_\gamma)}\right]^2 K_\nu(\Theta_i-\theta_0).$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

In this chapter, the Fisher scoring method was employed, and the maximization procedure is summarized in Algorithm 4.1.

Algorithm 4.1: Joint estimation of mean and dispersion functions

Initialization: Select starting values $\boldsymbol{\beta}_\mu^{(0)}$ and $\boldsymbol{\beta}_\gamma^{(0)}$. The initial values $\mu^{(0)} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and $\gamma^{(0)} = \varphi$ are used.

1. For a given $\gamma^{(l)}(\cdot)$, update $\boldsymbol{\beta}_\mu^{(l+1)} = \boldsymbol{\beta}_\mu^{(l)} + \boldsymbol{F}_\mu^{-1}\boldsymbol{v}_\mu$, where $\boldsymbol{v}_\mu$ and $\boldsymbol{F}_\mu$ are constructed with $\boldsymbol{\beta}_\mu = \boldsymbol{\beta}_\mu^{(l)}$. Get an estimate $\mu^{(l+1)}(\cdot)$ for $\mu(\cdot)$.

2. Using $\mu^{(l+1)}(\cdot)$, update $\boldsymbol{\beta}_\gamma^{(l+1)} = \boldsymbol{\beta}_\gamma^{(l)} + \boldsymbol{F}_\gamma^{-1}\boldsymbol{v}_\gamma$, where $\boldsymbol{v}_\gamma$ and $\boldsymbol{F}_\gamma$ are constructed with $\boldsymbol{\beta}_\gamma = \boldsymbol{\beta}_\gamma^{(l)}$. Obtain an estimate $\gamma^{(l+1)}(\cdot)$

Stopping criterion: $1-\varepsilon < \max\left(\left|\frac{\mu^{(l+1)}(\cdot)}{\mu^{(l)}(\cdot)}\right|, \left|\frac{\gamma^{(l+1)}(\cdot)}{\gamma^{(l)}(\cdot)}\right|\right) < 1+\varepsilon$, for a fixed $\varepsilon > 0$.

4.3 Simulation study

In this section, the performance of the estimation method presented in Section 4.2 is empirically assessed, in the particular case of count data. The simulation scenarios are described in Section 4.3.1, while the results are analyzed in Section 4.3.2. Additionally, Section 4.3.3 introduces a graphical representation for the joint visualization of both estimated functions.

4.3.1 Simulation scenarios

A description of the simulation scenarios is given in what follows. Data from a double Poisson model are simulated, where the mean and dispersion parameters depend on a circular covariate Θ (drawn from a circular uniform distribution) through the functions $\mu(\theta) = \exp[g(\theta)]$ and $\gamma(\theta) = \exp[\psi(\theta)]$. The mean and dispersion functions are given in Table 4.1 and depicted in Figure 4.1, where realizations of the data of size $n = 200$ are shown along the mean functions.

For model M1, $\psi(\theta) > 0 \ \forall\ \theta \in \mathbb{T}$, which implies $\gamma(\theta) > 1 \ \forall\ \theta \in \mathbb{T}$, *i.e.*, there is over-dispersion with respect to the Poisson model for all values of the predictor, although the amount of dispersion varies with the covariate. Contrarily, for model M2 $\psi(\theta) < 0 \ \forall\ \theta \in \mathbb{T}$, implying under-dispersion with respect to the Poisson. Lastly, for model M3, there are both over- and under-dispersion, depending on the value of Θ. For all models, sample sizes $n = 100, 200, 500$ are considered and 500 replications of the data are used. The quality of the estimators is obtained by approximating the Integrated Squared Error (ISE) of the mean and

Model	Model elements
M1	$g(\theta) = 3.75 + 0.05\sin\theta - 0.1\cos\theta + 0.07\sin(\theta) - 0.2\cos(2\theta) - 0.15\sin(3\theta)$ $\psi(\theta) = 1.5 - 0.4\sin\theta - 0.45\cos\theta$
M2	$g(\theta) = 15\exp[-2 + \sin\theta\cos^3\theta - 0.25\exp(\sin^3\theta)]$ $\psi(\theta) = -2.35 - 1.1\sin\theta + \cos\theta - 0.5\sin(2\theta) + 0.55\cos(2\theta)$
M3	$g(\theta) = 3 + \cos^3\theta/4 + \cos^4\theta\sin\theta\exp[\cos(\theta/2)]$ $\psi(\theta) = -0.85 + \exp[0.3\cos\theta + 0.05\sin(2\theta)] - 0.15\cos(2\theta)$

Table 4.1: Description of the simulated models.

dispersion functions $\mu(\cdot)$ and $\gamma(\cdot)$:

$$\text{ISE}[\hat{\mu}(\cdot)] = \frac{\int_0^{2\pi}[\hat{\mu}_{(b)}(\theta) - \mu(\theta)]^2 d\theta}{\int_0^{2\pi}\mu(\theta)^2 d\theta}, \quad \text{ISE}[\hat{\gamma}(\cdot)] = \frac{\int_0^{2\pi}[\hat{\gamma}_{(b)}(\theta) - \gamma(\theta)]^2 d\theta}{\int_0^{2\pi}\gamma(\theta)^2 d\theta}, \tag{4.13}$$

where $\hat{\mu}_{(b)}(\theta)$ represents the estimator of $\mu(\theta)$ for the bth replication of the data (respectively, $\hat{\gamma}_{(b)}(\theta)$ represents the estimator of $\gamma(\theta)$ for the bth replication of the data). The integrals in (4.13) are approximated numerically by Simpson's rule.

For the estimation of $\mu(\cdot)$ and $\gamma(\cdot)$, local sine-polynomials, with $p = 1$ in (4.9) and (4.10), are used. In addition, the smoothing parameters, κ and ν, must be selected. A simple two-step cross-validation approach is employed in this section: first, κ is selected by assuming that $\gamma(\theta) = 1\ \forall\ \theta \in \mathbb{T}$ and maximizing

$$\sum_{i=1}^{n}\{Y_i \log[\hat{\mu}_{-i}(\Theta_i)] - \hat{\mu}_{-i}(\Theta_i)\},$$

where $\hat{\mu}_{-i}(\cdot)$ denotes the estimator of $\mu(\cdot)$ constructed without the ith observation and computed with smoothing parameter κ. Next, ν is selected by maximizing

$$-\frac{1}{2}\sum_{i=1}^{n}\left(\hat{\gamma}_{-i}(\Theta_i) + \frac{D\{Y_i, \log[\hat{\mu}(\Theta_i)]\}}{\exp[\hat{\gamma}_{-i}(\Theta_i)]}\right),$$

where $D(\cdot,\cdot)$ is the deviance defined in (4.7), $\hat{\gamma}_{-i}(\cdot)$ denotes the estimator of $\gamma(\cdot)$ constructed without the ith observation and computed with smoothing parameter ν and $\hat{\mu}(\cdot)$ is estimated using parameter κ.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.3.2 Simulation results

Table 4.2 contains the averages of the numerically approximated ISE of the mean and dispersion functions, for the three models and all the considered sample sizes. In all cases, the

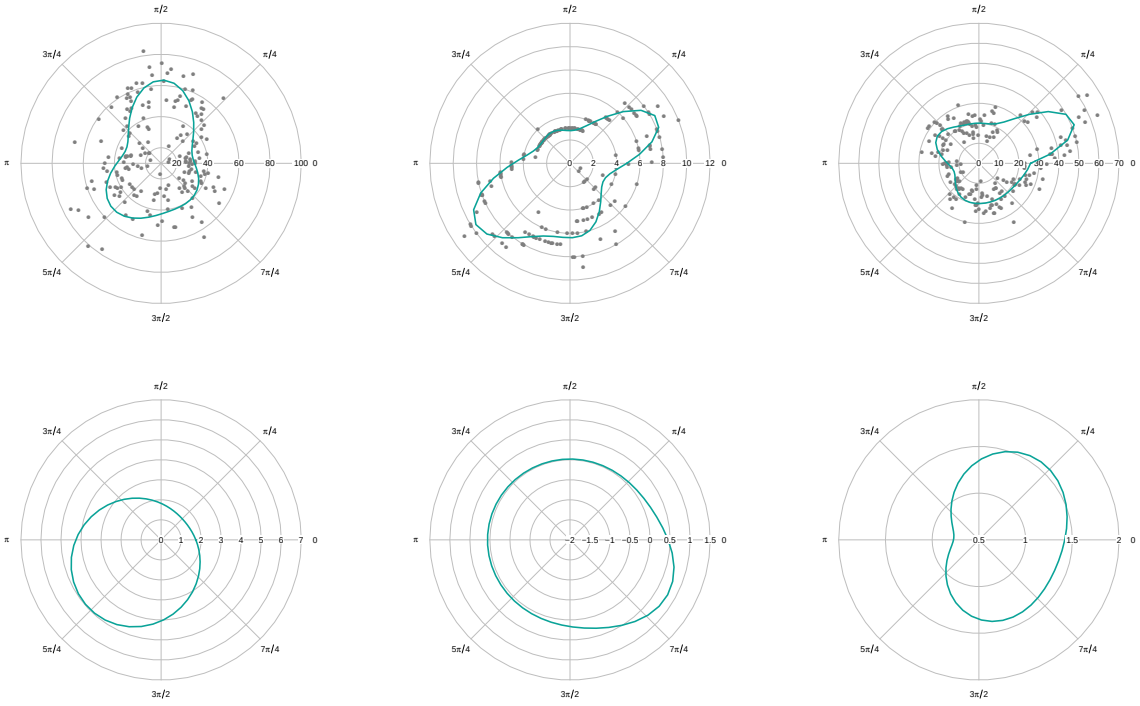


Figure 4.1: Top row: radial representations of the mean functions $\mu(\cdot)$ of models M1 (left), M2 (center) and M3 (right), with realizations of the data with $n = 200$. Bottom row: representations of the dispersion functions $\gamma(\cdot)$ of models M1 (left), M2 (center) and M3 (right).

results concerning the estimation of the mean function are of the order of 10^{-3}, and the average approximated ISE gets smaller as the sample size increases. Regarding the estimation of the dispersion function, the average approximated ISE is of the order of 10^{-2} for models M1 and M3, and 10^{-1} for M2. This is not surprising as the estimation of the dispersion is always more difficult than the estimation of the mean (see Gijbels et al., 2010). However, in all cases the errors decrease when increasing the sample size.

	Mean function $\mu(\cdot)$			Dispersion function $\gamma(\cdot)$		
n	M1	M2	M3	M1	M2	M3
100	$7.02 \cdot 10^{-3}$	$5.82 \cdot 10^{-3}$	$9.70 \cdot 10^{-3}$	$9.93 \cdot 10^{-2}$	$3.03 \cdot 10^{-1}$	$9.01 \cdot 10^{-2}$
200	$4.01 \cdot 10^{-3}$	$3.28 \cdot 10^{-3}$	$5.43 \cdot 10^{-3}$	$5.06 \cdot 10^{-2}$	$2.61 \cdot 10^{-1}$	$4.94 \cdot 10^{-2}$
500	$1.80 \cdot 10^{-3}$	$1.46 \cdot 10^{-3}$	$2.44 \cdot 10^{-3}$	$2.32 \cdot 10^{-2}$	$2.30 \cdot 10^{-1}$	$2.26 \cdot 10^{-2}$

Table 4.2: Averages of the numerically approximated ISE of the estimated mean and dispersion functions across the 500 replications.

In order to graphically ascertain the quality of the estimators, the top row of Figure 4.2 presents radial representations of the true mean functions and representatives of the 500 esti-

mators when $n = 200$. In particular, the estimators corresponding to the 5th, 50th and 95th percentiles of the approximated ISE are shown. It can be observed that the estimation of the mean is very satisfactory, specially for model M2, where even the estimator corresponding to the 95th percentile of the ISE performs quite well.

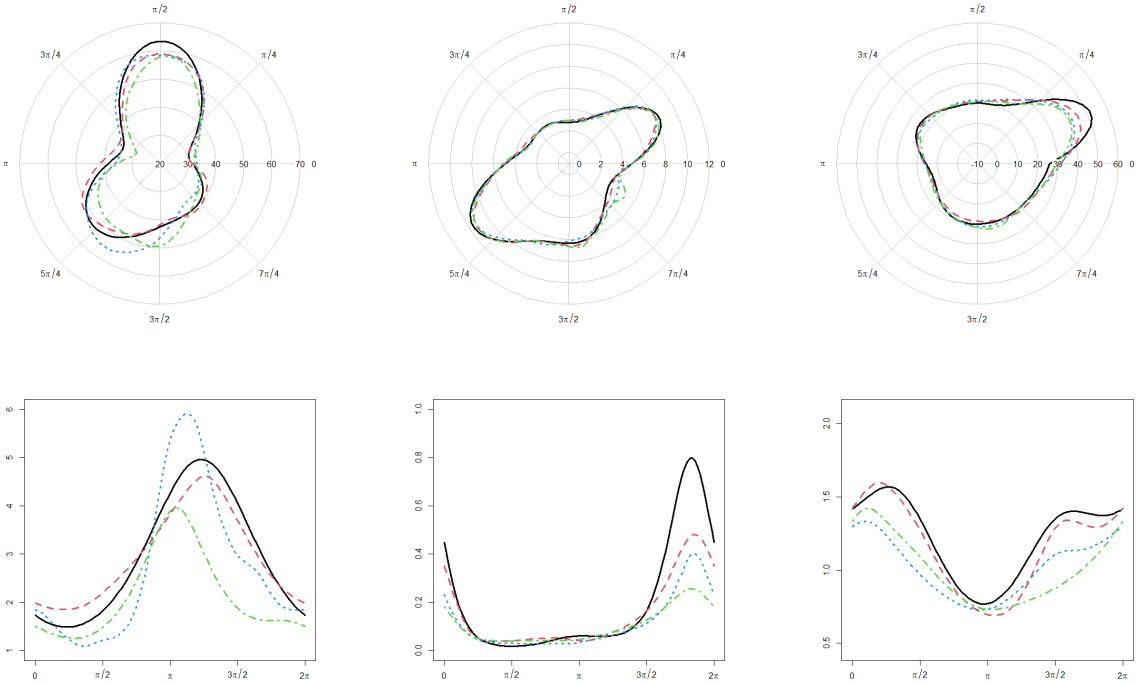


Figure 4.2: Top row: radial representations of the true mean functions $\mu(\cdot)$ (continuous line) of models M1 (left), M2 (center) and M3 (right), with representative estimates when $n = 200$ (5th ISE percentile with dashed line, 50th ISE percentile with dotted line and 95th ISE percentile with dotted-dashed line). Bottom row: same with the dispersion functions $\gamma(\cdot)$, but with a planar representation.

The true dispersion functions and their representative estimates for $n = 200$ are shown in the bottom row of Figure 4.2. The dispersion functions are shown in a planar representation to facilitate the comparison of the true function and the representative estimates. As mentioned above, the estimation of the dispersion is more difficult than the one of the mean, but all the representatives are able to capture the general shape of the true function. It is also observed that the higher approximated ISE results for model M2 in Table 4.2 are due to the fact that the estimators are not able to capture the height of the curve at the peak.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.3.3 Joint visualization of mean and dispersion functions

An important issue after the joint estimation of the mean and dispersion functions is the joint visualization of the estimated mean and dispersion. In this section, a way of visualizing both

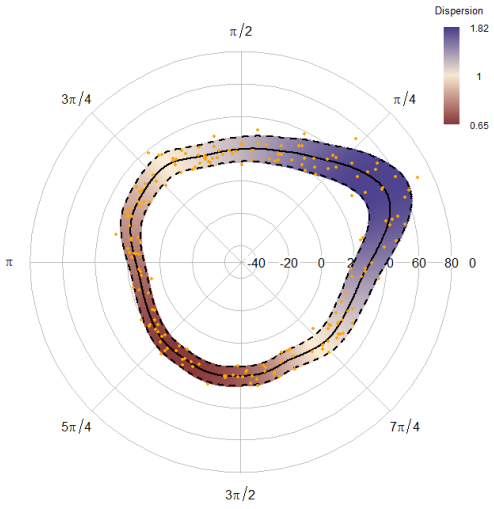


Figure 4.3: Radial representation of simulated data from model M3 with $\hat{\mu}(\theta) \pm 1.5 \cdot \widehat{SD}(\theta)$. *The colors of the band indicate the estimated dispersion, as indicated in the legend.*

functions is introduced, facilitating the detection of values of the predictor variable for which there is over- or under-dispersion. Figure 4.3 shows simulated data from model M3 with the estimated mean function $\hat{\mu}(\theta) = \exp\{\hat{g}(\theta)\}$ (continuous line) and $\pm 1.5 \cdot \widehat{\text{SD}}(\theta)$ (dashed lines), where $\widehat{\text{SD}}(\theta)$ is the estimated standard deviation function computed as $\widehat{\text{SD}}(\theta) = [\hat{\mu}(\theta)\hat{\gamma}(\theta)]^{1/2}$. This is not a band for inferential purposes, but just a representation of the variability of the data. Note that the factor 1.5 is present just with the objective of a better visualization. The colors inside the band indicate the dispersion, with darker shades of purple representing a stronger over-dispersion with respect to the Poisson distribution, and darker shades of brown indicating a stronger under-dispersion compared to the Poisson model.

4.4 Analysis of the number of spikes in a macaque's neuron

In this section, the spike count data described in Appendix A.5 is analyzed. The nonparametric double exponential method is considered to model the relationship between the variables, and its estimation is performed through the maximization of the circular local likelihood as described in Section 4.2. This approach is compared to the COM-Poisson regression model (see Section 4.1.1). To estimate the parameters, the efficient estimation procedure given by Chatla and Shmueli (2018) is used, but accounting for the circular nature of the predictor variable by modeling the parameters as a function of the tuning direction as in equation (4.1). The logarithm function is selected as the link functions $\eta(\cdot)$ and $h(\cdot)$ and, following the work of Stevenson (2016), the values $k = 2$ and $\tilde{k} = 1$ are employed in (4.1). Once the functions $\lambda(\theta)$ and $\zeta(\theta)$ are estimated, the mean function is estimated by using the approximation in

(4.2), leading to

$$\hat{\mu}(\theta) = \hat{\lambda}(\theta)^{1/\hat{\zeta}(\theta)} - \frac{\hat{\zeta}(\theta) - 1}{2\hat{\zeta}(\theta)}. \tag{4.14}$$

The dispersion function is obtained by computing the ratio between the estimated variance and mean function, and by employing the approximation in (4.3), which yields that, for the COM-Poisson model, the estimated dispersion is given by

$$\hat{\gamma}(\theta) = \frac{1}{\hat{\mu}(\theta)} \frac{\hat{\lambda}(\theta)^{1/\hat{\zeta}(\theta)}}{\hat{\zeta}(\theta)},$$

where $\hat{\mu}(\theta)$ is given in (4.14).

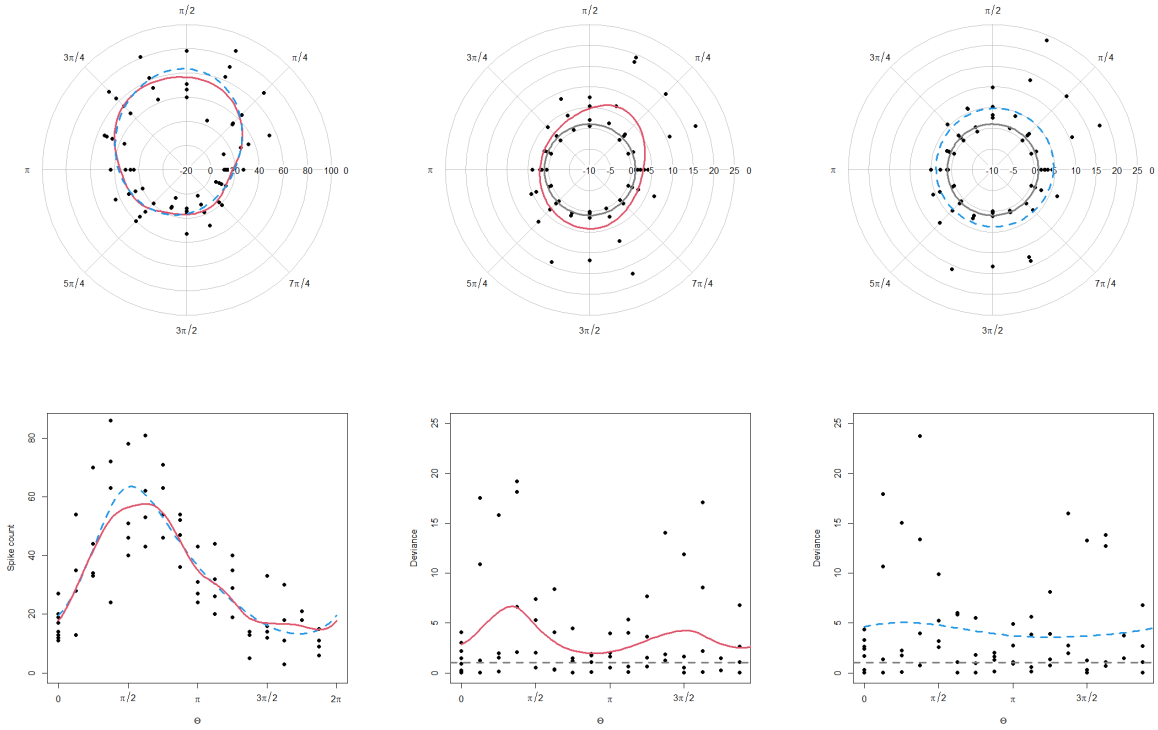


Figure 4.4: Left column: radial and planar representations of the spike count data with the estimated mean function with the nonparametric double exponential model (continuous line) and the COM-Poisson model with (dashed line). Center column: polar and planar representations of the deviance residuals obtained with the nonparametric double exponential model along with the estimated dispersion function. Right column: polar and planar representations of the deviance residuals obtained with the COM-Poisson model, along with the estimated dispersion function.

For the nonparametric double exponential model, by using the two-step cross-validation method described in Section 4.3.1, the obtained concentration parameters are $\kappa = 7.4$ for the estimation of the mean and $\nu = 4.5$ for the estimation of the dispersion functions. The

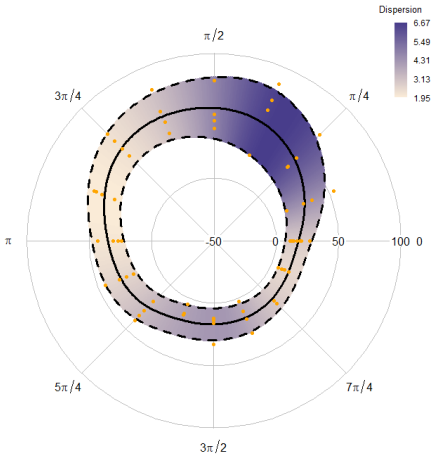


Figure 4.5: Radial representation of the spike count data with $\hat{\mu}(\theta) \pm 1.5 \cdot \widehat{SD}(\theta)$. *The colors of the band indicate the estimated dispersion, as indicated in the legend.*

nonparametric estimator of the mean is displayed as a continuous line in the left column of Figure 4.4, while the mean function estimated by employing the COM-Poisson model is shown in the same panels as a dashed line. Both estimators are fairly similar, showing that the highest number of spikes per trial is obtained for directions around $\pi/2$. Note that although the COM-Poisson estimator is obtained under a parametric framework, the mean of the COM-Poisson is given by a complex function of the two parameters of the distribution (see equation (4.2)).

Regarding the estimation of the dispersion, the nonparametric estimator is represented in the center panels of Figure 4.4, along with the deviance residuals obtained with the estimated mean. It was obtained that the estimated curve is always above 1, indicating over-dispersion for all the stimulus directions. However, it can be clearly seen that the dispersion changes with the stimulus direction, obtaining large values of the dispersion for tuning directions around $3\pi/8$ and $3\pi/2$. The deviance residuals of the COM-Poisson model and the estimated dispersion function are represented in the right panels of Figure 4.4. As it can be observed, the estimated curve is also greater than 1 for all the tuning directions, but it is very close to a horizontal line, indicating almost constant dispersion.

Finally, Figure 4.5 shows the joint illustration of the estimated mean and dispersion using the nonparametric double exponential model, with the estimated mean function, $\hat{\mu}(\theta) = \exp[\hat{g}(\theta)]$ and $\pm 1.5 \cdot \widehat{\mathrm{SD}}(\theta)$ indicating the variability of the data. It is clearly shown that, although there is over-dispersion for all tuning directions, for directions between $\pi/4$ and $\pi/2$ there is a strong over-dispersion with respect to the Poisson model.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 5

Multimodal circular regression

In the previous chapters, different nonparametric methods to estimate regression functions in the context of circular data were considered. In Chapter 2, estimators of the conditional mean were studied, while Chapters 3 and 4 focus on the estimation of general conditional characteristics, once the distribution of the response has been specified. However, these methods might not be suitable in situations where the conditional density is multimodal.

In order to explain why the previously studied methods might fail with multimodal data, consider, as a toy example, the classical periwinkle data set given in Fisher and Lee (1992), consisting of measurements of distance (in centimeters) and directions moved by small blue periwinkles after being relocated. The interest lies on studying the influence of the distance (real-valued predictor variable) on the escape direction (circular response variable). In the left panel of Figure 5.1 the data set was represented in a horizontal cylinder, where the height of the cylinder represents the magnitude of the distance. For clarity purposes, the right panel of Figure 5.1 shows a planar representation of the same data set. The curve in continuous trace represents a nonparametric estimation of the regression function, regarded as the expected circular mean direction given the covariate, exposed in Section 2.1.2. It can be seen that, for small distances, this estimation is located in a region without data, where there seems to exist two different groups or two different preferred directions.

A possible approach to model this data set, related to the methodology in Chapters 3 and 4, could be the selection of a multimodal conditional distribution for the escape direction, and the nonparametric estimation of the parameters determining the local modes through the maximization of the local likelihood. Nevertheless, the general approach for the obtainment of a multimodal distribution is the construction of mixtures, which would lead to a circular mixture of regressions. Nonparametric mixtures of regressions have been studied in the real-valued context by Huang et al. (2013) and Xiang (2018), among others. However, the simultaneous selection of the number of components and the smoothing parameter is not an easy task. In addition, situations such as the one in Figure 5.1, where it seems that the support of one of the regression *branches* or components is different from the other, is difficult to handle.

UCC
DEPARTAMENTO

Hence, the approach taken in this chapter is a fully nonparametric method to estimate the conditional local modes, known as multimodal regression. The idea of estimating the local modes of the conditional density in the euclidean context was first introduced by Scott (1992), who obtained the conditional modes by computing a linear blend of the bivariate averaged shifted histogram estimate. Einbeck and Tutz (2006) proposed using a conditional

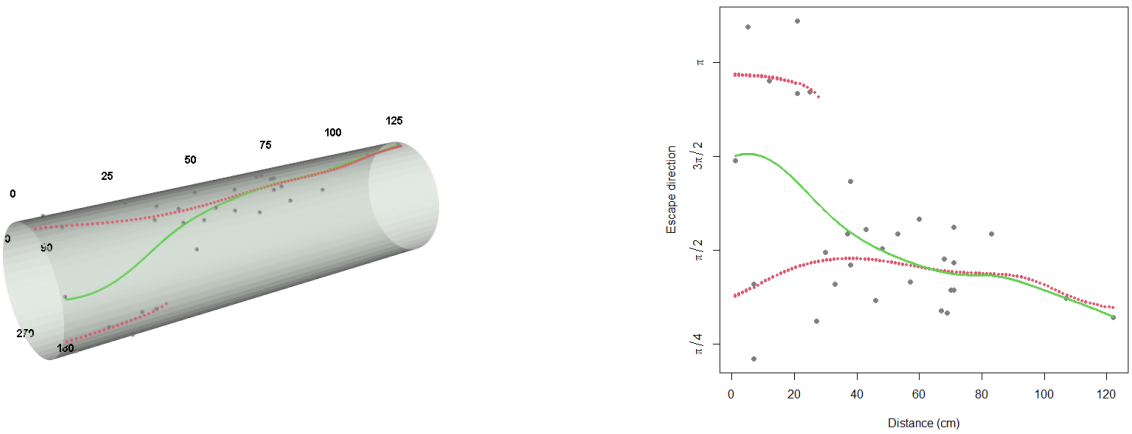


Figure 5.1: Representations on cylinder (left) and on the plane (right) of the periwinkle data with the kernel estimator of the mean regression function (continuous trace) and the proposed multimodal estimator (dotted trace).

version of the mean shift algorithm (Fukunaga and Hostetler, 1975; Cheng, 1995; Comaniciu and Meer, 2002) to estimate the regression multifunction based on a kernel-type conditional density estimator. The theoretical properties of this approach were studied by Chen et al. (2016a) and extended to measurement error problems by Zhou and Huang (2016). Recent reviews of multimodal regression for euclidean variables can be found in Chen (2018) and Alonso-Pena (2020).

The mean shift algorithm was extended to directional variables by Oba et al. (2005) in the context of nonparametric modal clustering and it was also studied by Kobayashi and Otsu (2010) in the same setting. The convergence of the algorithm and some other theoretical results were recently obtained by Zhang and Chen (2021a,c) and the extension to directional product spaces was studied by Zhang and Chen (2021b). However, modal regression for circular variables, where the circular nature may characterize the response, the covariate or both, has not been addressed in the literature.

The goal of this chapter is to introduce statistical tools for analyzing relationships between variables in the circular regression settings, by taking into account the multimodal structure of the data. In addition, an important objective of this chapter is to apply the proposed methodology to the analysis of escape responses in the zoological field. In particular, a study of the zebrafish data set, described in Appendix A.7, will be carried out by employing the methods studied in this chapter.

USC DE SANTIAGO DE COMPOSTELA

The organization of the chapter is as follows: Section 5.1 presents the regression scenarios and details the algorithms that allow to estimate the circular regression multifunctions. Section 5.2 deals with the error measurement in the circular multimodal regression context and gives some theoretical results about the asymptotic behavior of the estimators, while the

problem of bandwidth/concentration selection is discussed in Section 5.3. The finite sample performance of the estimators is studied in Section 5.4 through simulation experiments and Section 5.5 contains a complete analysis of the zebrafish data (see Appendix A.7) with the proposed methodology. The main contents of this chapter are collected in Alonso-Pena and Crujeiras (2022).

Contents

5.1 On the use of mean shift for circular modal regression 91
5.1.1 Real-valued response 92
5.1.2 Circular response 94
5.2 Some theoretical results 97
5.3 Smoothing parameter selection 100
5.3.1 Modal cross-validation for circular regression 102
5.4 Simulation study 102
5.4.1 Simulation scenarios 103
5.4.2 Simulation results 103
5.4.3 The unknown boundary condition 107
5.5 Analysis of the escape responses in larval zebrafish 110

5.1 On the use of mean shift for circular modal regression

Nonparametric multimodal regression aims for estimating the conditional density of a response over a covariate with a flexible nonparametric estimator and, afterwards, compute the conditional local modes with the so-called mean shift algorithm (Fukunaga and Hostetler, 1975; Cheng, 1995; Comaniciu and Meer, 2002). In this section, details of the algorithms designed for estimating the multimodal regression multifunctions in the different contexts involving circular variables exposed in Section 1.2.3 (circular-linear, linear-circular and circular-circular) will be given. The case of a real-valued response and a circular predictor will be contemplated first, since the methodology in this case is closer to the real-valued scenario (Einbeck and Tutz, 2006; Chen et al., 2016a) and is a good starting point for understanding the proposal in the circular response setting.

Consequently, two scenarios are considered next: the case where the response is a real-valued variable, exposed in Section 5.1.1, where the usual conditional mean shift algorithm with suitably adapted weights will be used, and the cases where the response is circular, contained in Section 5.1.2, where the circular conditional mean shift is proposed and applied to the regression context.

5.1.1 Real-valued response

Consider a circular predictor, Θ, with support on the circumference of the unit circle, $\mathbb{T}$, and a real-valued response, denoted by Y, with support on $\mathbb{R}$. Let $\{(\Theta_i, Y_i)\}_{i=1}^n$ be a bivariate sample from (Θ, Y) with conditional density function $f(y|\theta)$ and with joint density $f(\theta, y)$ satisfying assumption **(A1)** in Appendix C.3. In order to model the relationship between the predictor and the response, the modal regression multifunction (Einbeck and Tutz, 2006) is considered. For each $\theta \in \mathbb{T}$, the regression multifunction is defined as the set of local modes (or local maxima) of the conditional density function:

$$M(\theta) = \left\{ y \in \mathbb{R} : \frac{\partial}{\partial y} f(y|\theta) = 0, \frac{\partial^2}{\partial y^2} f(y|\theta) < 0 \right\}. \tag{5.1}$$

The estimation of the set $M(\theta)$ in (5.1) is carried out through an indirect approach: first, the conditional density is estimated and, next, the conditional local modes are computed. For the estimation of the conditional density of the response over the covariate, a kernel type estimator will be employed:

$$\hat{f}(y|\theta) = \frac{\sum_{i=1}^n K_\kappa(\Theta_i - \theta) L_h(Y_i - y)}{\sum_{i=1}^n K_\kappa(\Theta_i - \theta)}, \tag{5.2}$$

where $L_h(\cdot) = \frac{1}{h} L(\cdot/h)$ is a *linear* kernel with bandwidth h where $L(\cdot)$ satisfies condition **(A3)** in Appendix C.3, and, as in Chapter 3, $K_\kappa(\cdot)$ is a circular kernel with concentration κ of the form

$$K_\kappa(\cdot) = c_\kappa K\{\kappa[1 - \cos(\cdot)]\}, \tag{5.3}$$

where c_κ is a normalizing constant depending on κ and $K(\cdot)$ is a function satisfying assumption **(A4-1)** in Appendix C.3. Recall that κ plays a role opposite to h in the smoothing of the kernel density estimator: if κ is small, an oversmoothed estimator is obtained and, if the value of κ is large, the kernel density estimator will be undersmooothed. Note that the numerator in (5.2) is proportional to the product kernel estimator of the joint density in the cylinder (García-Portugués et al., 2013) and the denominator is proportional to the circular kernel density estimator in (1.10). Thus, the kernel density estimator of the conditional density depends on two smoothing parameters, κ and h which, as it will be discussed later, play a different role in the estimation of the regression multifunction.

In order to estimate the regression multifunction, the expression in (5.2) is plugged into (5.1), obtaining

$$\widehat{M}(\theta) = \left\{ y \in \mathbb{R} : \frac{\partial}{\partial y} \hat{f}(y|\theta) = 0, \frac{\partial^2}{\partial y^2} \hat{f}(y|\theta) < 0 \right\}. \tag{5.4}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The computation of the conditional local modes is not straightforward and, thus, the conditional mean shift algorithm is used. It will be assumed that $L_h(\cdot)$ is a radially symmetric kernel with profile $l(\cdot)$, so that $L(\cdot) = c_L \times l\left[(\cdot)^2\right]$, with $c_L \in \mathbb{R}^+$. A local maximum of

$\hat{f}(y|\theta)$ must satisfy

$$\frac{\partial}{\partial y}\hat{f}(y|\theta) = \frac{\frac{2c_L}{h^3}\sum_{i=1}^{n} K_\kappa\left(\Theta_i - \theta\right) l'\left[\left(\frac{Y_i - y}{h}\right)^2\right](Y_i - y)}{\sum_{i=1}^{n} K_\kappa(\Theta_i - \theta)} = 0.$$

Taking $g(\cdot) = -l'(\cdot)$ yields

$$\frac{\partial}{\partial y}\hat{f}(y|\theta) = \frac{-2c_L}{nh^3\hat{f}(\theta)}\sum_{i=1}^{n} K_\kappa\left(\Theta_i - \theta\right) g\left[\left(\frac{Y_i - y}{h}\right)^2\right](Y_i - y) = 0, \tag{5.5}$$

where $\hat{f}(\theta) = \sum_{i=1}^{n} K_\kappa(Y_i - y)/n$. Note that the last factor in the sum of the first equality in (5.5), $(Y_i - y)$, is the shift from each datum Y_i to the point y. Therefore, the partial derivative in (5.5) is a weighted sum of the shifts from the observations to the point y. Now, assuming that $G(\cdot)$ is a function proportional to $g\{(\cdot)^2\}$ implies

$$y = \frac{\sum_{i=1}^{n} K_\kappa(\Theta_i - \theta) G\left(\frac{Y_i - y}{h}\right) Y_i}{\sum_{i=1}^{n} K_\kappa(\Theta_i - \theta) G\left(\frac{Y_i - y}{h}\right)}.$$

However, this equation cannot be solved for y since it appears in both equation terms, *i.e.*, it is a fixed-point equation $y = \omega(y)$, where

$$\omega(\cdot) = \frac{\sum_{i=1}^{n} K_\kappa(\Theta_i - \theta) G\left(\frac{Y_i - \cdot}{h}\right) Y_i}{\sum_{i=1}^{n} K_\kappa(\Theta_i - \theta) G\left(\frac{Y_i - \cdot}{h}\right)}.$$

Note that when $L(\cdot)$ is a Gaussian kernel, $G(\cdot)$ is also a Gaussian kernel and that the function $\omega(y)$ is actually a weighted mean of the observations with weights depending on the proximity of the data to the point (θ, y). In order to obtain the solutions of the equation $y - \omega(y) = 0$, the conditional mean shift algorithm is applied. The unconditional version of the mean shift for euclidean variables was proposed by Fukunaga and Hostetler (1975) in the context of gradient density estimation and subsequently studied by Cheng (1995), Comaniciu and Meer (2002) and Li et al. (2007). The convergence of the algorithm was also investigated by Aliyari Ghassabeh et al. (2013), Aliyari Ghassabeh (2015) and Aliyari Ghassabeh (2013). The mean shift algorithm was generalized to a conditional version by Einbeck and Tutz (2006) in the regression setting and Chen et al. (2014) considered general weights in the context of mode and ridge estimation. Now, to obtain the local modes, the conditional mean shift is applied, where the weights are given by circular kernels. For each value $\theta \in \mathbb{T}$, the so-called

conditional mean shift function, defined as

$$\mathcal{M}(y) = \omega(y) - y, \tag{5.6}$$

is considered. This function represents the shift obtained when moving from y to the weighted mean of the observations with weights depending on y, where the function is evaluated. The key of the conditional mean shift algorithm is that, for a local maximum of the conditional density, the mean shift function takes the value zero. The conditional version of the mean shift algorithm with weights given by circular kernels is described in Algorithm 5.1.

Algorithm 5.1: Conditional mean shift

Data: Sample $\{(\Theta_i, Y_i)\}_{i=1}^n$, smoothing parameters κ and h.

1. Initialize mesh points $\mathcal{T} \subset \mathbb{T}$.
2. For each $\theta \in \mathcal{T}$, select starting points $y_0^{(1)}(\theta), ..., y_0^{(p)}(\theta)$.
3. For each $\theta \in \mathcal{T}$ and for $k = 1, ..., p$, iterate until convergence:

$$y_{l+1}^{(k)} = \frac{\sum_{i=1}^n K_\kappa(\Theta_i - \theta) G\left(\dfrac{Y_i - y_l^{(k)}}{h}\right) Y_i}{\sum_{i=1}^n K_\kappa(\Theta_i - \theta) G\left(\dfrac{Y_i - y_l^{(k)}}{h}\right)}, \quad \text{with } l = 0, 1, ...$$

It can be easily seen that the conditional mean shift is actually a gradient ascent method on $\hat{f}(y|\theta)$ for a fixed θ, where the step size is implicitly chosen, since $\mathcal{M}(y)$ in (5.6) is proportional to and has the same direction as $\frac{\partial}{\partial y}\hat{f}(y|\theta)$. The convergence of the algorithm is guaranteed by the convergence of the unconditional version of the algorithm and, following Comaniciu and Meer (2002), once $y_{l+1}^{(k)}$ gets sufficiently close to a local mode of the conditional density, it will converge to it. Note that the number of starting points in Algorithm 5.1, namely p, may be different for each value of θ. Einbeck and Tutz (2006) used a fixed set of starting points for all the values of the predictor variable and Chen et al. (2016a) took the whole sample as the starting points, as it is usually done in nonparametric modal clustering. However, in order to initialize the algorithm at regions close to the data, a *local initialization* is recommended, where, for each value $\theta \in \mathbb{T}$, the starting values are the quantiles of the sample responses closer to θ.

5.1.2 Circular response

Consider now a circular response variable Φ, with support on $\mathbb{T}$, and either a real-valued predictor X, with support on $\mathbb{R}$, or a circular predictor Θ, with support on $\mathbb{T}$. As in Chapter 2, Δ will denote a generic predictor, with support either on the real line or on the unit circumference. Let $\{(\Delta_i, \Phi_i)\}_{i=1}^n$ be a random sample from (Δ, Φ), with conditional density

function $f(\phi|\delta)$ and with joint density $f(\delta, \phi)$ satisfying assumption **(A2)** in Appendix C.3. Similarly to Section 5.1.1, for each value of the predictor variable, δ, the modal regression multifunction is defined as

$$M(\delta) = \left\{ \phi \in \mathbb{T} : \frac{\partial}{\partial \phi} f(\phi|\delta) = 0, \frac{\partial^2}{\partial \phi^2} f(\phi|\delta) < 0 \right\}, \tag{5.7}$$

with $f(\phi|\delta)$ being the conditional density of Φ given the value of Δ. Note that an interesting feature of multimodal regression for circular responses is that the parameter of interest is analogous to the euclidean setting and it is uniquely defined. This is not the case when considering *mean* regression methods, since one can define the mean direction through the circular mean in equation (1.6), but also through the Fréchet mean, and they are not always defined, making the problem of circular mean regression not easy to handle in some situations. Conversely, the local modes are always well defined as long as the conditional density is not uniform.

In order to estimate the multifunction in (5.7), the same indirect approach as in the previous section is taken: the conditional density is estimated and then its local maxima are computed for each value of the predictor. For the first step, the kernel density estimator of the conditional density for a circular response (Di Marzio et al., 2016) is employed. If the predictor variable is real-valued ($\Delta = X$), the estimator is given by

$$\hat{f}(\phi|x) = \frac{\sum_{i=1}^{n} L_h\left(X_i - x\right) K_\kappa\left(\Phi_i - \phi\right)}{\sum_{i=1}^{n} L_h\left(X_i - x\right)}.$$

In this case $L_h(\cdot)$ is a *linear* kernel with smoothing parameter h and $K_\kappa(\cdot)$ is a circular kernel with concentration parameter κ. If the predictor is circular ($\Delta = \Theta$), the conditional density is estimated as

$$\hat{f}(\phi|\theta) = \frac{\sum_{i=1}^{n} K_\upsilon\left(\Theta_i - \theta\right) K_\kappa\left(\Phi_i - \phi\right)}{\sum_{i=1}^{n} K_\upsilon\left(\Theta_i - \theta\right)},$$

where the expression in the numerator is proportional to an estimation of the joint density with a product of circular kernels (Di Marzio et al., 2011). Here, the kernel associated to the predictor has concentration parameter υ and the kernel associated to Φ has concentration κ. From now on, the weights corresponding to the predictor variable (X or Θ) at the point δ will be denoted as $w_\delta(\Delta_i)$, $i = 1, ..., n$. Note that such weights depend on the smoothing parameter h or υ (depending on Δ being circular or real-valued). Thus,

$$\hat{f}(\phi|\delta) = \frac{1}{n\hat{f}(\delta)} \sum_{i=1}^{n} w_\delta(\Delta_i) K_\kappa(\Phi_i - \phi), \quad \hat{f}(\delta) = \frac{1}{n} \sum_{i=1}^{n} w_\delta(\Delta_i).$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Consequently, the estimator of the multimodal regression function (5.7) is given by

$$\widehat{M}(\delta) = \left\{ \phi \in \mathbb{R} : \frac{\partial}{\partial \phi} \hat{f}(\phi|\delta) = 0, \frac{\partial^2}{\partial \phi^2} \hat{f}(\phi|\delta) < 0 \right\}. \tag{5.8}$$

To obtain the local maxima of $\hat{f}(\delta|\phi)$, the necessary condition for a critical point is established: $\frac{\partial}{\partial\phi}\hat{f}(\phi|\delta) = 0$. By applying (5.3), it holds

$$\frac{\partial}{\partial\phi}\hat{f}(\phi|\delta) = \frac{\kappa c_\kappa}{n\hat{f}(\delta)}\sum_{i=1}^{n} w_\delta(\Delta_i)K'\{\kappa[1-\cos(\Phi_i-\phi)]\}\sin(\Phi_i-\phi). \tag{5.9}$$

Hence, the derivative of the estimated conditional density with respect to ϕ is a weighted sum of the sines of the differences from each observation to the point ϕ. In this case, the sine function is used for measuring the shifts from the observations to ϕ. This is very intuitive since, if $\phi = \Phi_j$, then $\sin(\Phi_i - \phi) = 0$. In addition, if the difference $\Phi_i - \phi$ is very small, then $\sin(\Phi_i - \phi) \approx \Phi_i - \phi$. Therefore, the right side of equation (5.9) can be interpreted as a weighted sum of the shifts from the observations to ϕ. Expanding the last factor in the right side of (5.9) yields

$$\frac{\partial}{\partial\phi}\hat{f}(\phi|\delta) = \frac{\kappa c_\kappa}{n\hat{f}(\delta)}\sum_{i=1}^{n} w_\delta(\Delta_i)K'\{\kappa[1-\cos(\Phi_i-\phi)]\}(\sin\phi\cos\Phi_i - \cos\phi\sin\Phi_i),$$

and equating it to zero gives

$$\sin\phi\sum_{i=1}^{n} w_\delta(\Delta_i)T(\Phi_i-\phi)\cos\Phi_i = \cos\phi\sum_{i=1}^{n} w_\delta(\Delta_i)T(\Phi_i-\phi)\sin\Phi_i,$$

where $T(\cdot)$ is proportional to $K'\{\kappa[1-\cos(\cdot)]\}$. Therefore, by denoting

$$S_\delta(\phi) = \sum_{i=1}^{n} w_\delta(\Delta_i)T(\Phi_i-\phi)\sin\Phi_i \quad \text{and} \quad C_\delta(\phi) = \sum_{i=1}^{n} w_\delta(\Delta_i)T(\Phi_i-\phi)\cos\Phi_i,$$

it holds that if $S_\delta(\phi) \neq 0$ or $C_\delta(\phi) \neq 0$, then $\phi = \text{atan2}\left[S_\delta(\phi), C_\delta(\phi)\right]$, where the operator $\text{atan2}(\cdot,\cdot)$ was defined in equation (1.7). Thus, it is obtained that

$$\phi = \tilde{\omega}(\phi) = \text{atan2}\left[S_\delta(\phi), C_\delta(\phi)\right].$$

Note that the function $\tilde{\omega}(\phi)$ actually returns a weighted circular mean direction of the observations, since the quantity $S_\delta(\phi)$ is a weighted sum of $\sin\Phi_i$ and $C_\delta(\phi)$ is a weighted sum of $\cos\Phi_i$ (where the weights depend on the point (δ,ϕ)). Since the previous expression is a fixed-point equation, a mean shift-type algorithm is used in order to obtain the conditional mode estimator. As an analogy with the mean shift function in equation (5.6), the conditional circular mean shift function is defined as:

$$\widetilde{\mathcal{M}}(\phi) = \sin[\tilde{\omega}(\phi) - \phi].$$

As discussed before, given that for small values of $\tilde{\omega}(\phi) - \phi$ it holds that $\sin[\tilde{\omega}(\phi) - \phi] \approx$

$\tilde{\omega}(\phi) - \phi$, the sine function is used to measure the shift from ϕ to $\tilde{\omega}(\phi)$. In addition, for a local mode of the conditional density estimator, the circular mean shift function takes the value zero. Consequently, the estimated regression multifunction (5.8) is obtained by using the conditional circular mean shift algorithm, which is described in Algorithm 5.2. As in Algorithm 5.1, the number of starting points, p, can be different for each value of δ, and the same initialization as in in Algorithm 5.1, but using circular quantiles (see Fisher, 1993), is recommended.

Algorithm 5.2: Circular conditional mean shift

Data: Sample $\{(\Delta_i, \Phi_i)\}_{i=1}^n$, smoothing parameters κ and h/v.

1. Initialize mesh points $\mathcal{S} \subset \mathbb{R}$ if $\Delta = X$ or $\mathcal{T} \subset \mathbb{T}$ if $\Delta = \Theta$.
2. For each $\delta \in \mathcal{S}$ (or $\delta \in \mathcal{T}$), select starting points $\phi_0^{(1)}(\delta), ..., \phi_0^{(p)}(\delta)$.
3. For $k = 1, ..., p$ iterate until convergence:

$$\phi_{l+1}^{(k)} = \text{atan2}\left[\sum_{i=1}^{n} w_\delta(\Delta_i) T\left(\Phi_i - \phi_l^{(k)}\right) \sin \Phi_i, \sum_{i=1}^{n} w_\delta(\Delta_i) T\left(\Phi_i - \phi_l^{(k)}\right) \cos \Phi_i\right],$$

with $l = 0, 1, ...$

Remark 5.1. The unconditional version of Algorithm 5.2 actually coincides with the directional mean shift proposed by Oba et al. (2005) and studied by Kobayashi and Otsu (2010) and Zhang and Chen (2021a), in the particular case of data in the unit circumference. Note that, however, the derivations of the algorithms are different. The results in Zhang and Chen (2021a,c) guarantee the convergence of Algorithm 5.2 and the fact that the algorithm converges to a mode of the estimated conditional density.

5.2 Some theoretical results

The aim of this section is to give asymptotic error rates of the multimodal regression estimators introduced in Section 5.1. Note that usual error metrics in kernel regression (such as the MISE or ISE) are not adequate to measure the quality of the estimators since, in the multimodal regression context, the objective function is actually a multifunction, *i.e.*, for each value of the predictor, there might be a set of values of the response. Following Chen et al. (2016a), pointwise and mean integrated errors (based on distances between sets) are considered.

USC DE SANTIAGO DE COMPOSTELA

First, the Hausdorff distance is introduced: for two sets $A, B \subset \mathbb{R}^q$, it is defined as

$$\text{Haus}(A, B) = \max\left\{\sup_{x \in A} d(x, B), \sup_{x \in B} d(x, A)\right\}, \tag{5.10}$$

where, $d(x, A) = \inf_{z \in A} ||x - z||$. The Hausdorff distance measures how close two euclidean

sets are from each other. In the real-valued response case exposed in Section 5.1.1, the Hausdorff distance is an adequate measure of the distance from the true regression multifunction to the estimated multifunction, because $M(\theta)$ in equation (5.1) and its estimated version $\widehat{M}(\theta)$ are subsets of $\mathbb{R}$. However, in the circular response scenario, the multifunctions $M(\delta), \widehat{M}(\delta)$ are subsets of $\mathbb{T}$ and, thus, it is necessary to generalize the definition of the Hausdorff distance by considering, for $A, B \subset \mathbb{T}$,

$$\widetilde{\text{Haus}}(A, B) = \max\left\{ \sup_{x \in A} \tilde{d}(x, B), \sup_{x \in B} \tilde{d}(x, A) \right\}, \tag{5.11}$$

with $\tilde{d}(x, A) = \inf_{z \in A}[1 - \cos(x - z)]$. The Hausdorff distances defined in (5.10) and (5.11) are used to construct the pointwise error metrics, defined as

$$\Lambda(\theta) = \text{Haus}[M(\theta), \widehat{M}(\theta)], \quad \tilde{\Lambda}(\delta) = \widetilde{\text{Haus}}[M(\delta), \widehat{M}(\delta)], \tag{5.12}$$

where the first pointwise error is applied to the multifunctions with a circular predictor and real-valued response and the second is used for multifunctions with a circular response. The pointwise errors measure how close the estimated multifunction is from the true regression multi-valued function for each value of the predictor, θ or δ, respectively. In order to obtain a global error measure, the modal Mean Integrated Squared Error (MISE_m) and the modal Circular Mean Integrated Error (CMIE_m) are considered:

$$\text{MISE}_m(\widehat{M}) = \mathbb{E}\left[\int_{\theta \in \mathbb{T}} \Lambda^2(\theta) d\theta \right], \quad \text{CMIE}_m(\widehat{M}) = \mathbb{E}\left[\int_{\delta \in \text{Supp}(\Delta)} \tilde{\Lambda}(\delta) d\delta \right],$$

where $\text{Supp}(\Delta)$ denotes the support of Δ, which recall that can denote either a circular or real-valued predictor. The MISE_m is obtained by taking the expectation of the integrated squared pointwise error, and is a global error measure for the multimodal regression estimator in (5.4). The CMIE_m, constructed to measure the global error of the estimator in (5.8), is the expectation of the integrated pointwise error. The reason why the circular error is not squared is that it is constructed as an analogy with the euclidean case: for real-valued variables, the MSE of an estimator is the expectation of the squared distance between the estimator and the parameter while, for circular variables, the so-called *circular mean squared error*, defined by Kim and SenGupta (2017), is the expectation of the cosine distance between the estimator and the parameter.

The following propositions show the consistency of the estimators. The proofs can be found in Appendix C.3.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Proposition 5.1. *Let Θ be a circular random variable and Y a real-valued random variable. Consider the multimodal regression estimator in (5.4). Under assumptions **(A1)**, **(A3)**,*

*(**A4-1**), (**B1**) and (**C1**) in Appendix C.3, it holds that,*

$$\Lambda(\theta) = O(\kappa^{-1} + h^2) + O_P\left(\sqrt{\frac{\kappa^{1/2}}{nh^3}}\right)$$

and

$$\mathit{MISE}_m(\widehat{M}) = O(\kappa^{-2} + h^4) + O\left(\frac{\kappa^{1/2}}{nh^3}\right),$$

as $h \to 0$, $\kappa \to \infty$ *and* $\frac{nh^{1+2q}}{\kappa^{1/2}|\log(h\kappa^{-1/2})|} \to \infty$, *for* $q = 0, 1, 2$.

Proposition 5.2. *Let* X *be a real-valued random variable and* Φ *a circular random variable. Consider the multimodal regression estimator in (5.8). Under assumptions (**A2**), (**A3**), (**A4**), (**B2**) and (**C2**) in Appendix C.3, it holds that*

$$\tilde{\Lambda}(x) = O(h^2 + \kappa^{-1}) + O_P\left(\sqrt{\frac{\kappa^{3/2}}{nh}}\right)$$

and

$$\mathit{CMIE}_m(\widehat{M}) = O(h^2 + \kappa^{-1}) + O\left(\sqrt{\frac{\kappa^{3/2}}{nh}}\right),$$

as $\kappa \to \infty$, $h \to 0$ *and* $\frac{nh}{\kappa^{(1+2q)/2}|\log(h\kappa^{-1/2})|} \to \infty$, *for* $q = 0, 1, 2$.

Proposition 5.3. *Let* Θ *and* Φ *be circular random variables. Consider the multimodal regression estimator in (5.8). Under assumptions (**A2**), (**A3**), (**A4**), (**B2**) and (**C2**) in Appendix C.3, it holds that*

$$\tilde{\Lambda}(\theta) = O(\upsilon^{-1} + \kappa^{-1}) + O_P\left(\sqrt{\frac{\kappa^{3/2}\upsilon^{1/2}}{n}}\right)$$

and

$$\mathit{CMIE}_m(\widehat{M}) = O(\upsilon^{-1} + \kappa^{-1}) + O\left(\sqrt{\frac{\kappa^{3/2}\upsilon^{1/2}}{n}}\right),$$

as $\kappa \to \infty$, $\upsilon \to \infty$ *and* $\frac{n}{\kappa^{(1+2q)/2}\upsilon^{1/2}|\log(\upsilon^{-1/2}\kappa^{-1/2})|} \to \infty$, *for* $q = 0, 1, 2$.

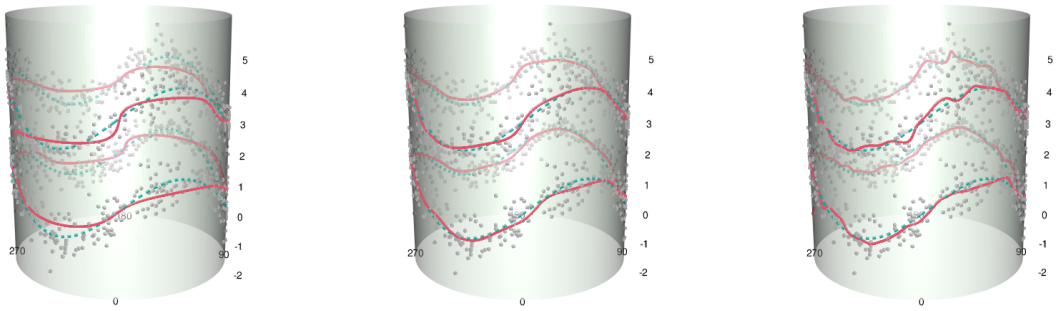
Roughly speaking, the derived rates show that the pointwise errors $\Lambda(\theta)$, $\tilde{\Lambda}(x)$ and $\tilde{\Lambda}(\theta)$ converge to zero at the same rate as the first partial derivative, with respect to the response variable, of the joint kernel density estimators, namely $\frac{\partial}{\partial y}\hat{f}(\theta, y)$, $\frac{\partial}{\partial \phi}\hat{f}(x, \phi)$ and $\frac{\partial}{\partial \phi}\hat{f}(\theta, \phi)$. Note that all the obtained rates are the same as in the euclidean case if one thinks of $\kappa \asymp h^{-2}$ and $\upsilon \asymp h^{-2}$. Note that the conditions **(C1)** and **(C2)** in Appendix C.3 imply that the *branches* of the regression multifunction cannot cross each other, as it happens in the real-valued context (Chen et al., 2016a).

5.3 Smoothing parameter selection

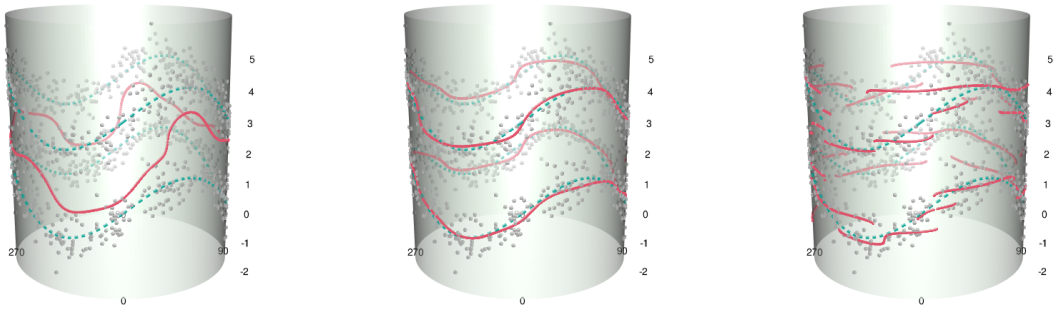
As it usually happens in kernel estimation, the selection of the smoothing parameters is a crucial issue in multimodal regression. As mentioned in Section 1.1, when performing classical kernel regression with real-valued covariates, a large value of the bandwidth h leads to an oversmoothed estimator, while a small value of h produces an undersmoothed estimation of the regression function. Similarly, and as exposed in the previous chapters, when a circular kernel is employed, the behavior of the concentration parameter is reversed: when κ is large, an undersmoothed estimation of the regression function is obtained, and a small value of κ leads to an oversmoothed estimator.

However, in the context of multimodal regression two smoothing parameters are required, one associated to the predictor variable and one associated to the response, and the role of the two parameters is very different. The parameter associated to the predictor variable controls the smoothing of the regression multifunction, having a similar role to the smoothing parameter in classical kernel regression. On the contrary, the parameter associated to the response variable influences the number of estimated modes. The reason behind this behavior is that an undersmoothed estimator of the conditional density will lead to many local modes, producing a large number of estimated *branches* in the regression multifunction. This behavior is illustrated in Figure 5.2 with simulated data from a multimodal regression model with a circular predictor and a real-valued response (so the data cloud lies on a cylinder), although similar interpretations are obtained for the scenarios involving a circular response variable. In the top row of Figure 5.2, the parameter h, associated to the response variable is fixed, while κ (the parameter associated to the covariate) takes different values. As the value of the concentration increases, the multimodal estimator turns more undersmoothed. On the other hand, in the bottom row, the value of κ was fixed and the value of h was changed. If h is too large, the two *branches* of the regression multifunction are merged into one, but if h is too small, many local modes are estimated for each value of the predictor, resulting into a too fragmented estimator.

The statistical literature on the selection of the smoothing parameters for multimodal regression in the euclidean setting is quite scarce. Since the estimation is carried out by obtaining the local maxima of the conditional density estimator, Einbeck and Tutz (2006) recommended using methods for bandwidth selection tailored for conditional density estimation. However, such methods may not be adequate in practice, because mode estimation is related but not equivalent to density estimation. As noted by Casa et al. (2020), an estimated density might be close to the true density in terms of ISE, but have many estimated local modes which, in the case of modal regression, would lead to many estimated *branches* of the regression multifunction. Zhou and Huang (2019) proposed two different methods for obtaining smoothing parameters in multimodal regression for real-valued response and covariate. The first of them, called modal cross-validation, aims to balance the number of estimated local modes and the distance from the estimation to the data. Although the method shows



(a) $\kappa = 5$ *(b)* $\kappa = 60$ *(c)* $\kappa = 300$



(d) $h = 1.5$ *(e)* $h = 0.5$ *(f)* $h = 0.2$

Figure 5.2: Simulated data from a multimodal regression model with a circular predictor and a real-valued response. The dashed curves represent the true regression multifunctions and the continuous trace shows the estimated multifunction. In the top row, the smoothing parameter is $h = 0.6$ *and the values of* κ *are specified below each panel. In the bottom row, the value of* κ *is 30 and the values of* h *are specified below each panel.*

a good performance in practice, nothing assures that minimizing the modal cross-validation function will minimize the MISE_m of the estimator or any other error criterion. The second approach in Zhou and Huang (2019) is to minimize the ISE_m of the estimator (the integrated Hausdorff distance between the estimator and the true multifunction) by using a parametric bootstrap based on a mixture of parametric regressions. Another approach was proposed by Chen et al. (2016a), which consists on constructing a prediction band for the regression multifunction and afterwards selecting the parameters minimizing a loss function defined as the volume of the prediction band. However, the authors assume that the smoothing parameter is the same for both variables, which is not well justified, especially given the different role of the two parameters. In addition, the choice of the smoothing parameters depends on the previously selected prediction level.

5.3.1 Modal cross-validation for circular regression

The modal cross-validation method in Zhou and Huang (2019) is easily adaptable to the context of circular multimodal regression. Although the theoretical properties of the modal cross-validation method are not well studied, the behavior of the method in practice is quite satisfactory. Thus, in this section, the cross-validation method is modified to deal with circular variables. For the scenario presented in Section 5.1.1 (circular predictor and real-valued response), the modal CV approach consists on selecting κ and h by minimizing

$$CV_1(\kappa, h) = \frac{1}{n}\sum_{i=1}^{n} d^2\left[\widehat{M}_{-i}^{\kappa,h}(\Theta_i), Y_i\right] N_{-i}^2(\Theta_i),$$

where $d(x, A) = \inf_{z\in A} ||x - z||$, $\widehat{M}_{-i}^{\kappa,h}(\cdot)$ is the estimator of $M(\cdot)$ using data $\{(\Theta_j, Y_j) : j \neq i\}$ and constructed with smoothing parameters κ and h and $N_{-i}(\Theta_i)$ is the number of estimated modes of $\widehat{M}_{-i}^{\kappa,h}(\Theta_i)$.

For the circular response case exposed in Section 5.1.2, the parameters ρ and κ (where ρ represents either h or υ, depending on the nature of the predictor) are selected by minimizing

$$CV_2(\rho, \kappa) = \frac{1}{n}\sum_{i=1}^{n} \tilde{d}\left[\widehat{M}_{-i}^{\rho,\kappa}(\Delta_i), \Phi_i\right] N_{-i}(\Delta_i),$$

where $\tilde{d}(x, A) = \inf_{z\in A} 1 - \cos(x - z)$ and $\widehat{M}_{-i}^{\rho,\kappa}(\cdot)$ is the estimator of $M(\cdot)$ using data $\{(\Delta_j, \Phi_j) : j \neq i\}$ and constructed with smoothing parameters ρ and κ. This approach does not rely on a theoretical basis and computational trials are needed in order to assess the efficacy of the method.

In the next section, a modification of this method is employed in order to accomplish more computational efficiency. Since the role of the two smoothing parameters is quite different, first, a pair of reference parameters is selected. Then, the parameter corresponding to the predictor variable is updated by minimizing the modal cross-validation function and, lastly, the second smoothing parameter is also updated. The detailed modified criteria can be found in Appendix B.2. Through simulation experiments (see Appendix B.2) it was verified that results using this modification of the modal cross-validation criteria are very similar to the ones obtained with the original procedures described in this section, but the modification is more efficient in computational terms.

5.4 Simulation study

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

In this section, the finite sample performance of the circular multimodal regression estimators is analyzed in the three different contexts (circular predictor and real-valued response, real-valued predictor and circular response and circular predictor and circular response). First, the simulation scenarios are presented and the obtained results are subsequently discussed.

5.4.1 Simulation scenarios

In the simulated examples, the sample is divided into two groups, with each group corresponding to a *branch* of the regression multifunction. For each observation, the subscript ji is used, where j denotes the group or *branch* number and i denotes the observation number within each group. In addition, n_j denotes the sample size for the jth group. Note that these underlying groups are not known in practice and there is no information about them. The simulated models are presented in Table 5.1. The first model in each scenario corresponds to two parallel regression curves or, seen from another perspective, to a regression curve with a bimodal error. As for the second model of each scenario, the regression curves are not parallel. The same happens with the third model, although this time the support of the first group is different (and smaller) from the support of the second group. For all scenarios and models the sample sizes are $(n_1, n_2) \in \{(100, 100), (100, 200), (200, 200), (200, 300), (300, 300)\}$. Examples of data from all simulated models are shown in Figure 5.3.

In order to assess the performance of the estimators, the MISE_m and CMIE_m are approximated by generating 500 Monte Carlo replicates of the simulated data and computing the average of the modal Integrated Error/modal Circular Integrated Error (ISE_m or CIE_m) of the multimodal estimator:

$$\text{ISE}_m(\widehat{M}) = \int_{\theta \in \mathbb{T}} \Lambda^2(\theta) d\theta, \quad \text{for euclidean response,}$$

$$\widetilde{\text{ISE}}_m(\widehat{M}) = \int_{\delta \in \text{Supp}(\Delta)} \widetilde{\Lambda}(\delta) d\delta, \quad \text{for circular response,}$$

where the integrals are numerically approximated by Simpson's rule and $\Lambda(\cdot), \widetilde{\Lambda}(\cdot)$ are defined in (5.12). In the simulation experiments, the smoothing parameters were selected by the modified version of the modal cross-validation method described in Appendix B.2.

5.4.2 Simulation results

The simulation results are presented in this section. Comments about the first and second models in each scenario are given, whereas the results of the third models (CL-3, LC-3 and CC-3) are discussed later as a special case.

The estimated MISE_m for models CL-1 and CL-2 is reported in Table 5.2. Results are obtained with the smoothing parameters selected by the modified modal cross-validation and also with the *optimal* parameters minimizing ISE_m (found through intensive search), which are used as benchmarks. As expected, the estimated MISE_m generally decreases both when increasing sample size and decreasing standard deviation. In addition, the performance of the modal cross-validation criterion to select the smoothing parameters is quite satisfactory, as the estimated MISE_m obtained using this method is reasonably close to the benchmark MISE_m. The larger differences are obtained for the smallest sample sizes, especially when

Table 5.1: Simulated models. CL stands for circular predictor and real-valued response, LC stands for real-valued predictor and circular response and CC stands for circular predictor and circular response. For CL models, $\varepsilon_{1i}, \varepsilon_{2i} \sim N(0, \sigma^2)$ with $\sigma \in \{0.75, 1, 1.25\}$. For LC and CC models, $\varepsilon_1, \varepsilon_2 \sim vM(0, \tau)$ with $\tau \in \{6, 8, 10\}$ for LC and $\tau \in \{10, 12, 14\}$ for CC.

Model	Data generation	Regression multifunction
CL-1	$Y_{1i} = 2\sin(2\Theta_{1i}) + \varepsilon_{1i}$ $Y_{2i} = 8 + 2\sin(2\Theta_{2i}) + \varepsilon_{2i}$ $\Theta_1, \Theta_2 \sim$ Circular Uniform	$M(\theta) = \{2\sin(2\theta), 8 + 2\sin(2\theta)\}$
CL-2	$Y_{1i} = -2\cos(\Theta_{1i}) + \varepsilon_{1i}$ $Y_{2i} = 2\sin(2\Theta_{2i}) - 10 + \varepsilon_{2i}$ $\Theta_1, \Theta_2 \sim$ Circular Uniform	$M(\theta) = \{2\sin(2\theta), 8 + 2\sin(2\theta)\}$
CL-3	$Y_{1i} = 2\cos(\Theta_{1i}) + \varepsilon_{1i}$ $Y_{2i} = 8 + 2\sin(2\Theta_{2i}) + \varepsilon_{2i}$ $\Theta_1 \sim U(0, \pi),\ \Theta_2 \sim$ Circular Uniform	$M(\theta) = \{2\sin(2\theta)\}$ if $\theta \in [0, \pi)$ $M(\theta) = \{8 + 2\sin(2\theta)\}$ otherwise
LC-1	$\Phi_{1i} = 6\tan^{-1}(2.5X_{1i} - 3) + \varepsilon_{1i}$ $\Phi_{2i} = \pi + 6\tan^{-1}(2.5X_{2i} - 3) + \varepsilon_{2i}$ $X_1, X_2 \sim U(0, 1)$	$M(x) = \{6\tan^{-1}(2.5x - 3),$ $\pi + 6\tan^{-1}(2.5x - 3)\}$
LC-2	$\Phi_{1i} = \text{atan2}(\sin(3X_{1i}^2), \cos(3X_{1i}^2)) + \varepsilon_{1i}$ $\Phi_{2i} = \pi/2 + 2\tan^{-1}(10X_{2i} - 1/2) + \varepsilon_{2i}$ $X_1, X_2 \sim U(0, 1)$	$M(x) = \{\text{atan2}(\sin(3x^2), \cos(3x^2)),$ $\pi/2 + 2\tan^{-1}(10x + 1/2)\}$
LC-3	$\Phi_{1i} = \pi/4 - 3\tan^{-1}(2.5X_{1i} - 3) + \varepsilon_{1i}$ $\Phi_{2i} = \pi/2 + 6\tan^{-1}(2.5X_{2i} - 3) + \varepsilon_{2i}$ $X_1 \sim U(0, 0.5),\ X_2 \sim U(0, 1)$	$M(x) = \{\pi/4 - 3\tan^{-1}(2.5x - 3), \pi/2 +$ $6\tan^{-1}(2.5x - 3)\}$ if $x \in [0, 0.5]$ $M(x) = \{\pi/2 + 6\tan^{-1}(2.5x - 3)\}$ otherwise
CC-1	$\Phi_{1i} = 2\cos\Theta_{1i} + \varepsilon_{1i}$ $\Phi_{2i} = 3\pi/4 + 2\cos\Theta_{2i} + \varepsilon_{2i}$ $\Theta_1, \Theta_2 \sim$ Circular Uniform	$M(\theta) = \{2\cos\theta, 3\pi/4 + 2\cos\theta\}$
CC-2	$\Phi_{1i} = 3/4\cos\Theta_{1i} - \pi/2 + \varepsilon_{1i}$ $\Phi_{2i} = \pi/2 - \cos\Theta_{2i} + \varepsilon_{2i}$ $\Theta_1, \Theta_2 \sim$ Circular Uniform	$M(\theta) = \{3/4\cos\theta - \pi/2, \pi/2 - 2\cos\theta\}$
CC-3	$\Phi_{1i} = 2\pi/3 + 3\sin(\Theta_{1i}/3) + \varepsilon_{1i}$ $\Phi_{2i} = 3\pi/4 + 1.75\cos\Theta_{2i} + \varepsilon_{2i}$ $\Theta_1 \sim U(2\pi/3, 3\pi/2),\ \Theta_2 \sim$ Circular Uniform	$M(\theta) = \{2\pi/3 + 3\sin(\theta/3), 3\pi/4 + 1.75\cos\theta\}$ if $\theta \in [2\pi/3, 3\pi/2]$ $M(\theta) = \{3\pi/4 + 1.75\cos\theta\}$ otherwise

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Note: The arithmetic in $\mathbb{T}$ is performed modulo 2π.

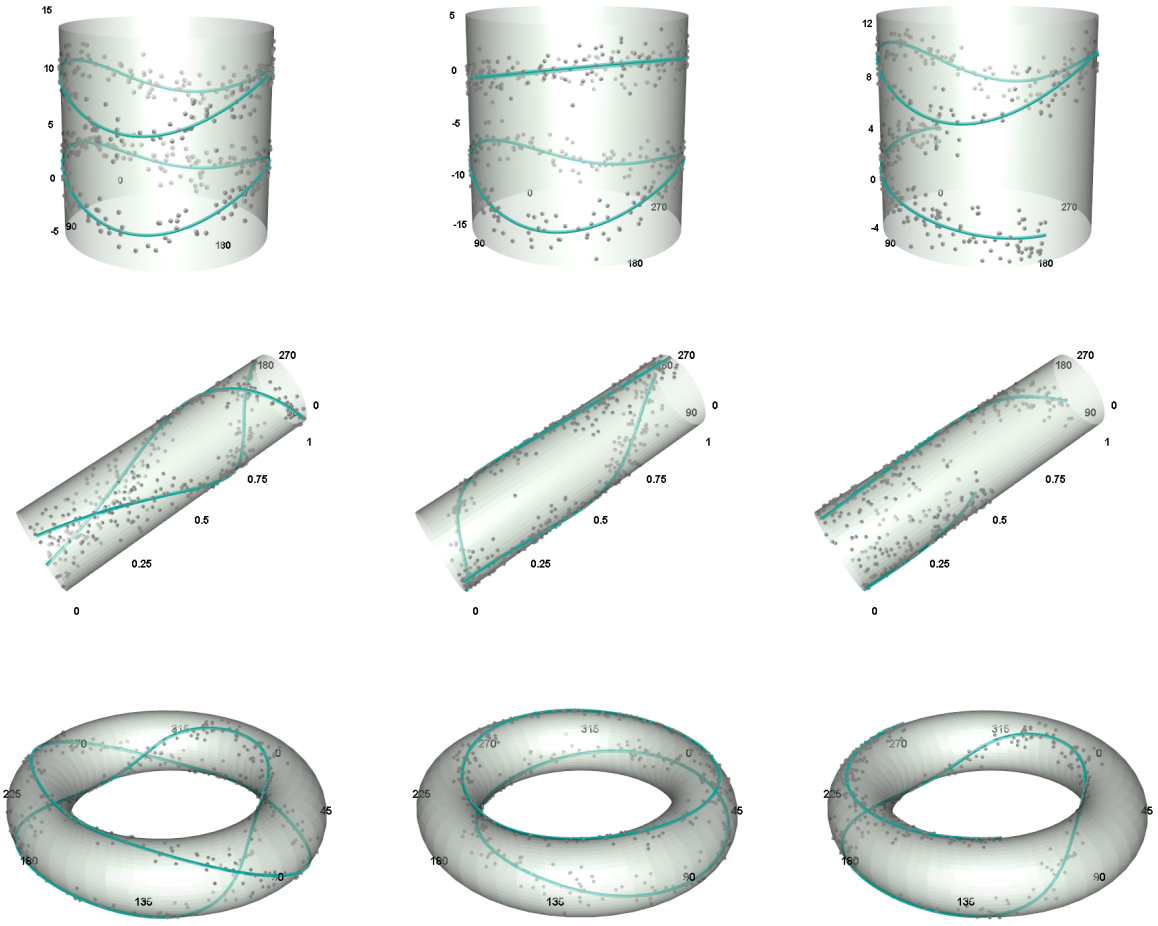


Figure 5.3: Representations on the cylinder and torus of simulated data and true regression multifunctions from models CL-1 (top left), CL-2 (top center) and CL-3 (top right) with standard deviation $\sigma = 1$; models LC-1 (middle left), LC-2 (middle center) and LC-3 (middle right) with concentration $\tau = 8$; and models CC-1 (bottom left), CC-2 (bottom center) and CC-3 with concentration $\tau = 12$. Sample sizes are $(n_1, n_2) = (200, 200)$ for all models.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$\sigma = 1.25$. Note that, although the estimations of MISE_m tend to decrease when increasing the sample size, a much remarkable decrease is obtained when the sample size is the same in both groups, *i.e.*, $(n_1, n_2) \in \{(100, 100), (200, 200), (300, 300)\}$. When the sample size in each group is different, it is observed that the estimated MISE_m does not decrease equally fast and is sometimes larger than with the previous sample size when using the modal cross-validation criterion. This behavior is due to the need of different degrees of smoothing in each *branch* of the multifunction when the sample sizes are highly unbalanced.

Table 5.2: Monte Carlo estimates of MISE_m for models CL-1, CL-2 and CL-3 with different values of the standard deviation σ and the sample sizes. Results under B (benchmark) correspond to the ones obtained with smoothing parameters minimizing ISE_m and results under CV_1 refer to the ones obtained with smoothing parameters selected by modified modal cross-validation.

Circular predictor and real-valued response							
		$\sigma = 1.25$		$\sigma = 1$		$\sigma = 0.75$	
Model	(n_1, n_2)	B	CV_1	B	CV_1	B	CV_1
CL-1	$(100, 100)$	2.308	3.644	1.456	1.877	0.928	1.012
	$(100, 200)$	1.935	3.382	1.191	1.653	0.729	1.021
	$(200, 200)$	1.172	1.321	0.776	0.851	0.480	0.509
	$(200, 300)$	1.031	1.215	0.650	0.706	0.411	0.438
	$(300, 300)$	0.835	0.940	0.527	0.574	0.340	0.359
CL-2	$(100, 100)$	1.578	3.194	1.076	1.777	0.697	0.951
	$(100, 200)$	1.270	3.252	0.836	1.439	0.528	0.903
	$(200, 200)$	0.880	1.143	0.596	0.686	0.381	0.432
	$(200, 300)$	0.735	0.912	0.489	0.583	0.309	0.345
	$(300, 300)$	0.613	0.710	0.412	0.473	0.261	0.290
CL-3	$(100, 100)$	1.489	3.673	0.921	2.080	0.749	1.144
	$(100, 200)$	0.959	2.321	0.590	1.244	0.358	0.731
	$(200, 200)$	0.785	1.917	0.491	0.849	0.299	0.448
	$(200, 300)$	0.570	1.179	0.359	0.607	0.219	0.356
	$(300, 300)$	0.534	1.307	0.338	0.515	0.206	0.278

Results for models LC-1 and LC-2 are given in Table 5.3, while Table 5.4 contains the results for models CC-1 and CC-2. Results are similar to those of the real-valued response scenario: the estimated value of CMIE_m generally diminishes as the sample size is larger. The few cases where this behavior is not observed is when the sample sizes are unbalanced, *i.e.*, $(n_1, n_2) \in \{(100, 200), (200, 300)\}$. In addition, a large value of the error concentration also leads to a smaller estimated CMIE_m. The performance of the cross-validation criterion

is also more than acceptable, as for large values of the sample size, the value of the estimated CMIE_m obtained with smoothing parameters selected by the modified modal cross-validation is very close to the values obtained with the *optimal* smoothing parameters. The poorest performance is achieved in model LC-2 when $\tau = 6$. As shown in the middle center panel of Figure 5.3, when the value of the predictor variable is around 1, the two *branches* are very close, which makes it difficult to discern the two groups when the error concentration is small.

Table 5.3: Monte Carlo estimates of CMIE_m for models LC-1, LC-2 and LC-3 with different values of the concentration τ and the sample sizes. Results under B (benchmark) correspond to the ones obtained with smoothing parameters minimizing CIE_m and results under CV_2 refer to the ones obtained with smoothing parameters selected by modified modal cross-validation.

Real-valued predictor and circular response							
		$\tau = 6$		$\tau = 8$		$\tau = 10$	
Model	(n_1, n_2)	B	CV_2	B	CV_2	B	CV_2
LC-1	$(100, 100)$	0.021	0.034	0.016	0.023	0.013	0.018
	$(100, 200)$	0.016	0.039	0.012	0.024	0.011	0.018
	$(200, 200)$	0.011	0.019	0.008	0.012	0.007	0.010
	$(200, 300)$	0.010	0.016	0.007	0.011	0.006	0.009
	$(300, 300)$	0.007	0.012	0.006	0.008	0.005	0.006
LC-2	$(100, 100)$	0.016	0.035	0.012	0.025	0.010	0.018
	$(100, 200)$	0.016	0.070	0.010	0.039	0.008	0.024
	$(200, 200)$	0.010	0.023	0.007	0.012	0.005	0.009
	$(200, 300)$	0.009	0.038	0.006	0.017	0.005	0.011
	$(300, 300)$	0.007	0.015	0.005	0.009	0.004	0.007
LC-3	$(100, 100)$	0.014	0.030	0.010	0.017	0.008	0.013
	$(100, 200)$	0.008	0.012	0.006	0.008	0.005	0.007
	$(200, 200)$	0.008	0.016	0.005	0.008	0.004	0.006
	$(200, 300)$	0.006	0.009	0.004	0.006	0.003	0.005
	$(300, 300)$	0.006	0.010	0.004	0.006	0.003	0.005

5.4.3 The unknown boundary condition

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

In this section, the results for models CL-3, LC-3 and CC-3 are analyzed. These models are studied as a special case since, unlike in the other models, here the support of the first group is different from the support of the other group. In fact, the support of the first group is always contained in the support of the second group. This issue makes the analysis of the

Table 5.4: Monte Carlo estimates of $CMIE_m$ for models CC-1, CC-2 and CC-3 with different values of the concentration τ and the sample sizes. Results under B (benchmark) correspond to the ones obtained with smoothing parameters minimizing CIE_m and results under CV_2 refer to the ones obtained with smoothing parameters selected by modified modal cross-validation.

Circular predictor and circular response							
		$\tau = 10$		$\tau = 12$		$\tau = 14$	
Model	(n_1, n_2)	B	CV_2	B	CV_2	B	CV_2
CC-1	$(100, 100)$	0.144	0.395	0.117	0.213	0.102	0.148
	$(100, 200)$	0.120	0.182	0.097	0.147	0.087	0.120
	$(200, 200)$	0.066	0.089	0.054	0.066	0.047	0.057
	$(200, 300)$	0.058	0.068	0.046	0.054	0.040	0.045
	$(300, 300)$	0.042	0.056	0.035	0.045	0.030	0.038
CC-2	$(100, 100)$	0.116	0.227	0.092	0.173	0.080	0.150
	$(100, 200)$	0.085	0.232	0.068	0.159	0.057	0.120
	$(200, 200)$	0.055	0.091	0.045	0.069	0.039	0.059
	$(200, 300)$	0.046	0.092	0.037	0.066	0.032	0.049
	$(300, 300)$	0.037	0.055	0.030	0.043	0.026	0.033
CC-3	$(100, 100)$	0.063	0.094	0.054	0.076	0.048	0.065
	$(100, 200)$	0.032	0.041	0.027	0.033	0.024	0.029
	$(200, 200)$	0.031	0.037	0.027	0.031	0.024	0.027
	$(200, 300)$	0.022	0.027	0.018	0.022	0.016	0.019
	$(300, 300)$	0.021	0.024	0.018	0.020	0.016	0.018

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

performance of the estimators rather tricky, as it is difficult to obtain the *optimal* smoothing parameters. This situation will be illustrated with model CL-3, but it also applies for the other two scenarios.

Figure 5.4 shows a representation on the plane of simulated data from model CL-3 with the true and estimated regression multifunctions. In the left panel, the smoothing parameters were selected by modal cross-validation, while the right panel shows the estimation where the smoothing parameters are the *optimal* ones minimizing ISE_m. Focusing on the left panel, it can be seen that the lower *branch* of the estimated regression multifunction includes a part falling outside of its support (represented with triangles). This is a general and expected behavior for all kernel methods, both for density and regression. Usually, one just restricts the estimation to the support of the variable but, in multimodal regression, the support of each group is unknown (recall that these groups are unknown in practice and there is no information about them). On the early stages of modal regression, Scott (1992) had already noticed this behavior, referring to it as the *unknown boundary condition*. Thus, the yellow part in Figure 5.4 leads to a huge ISE_m. Since the part represented with triangles is the main source of the error in these situations, when computing the *optimal* parameters which minimize ISE_m, very large concentrations (for a circular predictor) or very small values of the bandwidth (for a real-valued predictor) are obtained, leading to a heavily undersmoothed estimator as in the right panel of Figure 5.4. In order to show that the modal cross-validation criterion achieves a good performance within the support of the variables, in this simulation study, the error has been computed after having removed the estimations outside the support of each group (which cannot be done with real data where this is unknown). In addition, the *optimal* parameters were also selected by removing these extra parts of the estimators (such as the triangle parts in Figure 5.4).

The bottom rows in Tables 5.2, 5.3 and 5.4 report the estimated MISE_m and CIE_m for models CL-3, LC-3 and CC-3, respectively. For model CL-3, the estimated MISE_m generally decreases by increasing the sample size and diminishing the standard deviation. For $\sigma = 1$ and $\sigma = 0.75$, when the sample size is large, the results obtained with the modal cross-validation criterion are very close to the ones obtain with the *optimal* parameters. More difficulties are encountered when $\sigma = 1.25$, as the large dispersion makes the two *branches* really close to each other in some areas. Unlike in models CL-1 and CL-2, a better performance is achieved when the sample sizes within groups are unbalanced, which is also expected since the support of the first group is half of the support of the second group, so unbalanced sample sizes are needed to obtain a similar density of data points in each *branch*. For models LC-3 and CC-3, similar results are obtained, with the estimated CMIE_m generally decreasing with the sample size, although sometimes increasing when going from unbalanced sample sizes to equal sample sizes. Again, a larger error concentration leads to a decrease of the estimated CMIE_m.

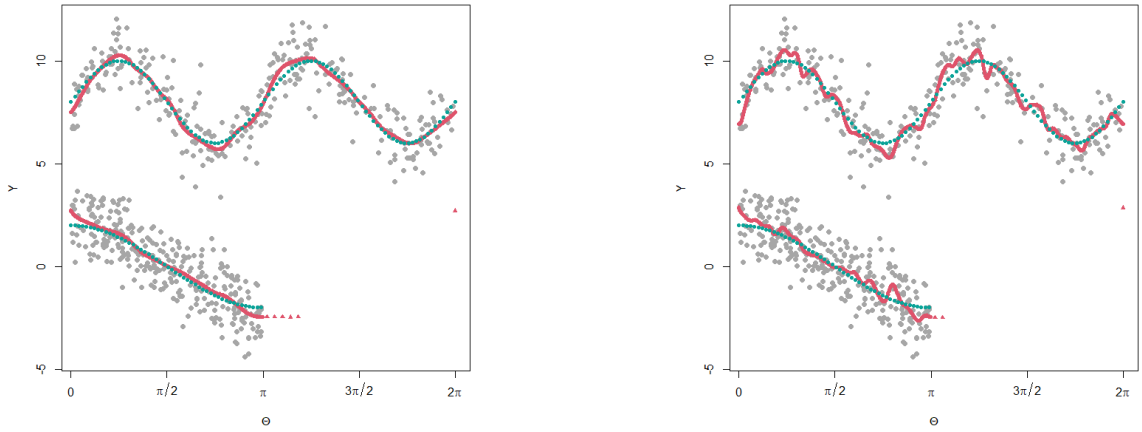


Figure 5.4: Simulated data from model CL-3 with the multimodal regression curve (dotted trace) and multimodal estimator (continuous trace and triangles) with smoothing parameters selected by cross-validation (left) and optimal smoothing parameters minimizing $MISE_m$ (right).

5.5 Analysis of the escape responses in larval zebrafish

In this section, the larval zebrafish data described in Appendix A.7 is analyzed. The angle of stimulus is considered as the predictor variable, Θ, and the direction of escape is taken to be the response variable, Φ. The observations corresponding to both variables are represented in Figure 5.5, both on the surface of the torus and on the plane.

In order to study how the escape response changes with the stimulus direction, one can consider the more classical approach of estimating the conditional mean of the escape direction, for example with the kernel regression estimator for a circular predictor and a circular response (see Section 2.1.2). This estimator was applied to the data, selecting the smoothing (concentration) parameter by cross-validation (see Oliveira et al., 2014b), and the estimation is represented in Figure 5.5 with a continuous trace. According to this estimator, the mean escape direction was ipsilateral when the robot approached the rostral side of the fish, contralateral when approaching the fish's left or right side of the bodies and around zero when approaching the caudal side of the bodies. This translates into the larval zebrafish escaping, on average, in a direction toward the predator when receiving the visual stimulus peripherally around the rostrum, escaping away from the predator if the stimulus was received laterally, and escaping in the direction the predator is coming if the stimulus is caught peripherally around the tail.

As it has been previously argued, estimating the conditional circular mean direction might not be the best approach for this kind of data, and multimodal regression might help to understand different features of the data. Thus, in order to obtain more insight into the relationship

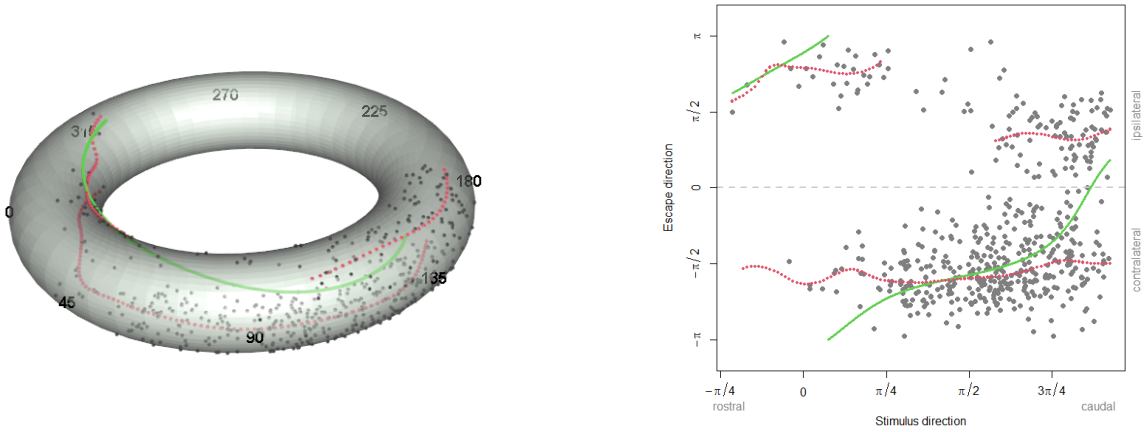


Figure 5.5: Representations on the torus (left) and on the plane (right) of the zebrafish data with the estimated regression multifunction (dotted trace) and the estimated mean regression function (continuous trace).

between the escape direction and the stimulus angle, the multimodal estimator for circular predictors and a circular response is applied and represented in a dotted trace in Figure 5.5. The smoothing parameters were obtained by applying the modal cross-validation criterion. Clearly from the bottom right panel of the figure, three different major trends in the data can be observed, feature which the more classical regression to the mean approach fails to capture. It can be seen that, when the robot appeared in the peripherical vision of the fish (both rostral and caudal sides), there were two preferred escape directions: one ipsilateral (toward the robot predator) and one contralateral (away from the threat). On the other hand, only one mode is estimated when the robot did not emerge through the peripherical vision of the fish, indicating a contralateral escape in this case.

Another advantage of the multimodal estimator is that it enables the construction of inferential tools for regression from a multimodal point of view, such as prediction sets. In what follows, a brief description is given on how to construct sets which predict, for a given prediction level $(1-\alpha)$, the regions where new data will fall. Adapting the methodology of Chen et al. (2016a) to the scenario with two circular variables, a theoretical uniform prediction set can be defined as

$$\mathcal{P}_{1-\alpha} = \left\{(\theta,\phi) : \theta \in \mathbb{T}, \phi \in \mathcal{A}_{\varepsilon_{1-\alpha}}(\theta)\right\},$$

where

$$\mathcal{A}_{\varepsilon_{1-\alpha}}(\theta) = \left\{\phi \in \mathbb{T} : \tilde{d}[\phi, M(\theta)] \leqslant \varepsilon_{1-\alpha}\right\}$$

and

$$\varepsilon_{1-\alpha} = \inf\left\{\varepsilon > 0 : \mathbb{P}\left(\tilde{d}[\Phi, M(\Theta)] > \varepsilon\right) \leqslant \alpha\right\}.$$

With this definition, it holds $\mathbb{P}(\Phi \in \mathcal{P}_{1-\alpha}) \geqslant 1-\alpha$. At sample level, following Chen et al.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

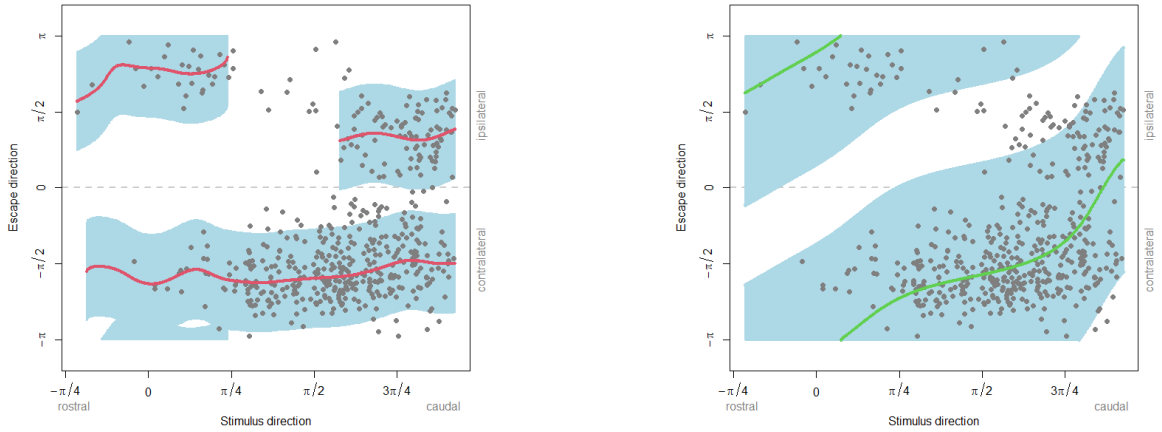


Figure 5.6: Left: planar representation of the zebrafish data with the estimated regression multifunction and uniform prediction sets for a prediction level of 0.90. Right: same representation with the nonparametric estimator of the mean regression function and uniform prediction bands with prediction level 0.90.

(2016a), these quantities can be estimated as

$$\widehat{\varepsilon}_{1-\alpha} = \text{Quantile}\left(\left\{\tilde{d}\left[\Phi_i, \widehat{M}(\Theta_i)\right], i = 1, \ldots, n\right\}, 1-\alpha\right),$$

$$\widehat{\mathcal{A}}_{\widehat{\varepsilon}_{1-\alpha}}(\theta) = \left\{\phi \in \mathbb{T} : \tilde{d}\left[\phi, \widehat{M}(\theta)\right] \leqslant \widehat{\varepsilon}_{1-\alpha}\right\}$$

and, then,

$$\widehat{\mathcal{P}}_{1-\alpha} = \left\{(\theta, \phi) : \theta \in \mathbb{T}, \phi \in \widehat{\mathcal{A}}_{\varepsilon_{1-\alpha}}(\theta)\right\}.$$

For a prediction level of $1 - \alpha = 0.90$, this uniform prediction set was estimated and represented in the left panel of Figure 5.6. Observe that the prediction set consists on three different prediction *bands*, each one corresponding to one of the estimated trends. The large dispersion of the data makes these prediction *bands* rather wide. The unknown boundary condition, discussed in Section 5.4.3, is clearly visible in the bottom left part of the plot, where a modal estimation and a prediction set are located in a region without any data. The resulting prediction set is even more satisfactory if it is compared to the uniform prediction band obtained with the circular mean regression estimator of Di Marzio et al. (2013), which is shown in the right panel of Figure 5.6 for a prediction level of 0.90. The band was computed by using 1000 bootstrap resamples of the data using bootstrapping by pairs. The plot clearly shows that the uniform band constructed with the estimator of the conditional circular mean is considerably wider than the set constructed with the modal estimator.

In summary, the multimodal regression estimator allows to estimate the two preferred directions of escape when a fish receives the visual stimulus of a predator through its periph-

erical vision, that is, coming toward the front or tail ends of its body. It also shows how, if the predator approaches the fish through the left or right sides of the body (where the fish's eyes are located), the direction of escape is in an opposite direction to the threat.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 6

Conclusions, discussion and further research

The focus of this dissertation was the introduction and study of nonparametric methods for regression, when dealing with circular data. More specifically, the main goals achieved within this thesis are the following:

- The construction and study of nonparametric tests to determine simplifying hypothesis in the context of *mean* regression with a circular covariate and/or a circular response variable. Concretely, the proposal of testing procedures to ascertain the significance of the predictor variable and to determine the equality or parallelism of the regression curves across different groups. This goal corresponds to the contents of Chapter 2.

- The nonparametric estimation of conditional characteristics of a random variable through the maximization of the local likelihood, when the covariate presents a circular nature. In addition, the study of the bias and variance of the proposed estimators, as well as the selection of the smoothing parameter in practice. The development of this point is collected in Chapter 3.

- The joint estimation of the mean and dispersion functions, conditioned by a circular covariate, in situations where the data presents over- or under-dispersion (or both) with respect to the exponential family. The development and computational analysis of this method, in addition to its application to a motivational real data problem, are contained in Chapter 4.

- The introduction and study of statistical tools to estimate the conditional local modes in the context of regression with a circular covariate and/or a circular response. Additionally, the investigation of the estimators' asymptotic properties and the analysis of their finite sample performance. The results concerning this goal are provided in Chapter 5.

Kernel regression methods which allow to flexibly estimate the conditional mean of a response variable in the context of circular data had been previously introduced by Di Marzio et al. (2009) and Di Marzio et al. (2013). These estimators are the building blocks of the contents of Chapter 2, where new proposals for significance and ANCOVA tests have been introduced. The circular nature of the response and/or the covariate implies important changes in the construction of the different tests with respect to the methods tailored for real-valued data. Throughout Chapter 2, the difficulties that arise from the consideration of this non-euclidean support have been overcome. As shown in the simulation study, a satisfactory performance

in practice has been assessed with all the testing proposals. In addition, it was shown that the smoothing parameter (bandwidth or concentration, depending on the type of kernel) may present a relevant impact on the outcome of the tests. This issue can be solved by exploring a range of smoothing parameters, as done in the illustration with the real data sets.

In Chapter 3, another estimation method for regression curves involving circular covariates was considered, consisting on the construction and maximization of the local log-likelihood function after the appropriate selection of the response variable's distribution. This method coincides with the estimation procedure considered in Chapter 2 for circular predictors and normally distributed responses. However, the approach of Chapter 3 allows to nonparametrically estimate any conditional parameter of interest in the response's distribution. Despite the fact that in most cases the proposed estimators do not have an explicit form, a way of computing the bias and variance of the estimators was derived, which allows to complete diverse inferential tasks, such as the construction of confidence bands for the regression functions. The computation of the bias and variance also allows to select an optimal concentration parameter in practice by minimizing an approximated MISE. However, the computation of the bias implies the selection of a pilot smoothing parameter. This task was also attained by constructing and minimizing an error criterion, and the properties of this new criterion were also studied. The performance of the estimators and the smoothing parameter selection were also studied in practice through an extensive simulation study, where it was shown that the minimization of the approximated MISE leads to less undersmoothed estimators than the ones obtained with a cross-validation smoothing parameter. The approach taking in Chapter 3 constitutes a general framework for regression with circular covariates since, apart from covering a wide range of practical situations, as stated in Section 3.6 this method can be easily adapted to more complex scenarios, such as partially linear or additive models.

In Chapter 4, the problem of studying the number of spikes in a macaque's brain cell, conditioned on the direction of a visual stimulus, illustrated and motivated the need of estimating, jointly with the mean, an additional dispersion parameter allowing to model over- or under-dispersion with respect to the exponential family. The double exponential family, introduced by Efron (1986), was presented and the local log-likelihood method, studied in Chapter 3, was employed to carry out a joint estimation of the mean and dispersion functions with circular covariates. The performance of the estimators in practice was analyzed in a simulation study, where a satisfactory performance was achieved, specially given the difficult task of estimating the dispersion.

Lastly, the estimation of the circular modal regression multifunction was considered in Chapter 5. A nonparametric method to study the local modes of a response variable given a covariate (where both the covariate and the response can be circular variables) was introduced. This multimodal regression method allows to estimate the most likely values of the response given the covariate, providing a more suitable insight on the variables relationship beyond mean regression settings. The properties of this proposal have been studied both

computationally through a simulation study and theoretically, where the asymptotic convergence rates of the estimators were derived. Despite the fact that the proposed estimators are tailored for a regression framework, the study of the properties of the cylindrical and toroidal kernel density estimators and their derivatives was the key constituent for the derivation of the asymptotic convergence rates for the multimodal regression estimators.

The importance in practice of the contributions of this thesis has been highlighted with a substantial number of real data illustrations and applications encompassing different fields, which manifests the need of flexible regression methods involving circular data. Owing to the direct applicability of many of the methods proposed in this thesis to different disciplines, `R` functions have been programmed for all the new techniques, and have been incorporated.

Connections between chapters. All the works that compose this dissertation deal with inference for regression models involving circular data where, in each chapter, the regression function (or multifunction) is regarded as a different conditional characteristic of the response. The local likelihood approach of Chapter 3 seems like a natural extension of the regression setting in Section 2.1.1 to the case of a more general response variable. The local likelihood estimation is also employed in Chapter 4 to perform a joint estimation of mean and dispersion functions. Contrastingly, the novelty of multimodal regression and its distinct estimation approach might make it seem like a disconnected topic from the rest of the thesis. Here, some remarkable interconnections with the other chapters are highlighted.

First, it is interesting to consider the relationship between the multimodal estimator of Chapter 5 and the estimation of conditional characteristics in Chapters 3 and 4. The modeling and estimation of the multimodal regression function can be considered as a joint estimation of conditional characteristics (the local modes), except that the distribution of the response variable is not specified. Thus, one can consider the multimodal regression approach as a fully nonparametric method, while the local likelihood method studied in Chapters 3 and 4 is actually a semiparametric approach. Moreover, it is worth noting that some of the most important elements of the proofs concerning the consistency of cylindrical and toroidal kernel density estimators in Chapter 5 are shared with the results derived in Chapter 3.

Finally, it can also be judged that, in practice, the different *branches* in the context of multimodal regression may be caused by the effect of an unknown categorical variable. Consequently, multimodal regression can be employed to model situations where an ANCOVA model (studied in Chapter 2) could be applied if the information about the different groups was known.

USC

Further research

Given the previous discussion, some further research lines, that would allow to broaden the studies carried out in this dissertation, have been identified. Some of the topics are common to all chapters, such as the extension to hyperspherical data or the derivation of other smooth-

ing parameter selection methods, while others are more specific to a certain methodological approach presented in this manuscript.

Extensions to higher dimensions. Hyperspherical data are defined on a sphere of arbitrary dimension, $\mathbb{S}^d \subset \mathbb{R}^{d+1}$, for $d \in \mathbb{N}$. A key tool inherent to all methodologies described in this dissertation is the use of circular kernels, which for the theoretical derivations were assumed to be of the form in (3.3). For an extension to hyperspherical variables, a directional kernel can be considered, where condition (3.3) can be replaced by equation (1) in García-Portugués et al. (2016). The von Mises-Fisher density, the generalization of the von Mises to the hyperspherical case, can be employed in practice. The theoretical derivations in Chapters 3 and 5 can be carried out in a similar way to account for the generality of hyperspherical variables, where the sine-polynomial expansion in equation (C.2) is substituted by a projected multivariate Taylor expansion, obtained after performing a radial projection of the target function from $\mathbb{S}^d$ to $\mathbb{R}^{d+1}$. Replacing the change of variables $r = \kappa(1 - \cos\varphi)$ in the proofs of Appendix C.2 and C.3 by the change of variables stated in pages 91-93 of Efthimiou and Frye (2014) and Lemma 2 of García-Portugués et al. (2013) leads to similar derivations.

Local smoothing parameters. Another interesting future topic, inherent to all models considered in this work, is the use of local smoothing parameters. Situations where the relationships between the variables present a complex structure would benefit from variable smoothing parameters. Consider, for example, the pm10 particle data set, described in Appendix A.6 and analyzed in Sections 3.5 and 3.6. In the center panel of Figure 3.13 it is clearly visible that the underlying density of the predictor variable is far from uniform, and the analysis of this data set could be more complete with the use of a variable concentration parameter. Nonetheless, it must be taken into account that the use of variable smoothing parameters would considerably increase the computational cost of the methods.

Hypothesis testing in the local likelihood setting. Also with the objective of improving the analysis of the real data sets considered in Section 3.5, a generalization of the no-effect test of Section 2.2 to scenarios with a general response would be of great interest. With the notation of Chapter 3, the formulation of the hypotheses in this context is

$$\begin{aligned} H_0 &: g(\Theta_i) = \gamma, \quad \gamma \in \mathbb{R}, \ \forall\, i \in \{1, ..., n\} \\ H_1 &: g(\Theta_i) \neq \gamma, \quad \text{for some } i \in \{1, ..., n\}. \end{aligned}$$

In order to construct the test statistic, a possibility is to replace the squared residuals in (2.11) by deviance residuals (see equation (4.7) for the exponential family).

Local likelihood estimation with a circular response. It is worth noting that, although in Chapters 3 and 4 the possibility of a circular response variable was not explicitly contemplated, the local likelihood approach could also be applied in this setting, after selecting a

suitably circular conditional distribution. Consider, as an example, a real-valued predictor X and a circular response Φ, such that $(\Phi|X = x_0) \sim \text{vM}[\mu(x_0), \tau]$ and $g(\cdot) = \eta[\mu(\cdot)]$, with $\eta(\cdot)$ being an invertible link function. Assuming that $g(\cdot)$ is at least p times differentiable, for data points X_i near x_0, the target function can be approximated as

$$g(X_i) \approx g(x_0) + g'(x_0)(X_i - x_0) + \ldots + \frac{g^{(p)}(x_0)}{p!}(X_i - x_0)^p = \boldsymbol{X}_i^T \boldsymbol{\beta},$$

where $\boldsymbol{X}_i = [1, X_i - x_0, \ldots, (X_i - x_0)^p]^T$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$, with $\beta_\nu = g^{(\nu)}(x_0)/\nu!$, for $\nu = 0, \ldots, p$. Maximizing the local log-likelihood $\mathcal{L}_p(\boldsymbol{\beta}; \boldsymbol{h}, x_0)$ would be equivalent to maximizing

$$\sum_{i=1}^{n} \cos\left[\Phi_i - \eta^{-1}(\boldsymbol{X}_i^T \boldsymbol{\beta})\right] L_h(X_i - x_0). \tag{6.1}$$

The maximization of (6.1) does not have an explicit solution in general, which is a drawback with respect to the regression method in Section 2.1.2. Nonetheless, interestingly, both methods coincide for $p = 0$ (a local constant fit). In this case, although the bias and variance of the estimator could be computed in an analogue way to the derivations of Section 3.2, the CRSC criterion studied in Section 3.3 should be modified. A deeper study of this proposal could lead to the derivation of a new rule for selecting the smoothing parameters in the context of regression with a circular response. In addition, other conditional distributions can be selected for the response variable. For instance, assuming a projected normal distribution would lead to a nonparametric version of the regression method proposed by Presnell et al. (1998) and discussed in Section 1.2.3.

Further study on multimodal regression. The research of Chapter 5 would also benefit from a deeper study of smoothing parameter selection methods. As it was discussed in Section 5.3, there are not any theoretical justifications for the modal cross-validation method. In this sense, a thorough study of its properties or a new approach for selecting the smoothing parameters in the multimodal regression context seems necessary, not just in the circular setting, but also when working with real-valued variables (see Chen, 2018; Alonso-Pena, 2020). For this aim, a possible approach is to investigate the use of local polynomial-type estimators of the conditional density (Fan et al., 1996; Cattaneo et al., 2022), that could allow the selection of adequate smoothing parameters for the estimation of the first partial derivative of the conditional density.

Finally, a further study of the unknown boundary condition of multimodal regression (see Section 5.4.3) would lead to an improvement of the practical use of these methods. In the simulation study of Section 5.4, the parts of the estimation outside of the support of each group were removed. This was only possible in the simulated setting, where the information about the underlying groups was known. For practical purposes, a potential approach in this case would estimating the α-shape of the data (Edelsbrunner et al., 1983; Edelsbrunner and

Mucke, 1984), accounting for the cylindrical/toroidal nature of the observations and then restrict the estimation to the interior of the estimated α-shape. Figure 6.1 presents a realization of simulated data form models CL-3 (left) and CC-3 (right) described in Section 5.4, with a two-dimensional α-shape estimation, obtained with R package `alphahull` (Pateiro-Lopez and Rodriguez-Casal, 2022). In the left panel, the chosen value of α is very large, producing a convex estimation. On the other side, the value of α in the right panel is selected as a smaller number, yielding a non-convex estimation.

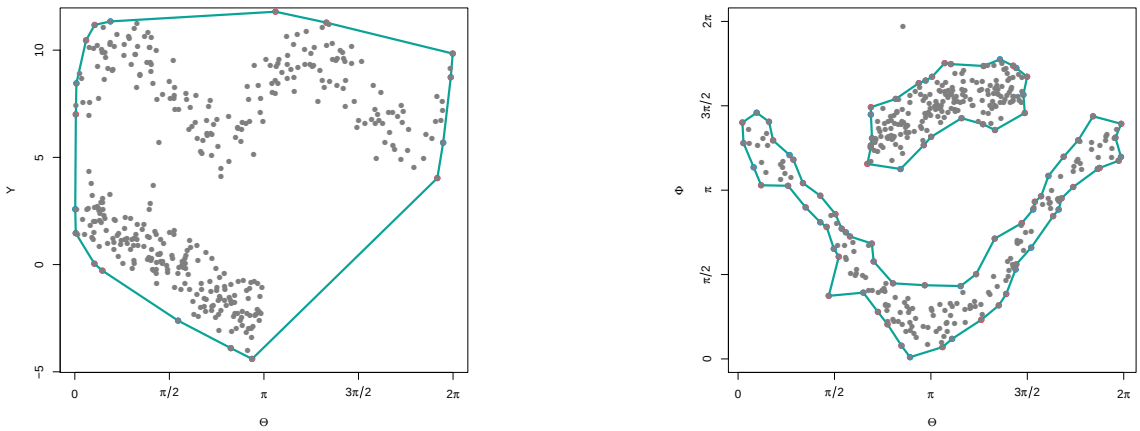


Figure 6.1: Simulated data from models CL-3 (left) and CC-3 (right) with the estimated α-shape, with different values of α.

However, the α-shape estimations presented in Figure 6.1 do not take into account the circular nature of the variables. In the left panel, the predictor variable is circular, and the left and right margins of the plot are actually the same, but the estimator in $\mathbb{R}^2$ creates artificial boundaries at $\theta = 0 = 2\pi$. For the right panel, both the predictor and the response are circular, so not only the left and right margins coincide, but also the top and bottom margins. It can be seen that the the estimator in the right panel divides the data into two sets, but one point is left out, since artificial boundaries are also created at $\phi = 0 = 2\pi$. Hence, a future line of research includes adapting this kind of estimators to include cylindrical and toroidal data. After a suitable α-shape estimator is obtained (with an appropriate selection of α), the multimodal estimators can be restricted to the interior of the sets, ensuring that there are not any estimated modes in regions without data.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Appendices

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

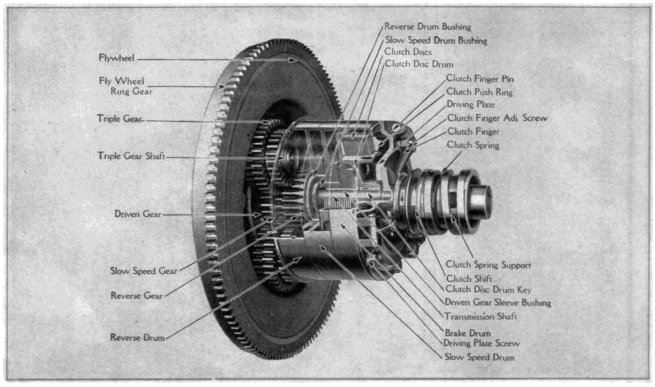
Appendix A

Real data sets

Throughout this dissertation, all the proposed methodology was illustrated with a variety of real data sets. This appendix contains descriptions of the seven data sets used along this manuscript, including some details on the experiments that lead to the collection of the data.

A.1 Flywheels

Flywheels are mechanical devices used in the the automotive industry, designed to regulate an engine's rotation. A flywheels is, essentially, a heavy wheel attached to a rotating shaft (see Figure A.1) and it is used to increase the machine's momentum, in order to produce stability and to store rotational energy in an efficient way. In vehicles production, it is necessary to balance flywheels to ensure that the rotation transmits minimal vibration to the vehicle, since such vibration can lead to the malfunctioning of the system. When correcting the balance, the response obtained is cylindrical: an angular component measuring the angle of imbalance and a linear component evaluating the magnitude of the correction required to balance the flywheel. Modeling the relationship between the angle and the magnitude of correction can be helpful for a better understanding of the balancing process, leading to the minimization of the costs by creating more efficient designs. In addition, a metallic molding is used in the production process. The metal's purity and density can influence both the angle and the correction magnitude.



USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figure A.1: Transmission system of a car, including a flywheel on the left, from Wikimedia Commons (2013).

The flywheel data set, given in Table 1 of Anderson-Cook (1999), contains measurements of the angles of imbalance of 60 flywheels, as well as the measurements of the corrections required (in inch-ounces). Four different kinds of metals were employed, with 15 flywheels corresponding to each type of metal. This data set has been studied in Section 2.5 considering a nonparametric ANCOVA model.

A.2 Sand hoppers

Sand hoppers (Talitridae family) are amphipod crustaceans (see left panel in Figure A.2) known for their typical "hopping" movement, from which their common name is derived. The behavior of sand hoppers has been studied for years: they usually remain burrowed in the wet sand during the day and go out to the surface from sunset to sunrise. When they are displaced during the day due to the drying up of the sand layers, they return to the wet sand near the water. Several authors have linked the direction of sand hoppers movements to different factors, such as the slope of the beach or landscape features. Pardi and Ercolini (1986) stated that in Mediterranean shores the movement on sand hoppers relies on sun orientation, while this does not happen in Atlantic shores. Differences in sand hoppers orientation also arise between those living on tidal and non-tidal beaches, because of the risk of being swept away.

The sand hoppers data set used in this manuscript was first studied in Scapini et al. (2002). It contains observations involving two sympatric sand hopper species (*Talitrus saltator* and *Talorchestia brito*). The study was conducted in an exposed and non-tidal beach in Tunisia, the Zouara beach. The main goal of the experiment was to analyze the direction of movement given different conditions. In order to record the data, the experimenters used two different circular arenas with 72 cross pit-fall traps each, placed at the circumference. The animals were released in the arenas, and once they made an orientation choice they were caught in one of the traps, which were separated from each other by an angle of 5°. One of the arenas allowed the view of both the sky and the landscape, while in the other the landscape was screened off, so that only the sky was visible. In addition to the direction of movement, other variables were recorded, such as sun azimuth, the species of sand hopper, the type of arena, the sex of the animal and several climatic variables (air temperature, air relative humidity and atmospheric pressure). The experiments were conducted in two different seasons (April and October). The total sample size was $n = 1828$.

With the objective of studying the variation of directions taken by the sand hoppers in the presence of several factors, Scapini et al. (2002) used the Projected Multivariate Linear Model (PMLM) proposed by Presnell et al. (1998) (see Section 1.2.3). However, when using this model the authors considered all the predictor variables as discrete variables, factorizing the continuous ones such as the sun azimuth or the climatic variables.

Marchetti and Scapini (2003) also analyzed this data set, focusing only on the data obtained for the sand hoppers of the species *Talitrus saltator*. In this case the authors considered the variable sun azimuth as continuous, and fitted regression models to explain the movement





Figure A.2: Left: sand hopper of the species Talitrus saltator, from Wikimedia Commons (2017). Right: image of a Gryllus bimaculatus, from Wikimedia Commons (2010).

direction of the animals. As an exploratory tool, the authors fitted a nonparametric regression model with the movement directions as a response variable and the sun azimuth as a predictor variable, and tried to graphically assess the differences between the screened and the unscreened groups. However, the regression smoothers used are meant for both linear responses and linear predictors. The authors also applied the PMLM as in Scapini et al. (2002), but this time considering the sun azimuth as a continuous real-valued variable. The data set has also been studied by Meilán-Vila et al. (2021, 2022) and Saavedra-Nieves and Crujeiras (2021b). The data set was kindly provided by Professor Felicita Scapini, from the University of Florence (Italy), and is now currently available at `R` package `HDiR` (Saavedra-Nieves and Crujeiras, 2021a).

A.3 Crickets

Gryllus bimaculatus is a species of cricket in the Gryllidae family, also known as *African* or *Mediterranean field cricket* (see right panel in Figure A.2). The oriented escape behavior of crickets from an air current stimulus is considered to represent escape behavior from an approaching predator. In addition, the escape behavior in crickets is similar to that of animals of other species, which makes the study of crickets useful in order to get more insight in escape responses for different animals.

The crickets data set used in Chapter 2, which was kindly provided by Professor Hiroto Ogawa from Hokkaido University (Japan), contains 182 measurements of the escape directions of crickets after an air-puff stimulus. Other variables included are the angle of the stimulus and the type of escape response (running or jumping). In the experimental study, described in Sato et al. (2019), crickets were placed in the center of a circular arena and an air current was applied horizontally from a nozzle in the wall. A high-speed digital camera was used to record the cricket's movement at 120 fps.

A.4 Human motor resonance

The primary motor cortex (M1) is a region of the brain that is located (in humans) in the dorsal portion of the frontal lobe (see the motor area in Figure A.3) and is the primary region of the motor system. The observation of others' movements evokes subliminal motor resonance response in the observers' M1 (Puglisi et al., 2017). In order to investigate the motor resonance responses in humans, Puglisi et al. (2017) carried out a behavioral experiment where subjects were requested to observe a movement of a rhythmic hand flexion-extension in front of them. A metallic and oscillating platform was used to record all hand movements. For each angular position of the hand, the H-reflex technique was used to quantitatively measure the resonance response. In total, the human motor resonance data set includes 70 observations containing the angular position (circular variable) and the resonance response (real-valued variable). The data set is available as online supplementary material of Puglisi et al. (2017).

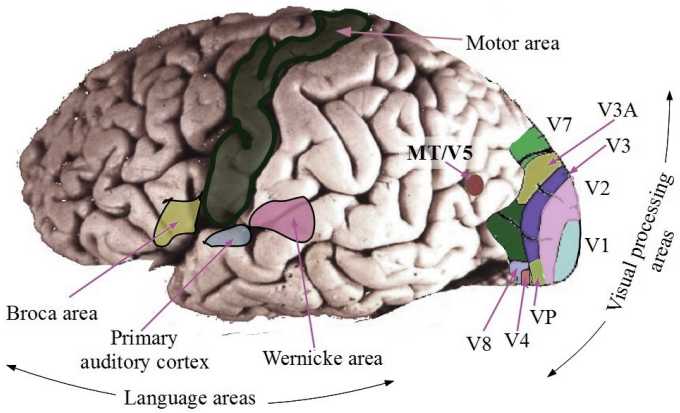


Figure A.3: Image of a human brain, with different areas of cortex including the motor area and the middle temporal visual area, from Wikimedia Commons (2006).

A.5 Spike count

The middle temporal visual area (known as MT or V5; see Figure A.3) is a region of the extrastriate visual cortex which is thought to play a key role in the perception of motion and in the conduct of some eye movements (Born and Bradley, 2005). Many authors in the neurological field have studied the response properties of neurons in the MT/V5 region, showing that this area contains a high proportion of direction-selective neurons (see, for example, Carandini and Ferster (2000) or Ringach et al. (2003)). The study of how the number of spikes in a MT/V5 neuron is tuned to the direction of a visual stimulus has attracted considerable attention over the years, and some more recent studies suggest that the variability of the number of spikes might also be tuned by the directional stimulus (Lombardo et al., 2018).

The spike count data set considered in this dissertation was obtained from the website of the Neural Signal Archive (http://www.neuralsignal.org/), which belong to an experimental study with an anesthetized and paralyzed adult male macaque monkey (*Macaca nemestrina*). This macaque was presented with a visual stimulus consisting on moving dots with different moving directions and the neural activity in some MT/V5 cells was measured. The objective of the study, which is fully described in Kohn and Movshon (2003), is to explore how the neural activity changes with the direction of the stimulus. More specifically, the data set employed in this thesis, which was also studied by Stevenson (2016), consists of 68 observations of the total number of spikes per trial in one MT/V5 cell and the direction of the visual stimulus.

A.6 Pm10 particles

Pm particles (particulate matter) are very small inhalable particles which are an indicator of air pollution. Particularly, pm10 particles are those with a diameter of 0.1 micrometers or less, which are small enough to get into people's throat and lungs, and can cause serious health issues. Common sources of pm10 particles include dust, smoke from fires, car exhausts and industry.

The pm10 data set used in this dissertation includes data of pm10 particles in the city of Pontevedra, Spain. Of special interest is the location of a controversial pulp factory in the surroundings of the city, approximately 2.5 km south-west of Pontevedra (see Figure A.4). Apart from the measurements of pm10 particles, the data set includes recordings of the wind direction and speed. All the measurements were taken in a meteorological station in Pontevedra during the year 2019. The concentration of pm10 particles is measured in $\mu g/m^3$, wind speed was measured in km/hour and wind direction was measured in degrees, with 0 degrees representing the north direction and a clockwise sense of rotation. Although the pollution data is recorded hourly and the meteorological data is measured on a 10-minute base, a subsample of the data was used, with the final sample size being $n = 1156$. The data set was obtained from Meteogalicia's web page (https://www.meteogalicia.gal/).

A.7 Zebrafish

Zebrafish (*Danio rerio*) are freshwater fish, native to South Asia, which are popular aquarium fish. They are named for their uniform horizontal stripes on the side of the body in adult age, which resemble zebra's stripes, although larval zebrafish are transparent (see Figure A.5). Zebrafish are widely used in scientific research, as a vertebrate model organism, and are specially valuable in the field of genetics.

However, zebrafish are also employed for research in other fields such as neuroscience, and more specifically the study of escape behavior in larval zebrafish when startled by different stimulus (see Colwill and Creton, 2011). In particular, the motivation of the analysis

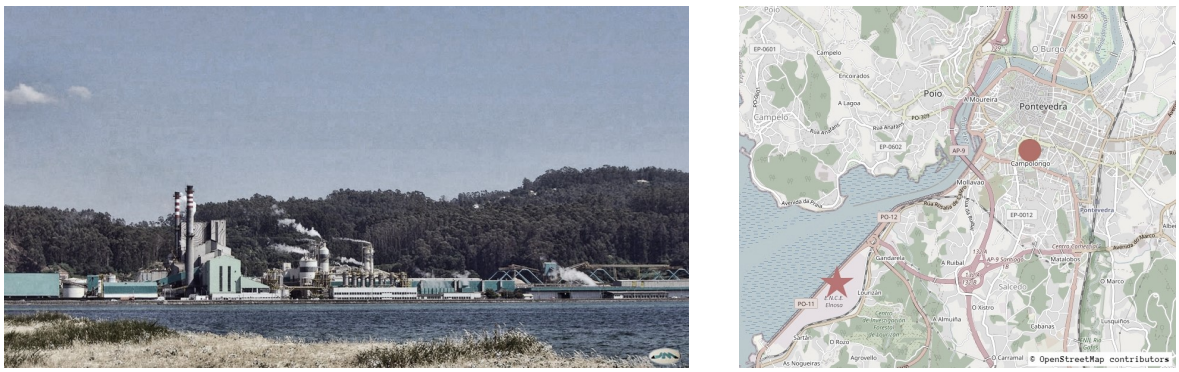


Figure A.4: Left: picture of the pulp factory in the surroundings of Pontevedra, from Wikimedia Commons (2014). Right: Map of the area surrounding the city of Pontevedra (obtained under license Open Database License, ODbL), with a circle indicating the position of the meteorological station and a star representing the position of the factory.

in Chapter 5 is to study the escape responses of larval zebrafish when sensing the visual stimulus of a chasing predator. Although logic would suggest that the animals should turn immediately away from predators, this is not always the case, and the study of their escape directions becomes necessary to understand the behavior of the fish and validate theoretical escape strategies. The data set employed in Chapter 5 is obtained from Nair et al. (2017a), concerning an experimental study on the escaping behavior of larval zebrafish, where the situation of a predator chasing the larvae was simulated.

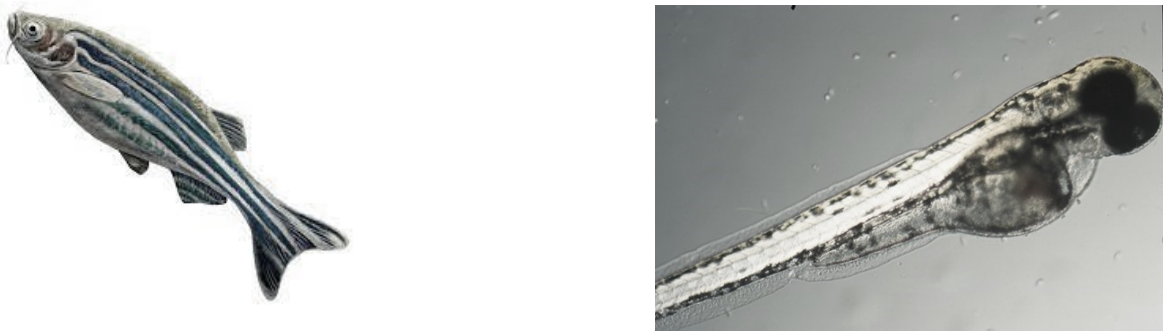


Figure A.5: Left: image of an adult zebrafish, from Wikimedia Commons (2014b). Right: image of a larval zebrafish, from Wikimedia Commons (2014a).

In this case study, a collection of $n = 502$ larval zebrafish were kept in an aquarium which disposed of high quality moving cameras used to collect the data regarding the movements of the fish. A robot was disguised with the body of a dead adult zebrafish and was used to imitate a real predator. This robot was kept still in the middle of the aquarium until the beginning of the experiment. At this point, and with the goal of studying the escape responses of the larval zebrafish when startled by a visual stimulus, the robot was moved at a constant speed, emulating a real predator. The variable determining the direction of escape from the

robot predator is the variable of interest in the study of Chapter 5, and the objective in this thesis is analyzing how this escape direction changes with the angle from which the fish were approached by the predator. Thus, the zebrafish data set is analyzed from a regression perspective, where the angle of approach or direction of stimulus is the predictor variable and the escape direction is the response variable.

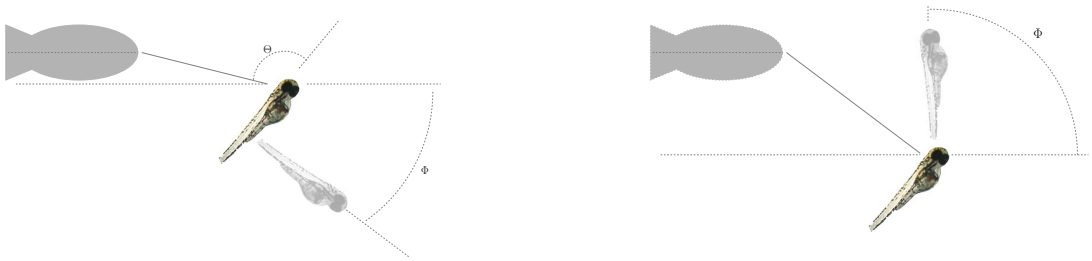


Figure A.6: Diagrams of the zebrafish experiment, where the gray shape represents the robot predator, the dark colored larvae represents the initial position of the fish and the light colored larvae represents its position after the escape response. Own elaboration using the image from Wikimedia Commons (2014a). Left: contralateral escape. Right: ipsilateral escape.

Figure A.6 contains illustrative diagrams of the experiment showing how the variables of the study were defined, where the gray shape represents the robot predator, the dark larval zebrafish represents the original position of the fish and the light colored larvae represents the direction of the fish after initiating the escape response. The angle of stimulus, Θ, which is the angle in which the fish were approached by the robot predator, was measured (anticlockwise) with respect to the initial orientation of the zebrafish's body, as indicated in the left panel of Figure A.6. The support of this variable is determined by the animal's vision field. Values of the stimulus direction close to $-\pi/4$ indicate that the robot approached the fish through the rostral (near the rostrum or front end) side of the body. On the other hand, stimulus directions close to π show that the robot approached the fish through the caudal side (near the tail). Approaching the fish through the rostral or caudal sides (front or tail) conveys that the robot appeared in the fish's peripherical vision. The direction of escape was considered as the response variable, Φ. This variable was measured (anticlockwise) with respect to the heading direction of the predator at the moment of the visual contact, as it can be seen in Figure A.6. An escape response in the range $(-\pi, 0)$, such as the one depicted in the left panel of Figure A.6, is considered to be a contralateral escape. On the contrary, an escape direction in the range $(0, \pi)$ is known as an ipsilateral escape, since the larvae would be escaping to the same side where its predator lies, as represented in the right panel of Figure A.6. More technical details about the experiment can be found in Nair et al. (2017b), where the zebrafish data set was initially analyzed and the authors tried to model the relationship between the escape direction and the angle of stimulus through simple linear regression.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Appendix B

Additional simulation results

This appendix gives additional simulation results complementing those of the main body of the text. Section B.1 supplies simulation results to support the conclusions in Chapter 2, where the considered simulation scenarios are the ones described in Section 2.4. Section B.2 contains supplementary material to Chapter 5: first, the modification of the modal cross-validation criterion employed in the main simulations is described; second, the empirical performance of this modification is explored via simulations.

B.1 Supplement to Chapter 2

The contents of this section concern the simulation study conducted in Section 2.4. Simulation results regarding the models analyzed for the no-effect tests (Section 2.4.1) and ANCOVA tests (Section 2.4.2) are provided. Tables B.1 and B.2 contain percentages of rejection of the no-effect tests with a significance level of $\alpha = 0.10$ and $\alpha = 0.01$, respectively. In addition, Tables B.3 and B.4 show percentages of rejection for the ANCOVA test for the same significance levels. Results for $\alpha = 0.05$ are presented in Section 2.4.

Additionally, for the *circular-linear* regression case, results for the no-effect and ANCOVA tests in Section 2.4 are obtained considering a shifted and scaled χ^2 distribution for calibration when using normal errors and results for those tests calibrated by the bootstrap approach when using exponential errors. In this section, percentages of rejection for the same tests when switching the distribution of the errors are presented: Table B.5 presents results for the no-effect test calibrated with the χ^2 distribution applied to exponential errors and the test calibrated by bootstrap applied to normal errors. In the same line, Table B.6 collects percentages of rejection for the ANCOVA tests calibrated with the χ^2 distribution applied to exponential errors and the bootstrap version of the tests applied to normal errors.

To conclude, the performance of the ANCOVA tests when using values of the smoothing parameter different from the ones obtained by cross-validation is also studied. Table B.7 contains percentages of rejection for the ANCOVA tests when using smoothing parameters which undersmooth the regression curves. The finite sample performance of the tests when oversmoothing the regression curves is displayed in Table B.8.

Table B.1: Percentages of rejection (for $\alpha = 0.10$) obtained with the no-effect tests using different smoothing parameters. Results for the first value of β show empirical size, whereas results for the other values of β show empirical power.

Circular-linear regression. χ^2 calibration. Normal errors									
	$4cv$			cv			$cv/8$		
n	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$
50	0.074	0.384	0.898	0.166	0.584	0.980	0.100	0.280	0.736
100	0.084	0.726	1	0.174	0.852	1	0.108	0.430	0.990
250	0.132	0.998	1	0.200	1	1	0.114	.920	1
400	0.128	1	1	0.204	1	1	0.124	0.996	1
Circular-linear regression. Bootstrap calibration. Exponential errors									
	$4cv$			cv			$cv/8$		
n	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$
50	0.110	0.406	0.902	0.164	0.762	0.994	0.090	0.584	0.972
100	0.098	0.712	0.998	0.182	0.956	1	0.098	0.896	1
250	0.126	0.992	1	0.188	1	1	0.118	1	1
400	0.120	1	1	0.188	1	1	0.128	1	1
Linear-circular regression. Von Mises errors									
	$cv/4$			cv			$4cv$		
n	$\beta=0$	$\beta=0.5$	$\beta=1$	$\beta=0$	$\beta=0.5$	$\beta=1$	$\beta=0$	$\beta=0.5$	$\beta=1$
50	0.090	0.522	0.996	0.142	0.618	0.868	0.108	0.590	0.856
100	0.124	0.836	1	0.174	0.894	0.990	0.126	0.882	0.986
250	0.110	0.990	1	0.146	0.996	1	0.100	0.996	1
400	0.110	1	1	0.164	1	1	0.122	1	1
Circular-circular regression. Von Mises errors									
	$4cv$			cv			$cv/8$		
n	$\beta=0$	$\beta=0.35$	$\beta=0.5$	$\beta=0$	$\beta=0.35$	$\beta=0.5$	$\beta=0$	$\beta=0.35$	$\beta=0.5$
50	0.102	0.408	0.670	0.216	0.598	0.828	0.144	0.424	0.614
100	0.094	0.746	0.976	0.158	0.874	0.994	0.118	0.626	0.910
250	0.088	0.996	1	0.168	1	1	0.110	0.976	1
400	0.138	1	1	0.210	1	1	0.122	1	1

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table B.2: Percentages of rejection (for $\alpha = 0.01$) obtained with the no-effect tests using different smoothing parameters. Results for the first value of β show empirical size, whereas results for the other values of β show empirical power.

Circular-linear regression. χ^2 calibration. Normal errors									
	$4cv$			cv			$cv/8$		
n	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$
50	0	.018	0.832	0.008	0.196	0.832	0.006	0.036	0.370
100	0.004	0.296	0.998	0.020	0.488	0.998	0.010	0.134	0.872
250	0.008	0.928	1	0.020	0.978	1	0.012	0.648	1
400	0.010	0.998	1	0.028	1	1	0.012	0.898	1
Circular-linear regression. Bootstrap calibration. Exponential errors									
	$4cv$			cv			$cv/8$		
n	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$
50	0.010	0.100	0.582	0.020	0.370	0.930	0.006	0.162	0.774
100	0.004	0.284	0.980	0.014	0.770	1	0.004	0.544	0.996
250	0.008	0.916	1	0.024	1	1	0.016	0.992	1
400	0.008	0.994	1	0.020	1	1	0.008	1	1
Linear-circular regression. Von Mises errors									
	$cv/4$			cv			$4cv$		
n	$\beta=0$	$\beta=0.5$	$\beta=1$	$\beta=0$	$\beta=0.5$	$\beta=1$	$\beta=0$	$\beta=0.5$	$\beta=1$
50	0.006	0.204	0.934	0.018	0.260	0.512	0.014	0.252	0.504
100	0.008	0.494	1	0.016	0.570	0.906	0.010	0.560	0.900
250	0.006	0.928	1	0.020	0.970	1	0.014	0.962	1
400	0.008	0.990	1	0.016	0.994	1	0.008	0.994	1
Circular-circular regression. Von Mises errors									
	$4cv$			cv			$cv/8$		
n	$\beta=0$	$\beta=0.35$	$\beta=0.5$	$\beta=0$	$\beta=0.35$	$\beta=0.5$	$\beta=0$	$\beta=0.35$	$\beta=0.5$
50	0.008	0.068	0.198	0.026	0.186	0.448	0.010	0.106	0.266
100	0.012	0.268	0.736	0.014	0.484	0.902	0.004	0.242	0.622
250	0.008	0.918	1	0.008	0.964	1	0.008	0.784	0.998
400	0.014	1	1	0.020	1	1	0.010	0.990	1

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table B.3: Percentages of rejection (for $\alpha = 0.10$) obtained with the ANCOVA tests using the smoothing parameters obtained by cross-validation. Results for the first value of β show empirical size, whereas results for the other values of β show empirical power.

Circular-linear regression. χ^2 calibration. Normal errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
50	50	0.090	0.646	0.952	0.094	0.698	0.966
50	100	0.090	0.822	0.992	0.104	0.826	0.992
100	100	0.116	0.968	1	0.096	0.972	1
100	250	0.102	0.998	1	0.104	0.992	1
250	250	0.090	1	1	0.100	1	1
Circular-linear regression. Bootstrap calibration. Exponential errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
50	50	0.102	0.844	0.994	0.098	0.192	0.326
50	100	0.096	0.922	1	0.096	0.202	0.332
100	100	0.106	0.996	1	0.084	0.672	0.854
100	250	0.110	1	1	0.106	1	1
250	250	0.098	1	1	0.126	1	1
Linear-circular regression. Von Mises errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 2$	$\beta = 1.75$	$\beta = 1.5$	$\beta = 2$	$\beta = 1.75$	$\beta = 1.5$
50	50	0.104	0.420	0.950	0.114	0.492	0.964
50	100	0.108	0.570	0.988	0.114	0.622	0.996
100	100	0.114	0.760	1	0.120	0.764	1
100	250	0.110	0.860	1	0.092	0.908	1
250	250	0.098	0.988	1	0.106	0.998	1
Circular-circular regression. Von Mises errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 2$	$\beta = 2.5$	$\beta = 3$	$\beta = 2$	$\beta = 2.5$	$\beta = 3$
50	50	0.102	0.570	0.994	0.118	0.518	0.982
50	100	0.124	0.658	1	0.126	0.684	0.996
100	100	0.104	0.902	1	0.112	0.898	1
100	250	0.128	0.962	1	0.126	0.976	1
250	250	0.122	1	1	0.094	0.998	1

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table B.4: Percentages of rejection (for $\alpha = 0.01$) obtained with the ANCOVA tests using the smoothing parameters obtained by cross-validation. Results for the first value of β show empirical size, whereas results for the other values of β show empirical power.

Circular-linear regression. χ^2 calibration. Normal errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
50	50	0.014	0.310	0.786	0.018	0.350	0.816
50	100	0.006	0.478	0.926	0.008	0.504	0.932
100	100	0.020	0.766	1	0.014	0.834	1
100	250	0.020	0.944	1	0.014	0.960	1
250	250	0.012	1	1	0.012	1	1
Circular-linear regression. Bootstrap calibration. Exponential errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
50	50	0.020	0.482	0.900	0.036	0.040	0.090
50	100	0.018	0.644	0.964	0.008	0.050	0.074
100	100	0.006	0.946	1	0.026	0.396	0.618
100	250	0.014	0.996	1	0.010	0.992	1
250	250	0.004	1	1	0.006	1	1
Linear-circular regression. Von Mises errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 2$	$\beta = 1.75$	$\beta = 1.5$	$\beta = 2$	$\beta = 1.75$	$\beta = 1.5$
50	50	0.006	0.106	0.752	0.006	0.190	0.814
50	100	0.012	0.226	0.914	0.016	0.274	0.938
100	100	0.010	0.392	0.994	0.022	0.408	0.994
100	250	0.006	0.622	1	0.012	0.644	1
250	250	0.010	0.890	1	0.014	0.894	1
Circular-circular regression. Von Mises errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 2$	$\beta = 2.5$	$\beta = 3$	$\beta = 2$	$\beta = 2.5$	$\beta = 3$
50	50	0.006	0.184	0.866	0.012	0.172	0.852
50	100	0.012	0.244	0.970	0.012	0.310	0.982
100	100	0.006	0.646	1	0.008	0.644	1
100	250	0.012	0.834	1	0.022	0.870	1
250	250	0.014	1	1	0.006	0.996	1

USC UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

Table B.5: Percentages of rejection (for $\alpha = 0.05$) obtained with the two versions of the no-effect test for circular-linear regression using different smoothing parameters. The shifted and scaled χ^2 calibrated test is applied to exponential errors and the bootstrap version is applied to normal errors. Results for the first value of β show empirical size, whereas results for the other values of β show empirical power.

Circular-linear regression. χ^2 calibration. Exponential errors									
	$4cv$			cv			$cv/8$		
n	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$
50	0.030	0.344	0.920	0.106	0.614	0.992	0.072	0.256	0.828
100	0.052	0.858	1	0.110	0.920	1	0.064	0.598	0.998
250	0.074	1	1	0.112	1	1	0.046	0.986	1
400	0.074	1	1	0.098	1	1	0.046	1	1
Circular-linear regression. Bootstrap calibration. Normal errors									
	$4cv$			cv			$cv/8$		
n	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$	$\beta=0$	$\beta=0.2$	$\beta=0.3$
50	0.070	0.236	0.858	0.104	0.412	0.962	0.060	0.198	0.582
100	0.046	0.578	0.998	0.094	0.782	1	0.052	0.292	0.954
250	0.060	0.984	1	0.100	0.996	1	0.054	0.862	1
400	0.050	0.998	1	0.078	1	1	0.044	0.988	1

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table B.6: Percentages of rejection (for $\alpha = 0.05$) obtained with the two versions of the ANCOVA tests for circular-linear regression. The shifted and scaled χ^2 calibrated tests are applied to exponential errors and the bootstrap versions are applied to normal errors. Results for the first value of β show empirical size, whereas results for the other values of β show empirical power.

Circular-linear regression. χ^2 calibration. Exponential errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
50	50	0.074	0.700	0.984	0.052	0.768	0.984
50	100	0.076	0.880	0.998	0.060	0.894	1
100	100	0.056	0.986	1	0.064	0.996	1
100	250	0.078	1	1	0.064	1	1
250	250	0.082	1	1	0.072	1	1
Circular-linear regression. Bootstrap calibration. Normal errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
50	50	0.048	0.524	0.898	0.042	0.086	0.136
50	100	0.044	0.650	0.974	0.050	0.070	0.142
100	100	0.038	0.914	1	0.040	0.464	0.694
100	250	0.040	0.986	1	0.044	0.992	1
250	250	0.052	1	1	0.046	1	1

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table B.7: Percentages of rejection (for $\alpha = 0.05$) obtained with the ANCOVA tests using the smoothing parameters which undersmooth the regression curves ($4cv$ for circular predictors, $cv/4$ for real-valued predictors). Results for the first value of β show empirical size, whereas results for the other values of β show empirical power.

Circular-linear regression. χ^2 calibration. Normal errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
50	50	0.048	0.364	0.766	0.052	0.356	0.782
50	100	0.058	0.534	0.914	0.044	0.542	0.930
100	100	0.030	0.802	0.990	0.062	0.770	0.993
100	250	0.054	0.932	1	0.050	0.948	1
250	250	0.042	1	1	0.062	1	1
Circular-linear regression. Bootstrap calibration. Exponential errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
50	50	0.050	0.574	0.908	0.070	0.120	0.218
50	100	0.052	0.710	0.962	0.056	0.138	0.208
100	100	0.052	0.926	0.998	0.040	0.530	0.784
100	250	0.062	0.982	1	0.056	0.992	1
250	250	0.064	1	1	0.062	1	1
Linear-circular regression. Von Mises errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 2$	$\beta = 1.75$	$\beta = 1.5$	$\beta = 2$	$\beta = 1.75$	$\beta = 1.5$
50	50	0.062	0.198	0.686	0.076	0.220	0.712
50	100	0.048	0.262	0.830	0.044	0.260	0.844
100	100	0.048	0.358	0.928	0.050	0.382	0.938
100	250	0.054	0.534	0.998	0.072	0.578	0.996
250	250	0.042	0.842	1	0.052	0.866	1
Circular-circular regression. Von Mises errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 2$	$\beta = 2.5$	$\beta = 3$	$\beta = 2$	$\beta = 2.5$	$\beta = 3$
50	50	0.044	0.244	0.790	0.050	0.242	0.822
50	100	0.054	0.374	0.970	0.066	0.354	0.956
100	100	0.066	0.596	1	0.048	0.646	1
100	250	0.062	0.884	1	0.060	0.866	1
250	250	0.056	1	1	0.048	1	1

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table B.8: Percentages of rejection (for $\alpha = 0.05$) obtained with the ANCOVA tests using the smoothing parameters which oversmooth the regression curves ($cv/8$ for circular predictors, $4cv$ for real-valued predictors). Results for the first value of β show empirical size, whereas results for the other values of β show empirical power.

Circular-linear regression. χ^2 calibration. Normal errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
50	50	0.160	0.460	0.780	0.290	0.648	0.900
50	100	0.152	0.534	0.884	0.220	0.714	0.968
100	100	0.154	0.846	0.992	0.202	0.912	0.998
100	250	0.104	0.962	1	0.142	0.982	1
250	250	0.114	1	1	0.152	1	1
Circular-linear regression. Bootstrap calibration. Exponential errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$	$\beta = 1$	$\beta = 1.5$	$\beta = 1.75$
50	50	0.202	0.690	0.890	0.068	0.102	0.170
50	100	0.172	0.740	0.978	0.072	0.116	0.158
100	100	0.126	0.964	1	0.068	0.442	0.608
100	250	0.092	0.998	1	0.110	1	1
250	250	0.052	1	1	0.116	1	1
Linear-circular regression. Von Mises errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 2$	$\beta = 1.75$	$\beta = 1.5$	$\beta = 2$	$\beta = 1.75$	$\beta = 1.5$
50	50	0.122	0.440	0.912	0.132	0.486	0.956
50	100	0.076	0.614	0.994	0.120	0.642	0.992
100	100	0.098	0.694	1	0.088	0.780	1
100	250	0.064	0.892	1	0.102	0.920	1
250	250	0.084	0.984	1	0.100	0.992	1
Circular-circular regression. Von Mises errors							
		Equality			Parallelism		
n_1	n_2	$\beta = 2$	$\beta = 2.5$	$\beta = 3$	$\beta = 2$	$\beta = 2.5$	$\beta = 3$
50	50	0.352	0.688	0.972	0.402	0.706	0.960
50	100	0.322	0.608	0.980	0.390	0.674	0.984
100	100	0.294	0.886	1	0.372	0.896	1
100	250	0.214	0.824	1	0.246	0.846	1
250	250	0.212	0.998	1	0.248	0.996	1

USC UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

B.2 Supplement to Chapter 5

In the simulation experiments of Section 5.4, a modification of the modal cross-validation criterion was used in order to attain more computational efficiency. The motivation behind this modification is the different role of the two smoothing parameters in the estimation of the regression multifunction. In this section, the modified modal cross-validation algorithm to select the smoothing parameters is described, and some simulation experiments are provided to study the practical performance of the modification.

Modification of the modal cross-validation criterion

The modified version of the modal cross-validation criterion is described in what follows. Three scenarios are considered: the setting with a circular predictor and a real-valued response, the situation where the response is circular and the predictor is real-valued and the scenario where both variables are circular.

Circular predictor and real-valued response Consider, in the first place, the case with a real-valued response, Y, and a circular predictor, Θ. The multimodal regression estimator in equation (5.4) will be denoted as $\widehat{M}^{\kappa,h}(\theta)$, where κ is the concentration parameter and h is the bandwidth. The cross-validation criterion presented in Section 5.3 can be modified as follows:

1. Select a pair of reference smoothing parameters by minimizing the MISE of the joint density estimator $\hat{f}(\theta, y)$ by assuming that the joint density is a mixture of independent von Mises and Gaussian densities (García-Portugués et al., 2013). These reference parameters will be denoted as $(\kappa_{\text{ref}}, h_{\text{ref}})$.

2. Set $\kappa = \kappa_{\text{ref}}$ and select h_{cv} minimizing

$$CV_1(h) = \frac{1}{n}\sum_{i=1}^{n} d^2\left[\widehat{M}_{-i}^{\kappa_{\text{ref}},h}(\Theta_i), Y_i\right] N_{-i}^2(\Theta_i),$$

   where $\widehat{M}_{-i}^{\kappa_{\text{ref}},h}(\cdot)$ represents the estimator computing without observations (Θ_i, Y_i), $N_{-i}(\Theta_i)$ represents the number of estimated local modes of $\widehat{M}_{-i}^{\kappa_{\text{ref}},h}(\Theta_i)$ and $d(x, A) = \inf_{z\in A} ||x - z||$.

3. Set $h = h_{\text{cv}}$ and select κ minimizing

$$CV_1(\kappa) = \frac{1}{n}\sum_{i=1}^{n} d^2\left[\widehat{M}_{-i}^{\kappa,h_{\text{cv}}}(\Theta_i), Y_i\right] N_{-i}^2(\Theta_i).$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Real-valued predictor and circular response Consider now the case with a circular response, Φ, and a real-valued predictor, X. Denote by $\widehat{M}^{h,\kappa}(x)$ the multimodal estimator described in equation (5.8), computed with smoothing parameters h and κ. In order to automatically select the smoothing parameters, the following modification of the cross-validation criterion will be used:

1. Select a pair of reference smoothing parameters by minimizing the MISE of the joint density estimator $\hat{f}(x,\phi)$ by assuming that the joint density is a mixture of independent von Mises and Gaussian densities (García-Portugués et al., 2013). These reference parameters will be denoted as $(h_{\text{ref}}, \kappa_{\text{ref}})$.

2. Set $h = h_{\text{ref}}$ and select κ_{cv} minimizing

$$CV_2(\kappa) = \frac{1}{n}\sum_{i=1}^{n} \tilde{d}\left[\widehat{M}_{-i}^{h_{\text{ref}},\kappa}(\Delta_i), Y_i\right] N_{-i}(\Delta_i),$$

   where $\tilde{d}(x, A) = \inf_{z\in A} 1 - \cos(x - z)$.

3. Set $\kappa = \kappa_{\text{cv}}$ and select h minimizing

$$CV_2(h) = \frac{1}{n}\sum_{i=1}^{n} \tilde{d}\left[\widehat{M}_{-i}^{h,\kappa_{\text{cv}}}(\Delta_i), Y_i\right] N_{-i}(\Delta_i).$$

Circular predictor and circular response Finally, when there is a circular response, Φ, and a circular predictor, Θ, the estimated regression multifunction will be denoted as $\widehat{M}^{\upsilon,\kappa}(\theta)$ when using concentration parameters υ and κ. The following modification of the algorithm will be employed:

1. Select a pair of reference smoothing parameters by minimizing the MISE of the joint density estimator $\hat{f}(\theta,\phi)$ by assuming that the joint density is a mixture of products of von Mises densities. These reference parameters will be denoted as $(\upsilon_{\text{ref}}, \kappa_{\text{ref}})$.

2. Set υ_{ref} and select κ_{cv} minimizing

$$CV_2(\kappa) = \frac{1}{n}\sum_{i=1}^{n} \tilde{d}\left[\widehat{M}_{-i}^{\upsilon_{\text{ref}},\kappa}(\Delta_i), Y_i\right] N_{-i}(\Delta_i).$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3. Set $\kappa = \kappa_{\text{cv}}$ and select υ minimizing

$$CV_2(g) = \frac{1}{n}\sum_{i=1}^{n} \tilde{d}(\widehat{M}_{-i}^{\upsilon,\kappa_{\text{cv}}}(\Delta_i), Y_i) N_{-i}(\Delta_i).$$

Performance of the modified modal cross-validation

Now, the performances of the modal cross-validation method described in Section 5.3 and the modification introduced in the previous section are compared through a brief simulation study. For that aim, data from models CL-1, LC-1 and CC-1 of Section 5.4 were simulated, where the dispersion/concentration parameters were set to $\sigma = 1$ for CL-1, $\tau = 8$ for LC-1 and $\tau = 12$ for CC-1. The MISE_m (for model CL-1) and the CMIE_m (for models LC-1 and CC-1) were estimated using the same sample sizes and number of replicates as in the simulation study of Section 5.4. Results when selecting the parameters by the original cross-validation procedure (O) and with its modified version (M) are shown in Table B.9. Both methods yield a very similar performance in terms of the estimated MISE_m and estimated CMIE_m. Regarding the computational efficiency, it is clear that the modified version requires less computational time, since it performs an optimization over one variable (twice) while the modal cross-validation procedure described in Section 5.3 implies optimization over two variables.

Table B.9: Monte Carlo estimates of MISE_m for model CL-1 ($\sigma = 1$) and CMIE_m for models LC-1 ($\tau = 8$) and CC-1 ($\tau = 12$), with different values of the sample sizes. Results under O correspond to the ones obtained with smoothing parameters selected by the original modal cross-validation and results under M refer to the modified modal cross-validation method.

	CL-1		LC-1		CC-1	
(n_1, n_2)	O	M	O	M	O	M
$(100, 100)$	1.877	1.865	0.023	0.024	0.213	0.201
$(100, 200)$	1.653	1.653	0.024	0.024	0.147	0.150
$(200, 200)$	0.851	0.840	0.012	0.012	0.066	0.065
$(200, 300)$	0.706	0.702	0.011	0.011	0.054	0.056
$(300, 300)$	0.574	0.574	0.008	0.008	0.045	0.043

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Appendix C

Theoretical results

This appendix contains the proofs regarding the theoretical results included in Chapters 2, 3 and 5, which were omitted from the main body of the document for facilitating the readability of the main ideas and concepts without interfering with technical details. Section C.1 gives the proof of Proposition 2.1. Section C.2 contains the proof of Proposition 3.1 and the derivation of the expression in equation (3.16), in addition to some preliminary results needed to derive the two proofs. Lastly, Section C.3 supplies the required assumptions to give the proofs of the propositions stated in Section 5.2, in addition to some results (with their corresponding proofs) needed to derive the final results.

C.1 Proofs of Chapter 2

Proof of Proposition 2.1. The minimization problem in equation (2.13) is equivalent to the next maximization problem:

$$\begin{aligned} &\arg \max_{\gamma_1,...,\gamma_J} \sum_{j=1}^{J} \sum_{i=1}^{n_j} \cos[\Phi_{ji} - \gamma_j - \hat{m}^1(\Delta_{ji})] \\ &\text{s.t.} \qquad \gamma_j \in [0, 2\pi), \ \forall\, j \in \{1, ..., J\}. \end{aligned}$$

The objective function can be written as

$$\begin{aligned} \Psi(\gamma_1, ..., \gamma_J) &= \sum_{j=1}^{J} \sum_{i=1}^{n_j} \cos[\Phi_{ji} - \hat{m}^1(\Delta_{ji}) - \gamma_j] \\ &= \sum_{j=1}^{J} \cos\gamma_j \sum_{i=1}^{n_j} \cos[\Phi_{ji} - \hat{m}^1(\Delta_{ji})] + \sum_{j=1}^{J} \sin\gamma_j \sum_{i=1}^{n_j} \sin[\Phi_{ji} - \hat{m}^1(\Delta_{ji})]. \end{aligned}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The partial derivative of $\Psi(\gamma_1, ..., \gamma_J)$ with respect to γ_t, $t \in \{1, ..., J\}$, is

$$\frac{\partial}{\partial \gamma_t} \Psi(\gamma_1, ..., \gamma_J) = -\sin\gamma_t \sum_{i=1}^{n_t} \cos[\Phi_{ti} - \hat{m}^1(\Delta_{ti})] + \cos\gamma_t \sum_{i=1}^{n_t} \sin[\Phi_{ti} - \hat{m}^1(\Delta_{ti})].$$

By denoting

$$\bar{C}_t = \sum_{i=1}^{n_t} \cos[\Phi_{ti} - \hat{m}^1(\Delta_{ti})] \quad \text{and} \quad \bar{S}_t = \sum_{i=1}^{n_t} \sin[\Phi_{ti} - \hat{m}^1(\Delta_{ti})],$$

it holds

$$\frac{\partial}{\partial \gamma_t}\Psi(\gamma_1, ..., \gamma_J) = 0 \iff \sin\gamma_t \bar{C}_t = \cos\gamma_t \bar{S}_t. \tag{C.1}$$

Therefore, if $\bar{C}_t \neq 0$ or $\bar{S}_t \neq 0$, it is obtained that

$$\hat{\gamma}_t = \text{atan2}(\bar{S}_t, \bar{C}_t).$$

In the case where $\bar{C}_t = \bar{S}_t = 0$, it holds that $\frac{\partial}{\partial \gamma_t}\Psi(\gamma_1, ..., \gamma_J) = 0 \ \forall \ (\gamma_1, \gamma_2, \ldots, \gamma_J) \in \mathbb{T}^J$. Consequently, if for all $j = 1, \ldots, J$, $\bar{C}_j \neq 0$ or $\bar{S}_j \neq 0$, equation (C.1) and Fermat's Theorem imply that $(\hat{\gamma}_1, \hat{\gamma}_2, ..., \hat{\gamma}_J)$ is the only critical point of Ψ.

To asses that $(\hat{\gamma}_1, \hat{\gamma}_2, ..., \hat{\gamma}_J)$ is a maximum, the Hessian matrix of Ψ is derived. If $l \in \{1, ..., J\}$ and $l \neq t$, it holds

$$\frac{\partial^2}{\partial \gamma_t \gamma_l}\Psi(\gamma_1, ..., \gamma_J) = 0.$$

On the other side,

$$\frac{\partial^2}{\partial \gamma_t^2}\Psi(\gamma_1, ..., \gamma_J) = -\cos\gamma_t \bar{C}_t - \sin\gamma_t \bar{S}_t,$$

and

$$\frac{\partial^2}{\partial \gamma_t^2}\Psi(\gamma_1, ..., \gamma_J)\Big|_{(\hat{\gamma}_1, \hat{\gamma}_2, ..., \hat{\gamma}_J)} = -\cos\hat{\gamma}_t \bar{C}_t - \sin\hat{\gamma}_t \bar{S}_t$$
$$= -\frac{\bar{C}_t}{\sqrt{\bar{S}_t^2 + \bar{C}_t^2}}\bar{C}_t - \frac{\bar{S}_t}{\sqrt{\bar{S}_t^2 + \bar{C}_t^2}}\bar{S}_t = \frac{-(\bar{S}_t^2 + \bar{C}_t^2)}{\sqrt{\bar{S}_t^2 + \bar{C}_t^2}}.$$

As a result, the Hessian matrix at $(\hat{\gamma}_1, \hat{\gamma}_2, ..., \hat{\gamma}_J)$ is given by

$$\boldsymbol{H}(\hat{\gamma}_1, \hat{\gamma}_2, ..., \hat{\gamma}_J) = \begin{pmatrix} \frac{-(\bar{C}_1^2 + \bar{S}_1^2)}{\sqrt{\bar{S}_1^2 + \bar{C}_1^2}} & 0 & ... & 0 \\ 0 & \frac{-(C_2^2 + \bar{S}_2^2)}{\sqrt{\bar{S}_2^2 + \bar{C}_2^2}} & ... & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ... & \frac{-(\bar{C}_J^2 + \bar{S}_J^2)}{\sqrt{\bar{S}_J^2 + \bar{C}_J^2}} \end{pmatrix}.$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Then, $(\hat{\gamma}_1, \hat{\gamma}_2, ..., \hat{\gamma}_J)$ is a maximum of Ψ as all the eigenvalues of $\boldsymbol{H}(\hat{\gamma}_1, \hat{\gamma}_2, ..., \hat{\gamma}_J)$ are negative. Since it was obtained that, if for all $j = 1, \ldots, J$ $\bar{C}_j \neq 0$ or $\bar{S}_j \neq 0$, the vector $(\hat{\gamma}_1, \hat{\gamma}_2, ..., \hat{\gamma}_J)$ was the only critical point, it holds that, under that hypothesis, it is the global maximum. □

C.2 Proofs of Chapter 3

The goal of this section is to give the proof of Proposition 3.1, which is the main result of Chapter 3, in addition to the derivation of the expression in (2.5). For both proofs, some preliminary results are needed, which are exposed in what follows. The proof of Proposition 3.1 and the derivation of (2.5) are given afterwards.

Preliminary results

For the following proofs, extensive use of the Taylor-like expansion employed by Di Marzio et al. (2009) will be made. The sine-polynomial expansion is based on the fact that, for small values of α, it holds $\sin\alpha \approx \alpha$, and then, for a function v,

$$v(\theta+\alpha) = v(\theta) + v'(\theta)\sin\alpha + \ldots + \frac{1}{k!}v^{(k)}(\theta)\sin^k\alpha + O(\sin^{k+1}\alpha). \tag{C.2}$$

In addition, some preliminary results will be employed in the proofs of this section. The first result was given by García-Portugués et al. (2013) and is summarized in Lemma C.1, while the proof of the second result, collected in Lemma C.2, is provided in this section.

Lemma C.1 (Lemma 1 and Remark 5 in García-Portugués et al. (2013))**.** *Under assumption* (3.4)*, when* $\kappa \to \infty$*,*

$$\lim_{\kappa\to\infty}\int_0^{2\kappa} r^{\frac{j-1}{2}}\left(2-\frac{r}{\kappa}\right)^{\frac{k-1}{2}}K^l(r)dr = 2^{\frac{k-1}{2}}\int_0^{\infty} r^{\frac{j-1}{2}}K^l(r)dr,$$

for $j,k\in\mathbb{N}$ *and* $l=1,2,4$*.*

Lemma C.2. *Given a sample of circular data* $\Theta_1,\ldots,\Theta_n$ *with a twice continuously density* $f(\cdot)$ *with* $f(\theta_0)>0$ *and a kernel* $K_\kappa(\cdot)$ *satisfying equations* (3.3) *and* (3.4)*, it holds*

a) $s_{n,j} = \sum_{i=1}^n \sin^j(\Theta_i-\theta_0)K_\kappa(\Theta_i-\theta_0) = nf(\theta_0)2^{\frac{j}{2}}\kappa^{-\frac{j}{2}}[b_j^*(K)+o_P(1)]$.

b) $\gamma_{n,j} = \sum_{i=1}^n \sin^j(\Theta_i-\theta_0)K_\kappa^2(\Theta_i-\theta_0) = nf(\theta_0)2^{\frac{j-1}{2}}\kappa^{-\frac{j-1}{2}}[d_j^*(K)+o_P(1)]$.

Proof. The proof of the statement in a) is obtained first and the derivations for statement b) are given afterwards.

Statement a)

It holds that

$$s_{n,j} = \mathbb{E}(s_{n,j}) + O_P\left(\sqrt{\mathrm{Var}(s_{n,j})}\right).$$

For the expectation part,

$$\mathbb{E}(s_{n,j}) = n\int_0^{2\pi} \sin^j(\alpha-\theta_0)K_\kappa(\alpha-\theta_0)f(\alpha)d\alpha = nc_\kappa \int_0^{2\pi} \sin^j(\varphi)K[\kappa(1-\cos\varphi)]f(\theta_0+\varphi)d\varphi,$$

where the second equality was obtained by applying the change of variables $\varphi = \alpha - \theta_0$ and using (3.3). Now, by splitting the integral, it holds

$$\begin{aligned}\mathbb{E}(s_{n,j}) = nc_\kappa \Bigg(&\int_0^{\pi} \sin^j(\varphi)K[\kappa(1-\cos\varphi)]f(\theta_0+\varphi)d\varphi \\ &+ \int_\pi^{2\pi} \sin^j(\varphi)K[\kappa(1-\cos\varphi)]f(\theta_0+\varphi)d\varphi\Bigg).\end{aligned} \tag{C.3}$$

For the first integral in (C.3), the change of variables $r = \kappa(1-\cos\varphi)$ is applied, noting that, since $\varphi \in [0,\pi]$, it holds $\varphi = \arccos(1-r/\kappa)$, and thus obtaining, for the first integral in (C.3):

$$\int_0^{\pi} \sin^j(\varphi)K[\kappa(1-\cos\varphi)]d\varphi = \frac{1}{\kappa}\int_0^{2\kappa} \left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{j-1}{2}} K(r)f[\theta_0+\arccos(1-r/\kappa)]dr.$$

Applying the expansion in (C.2), it holds that $f[\theta_0+\arccos(1-r/\kappa)] = f(\theta_0)+o(1)$ and, then, the first integral in (C.3) can be expressed as

$$\frac{1}{\kappa}f(\theta_0)\int_0^{2\kappa} \left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{j-1}{2}} K(r)dr[1+o(1)]. \tag{C.4}$$

Now, the focus is switched to the second integral in (C.3). By employing again the change of variables $r = \kappa(1-\cos\varphi)$, but taking into account that now $\varphi \in [\pi, 2\pi]$ and, thus, having $\varphi = -\arccos(1-r/\kappa)$, it is obtained that

$$\begin{aligned}&\int_\pi^{2\pi} \sin^j(\varphi)K[\kappa(1-\cos\varphi)]d\varphi \\ &= -\frac{1}{\kappa}\int_0^{2\kappa} \left[-\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{1}{2}}\right]^{j-1} K(r)f\left[\theta_0-\arccos(1-r/\kappa)\right]dr.\end{aligned} \tag{C.5}$$

Note that there are two different scenarios depending on j being odd or even. When j is even ($j-1$ is odd), the right term in (C.5) can be expressed as

$$\begin{aligned}&\frac{1}{\kappa}\int_0^{2\kappa} \left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{j-1}{2}} K(r)f[\theta_0-\arccos(1-r/\kappa)]dr \\ &= \frac{1}{\kappa}f(\theta_0)\int_0^{2\kappa} \left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{j-1}{2}} K(r)dr[1+o(1)].\end{aligned}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Recall now the asymptotic approximation of c_κ, given by

$$c_\kappa^{-1} \sim \kappa^{-1/2}\lambda(K) \quad \text{with} \quad \lambda(K) = 2^{1/2}\int_0^\infty r^{-1/2}K(r)dr. \tag{C.6}$$

Then, equation (C.3) for an even j becomes

$$\begin{aligned}
\mathbb{E}(s_{n,j}) &= nc_\kappa \frac{2}{\kappa} f(\theta_0)\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{j-1}{2}} K(r)dr[1+o(1)] \\
&= nc_\kappa 2\kappa^{-\frac{j+1}{2}} f(\theta_0)\int_0^{2\kappa} r^{\frac{j-1}{2}}\left(2-\frac{r}{\kappa}\right)^{\frac{j-1}{2}} K(r)dr[1+o(1)] \\
&= n2\kappa^{-\frac{j}{2}} f(\theta_0)\left[\lambda(K)^{-1}2^{\frac{j-1}{2}}\int_0^\infty r^{\frac{j-1}{2}}K(r)dr + o(1)\right][1+o(1)] \\
&= nf(\theta_0)2^{\frac{j}{2}}\kappa^{-\frac{j}{2}} b_j(K) + o\left(n\kappa^{-\frac{j}{2}}\right),
\end{aligned} \tag{C.7}$$

where $b_j(K)$ is given by

$$b_j(K) = \frac{\int_0^\infty r^{\frac{j-1}{2}}K(r)dr}{\int_0^\infty r^{-\frac{1}{2}}K(r)dr}. \tag{C.8}$$

Note that in the third equality of (C.7), equation (3.5) and Lemma C.1 were used, in addition to the assumption in equation (3.4).

On the other hand, when j is odd ($j-1$ is even), the right term in (C.5) is given by

$$\begin{aligned}
&-\frac{1}{\kappa}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{j-1}{2}} K(r)f[\theta_0-\arccos(1-r/\kappa)]dr \\
&= -\frac{1}{\kappa}f(\theta_0)\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{j-1}{2}} K(r)dr[1+o(1)],
\end{aligned}$$

which coincides with the expression in (C.4) but with opposite sign. Then, it holds that for an odd j,

$$\mathbb{E}(s_{n,j}) = o\left(n\kappa^{-\frac{j}{2}}\right).$$

In order to have a more compact expression, for a general j, it can be written

$$\mathbb{E}(s_{n,j}) = nf(\theta_0)2^{\frac{j}{2}}\kappa^{-\frac{j}{2}}[b_j^*(K)+o(1)], \tag{C.9}$$

where

$$b_j^*(K) = \begin{cases} 0 & \text{if } j \text{ is odd,} \\ b_j(K) & \text{if } j \text{ is even.} \end{cases}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Now, for the variance of $s_{n,j}$, it is obtained, by analogous computations, that

$$\begin{aligned}\mathrm{Var}(s_{n,j}) \leqslant & n\mathbb{E}\left[\sin^{2j}(\Theta_1-\theta_0)K_\kappa^2(\Theta_1-\theta_0)\right]\\ =& n\int_0^{2\pi}\sin^{2j}(\alpha-\theta_0)K_\kappa^2(\alpha-\theta_0)f(\alpha)d\alpha\\ =& nc_\kappa^2(K)\int_0^{2\pi}\sin^{2j}(\varphi)K^2[\kappa(1-\cos\varphi)]f(\theta_0+\varphi)d\varphi.\end{aligned}$$

Employing the change of variables $r=\kappa(1-\cos\varphi)$ yields

$$\begin{aligned}\mathrm{Var}(s_{n,j}) \leqslant & nc_\kappa^2(K)\frac{1}{\kappa}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{2j-1}{2}}K^2(r)f[\theta_0+\arccos(1-r/k)]dr\\ & - nc_\kappa^2(K)\frac{1}{\kappa}\int_0^{2\kappa}\left[-\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{1}{2}}\right]^{2j-1}K^2(r)f[\theta_0-\arccos(1-r/k)]dr\\ =& O(n\kappa^{-\frac{2j-1}{2}}).\end{aligned} \tag{C.10}$$

Note that equation (3.5), the approximation in (C.6) and the assumption in equation (3.4) have been employed. Finally, putting (C.9) and (C.10) together, it is derived that

$$\begin{aligned}s_{n,j} =& nf(\theta_0)2^{\frac{j}{2}}\kappa^{-\frac{j}{2}}\left[b_j^*(K)+o(1)\right]+O_P\left(n^{\frac{1}{2}}\kappa^{-\frac{2j-1}{4}}\right)\\ =& nf(\theta_0)2^{\frac{j}{2}}\kappa^{-\frac{j}{2}}\left[b_j^*(K)+o(1)+O_P\left(\frac{\kappa^{1/4}}{n^{1/2}}\right)\right]\\ =& nf(\theta_0)2^{\frac{j}{2}}\kappa^{-\frac{j}{2}}[b_j^*(K)+o_P(1)].\end{aligned}$$

Statement b)

By using the same changes of variables as in the previous calculations, it holds that the expectation of $\gamma_{n,j}$ can be expressed as

$$\begin{aligned}\mathbb{E}(\gamma_{n,j}) =& n\int_0^{2\pi}\sin^j(\alpha-\theta_0)K_\kappa^2(\alpha-\theta_0)f(\alpha)d\alpha\\ =& nc_\kappa^2(K)\int_0^{2\pi}\sin^j(\varphi)K^2[\kappa(1-\cos\varphi)]f(\theta_0+\varphi)d\varphi\\ =& nc_\kappa^2(K)\int_0^{\pi}\sin^j(\varphi)K^2[\kappa(1-\cos\varphi)]f(\theta_0+\varphi)d\varphi\\ & + nc_\kappa^2(K)\int_\pi^{2\pi}\sin^j(\varphi)K^2[\kappa(1-\cos\varphi)]f(\theta_0+\varphi)d\varphi.\end{aligned}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Employing now the change of variables $r = \kappa(1-\cos\varphi)$, it can be obtained that

$$\begin{aligned}\mathbb{E}(\gamma_{n,j}) =& n\frac{c_\kappa^2(K)}{\kappa}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{j-1}{2}}K^2(r)f[\theta_0+\arccos(1-r/\kappa)]dr\\ &-n\frac{c_\kappa^2(K)}{\kappa}\int_0^{2\kappa}\left[-\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{1}{2}}\right]^{j-1}K^2(r)f[\theta_0-\arccos(1-r/\kappa)]dr.\end{aligned}$$

Now, if j is even, by applying the expansion in (C.2), the expectation of $\gamma_{n,j}$ is given by

$$\begin{aligned}\mathbb{E}(\gamma_{n,j}) =& 2nf(\theta_0)\frac{c_\kappa^2(K)}{\kappa^{\frac{j+1}{2}}}\int_0^{2\kappa}r^{\frac{j-1}{2}}\left(2-\frac{r}{\kappa}\right)^{\frac{j-1}{2}}K^2(r)dr[1+o(1)]\\ =&2nf(\theta_0)\kappa^{-\frac{j-1}{2}}\left[\lambda(K)^{-2}2^{\frac{j-1}{2}}\int_0^{\infty}r^{\frac{j-1}{2}}K^2(r)dr+o(1)\right][1+o(1)]\\ =&nf(\theta_0)2^{\frac{j-1}{2}}\kappa^{-\frac{j-1}{2}}d_j(K)+o\left(n\kappa^{-\frac{j-1}{2}}\right),\end{aligned} \tag{C.11}$$

where

$$d_j(K)=\frac{\int_0^\infty r^{\frac{j-1}{2}}K^2(r)dr}{\left(\int_0^\infty r^{-\frac{1}{2}}K(r)dr\right)^2}.$$

Note that in the second equality of (C.11), equations (3.5) and Lemma C.1 have been used, and the quantities $d_j(K)$ exist due to assumption (3.4).

On the other hand, by analogous derivations, when j is odd,

$$\mathbb{E}(\gamma_{n,j})=n\frac{c_\kappa^2(K)}{\kappa}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{j-1}{2}}K^2(r)dr\cdot o(1)=o\left(n\kappa^{-\frac{j-1}{2}}\right).$$

Then, for a general j, it can be written that

$$\mathbb{E}(\gamma_{n,j})=nf(\theta_0)2^{\frac{j-1}{2}}\kappa^{-\frac{j-1}{2}}[d_j^*(K)+o(1)],\quad\text{with}\quad d_j^*(K)=\begin{cases}0 & \text{if } j \text{ is odd}\\ d_j(K) & \text{if } j \text{ is even.}\end{cases}$$

Further, the variance of $\gamma_{n,j}$ can be expressed as

$$\begin{aligned}\mathrm{Var}(\gamma_{n,j}) \leqslant& \mathbb{E}\left[\sin^{2j}(\Theta_1-\theta_0)K_\kappa^4(\Theta_1-\theta_0)\right]\\ =&nc_\kappa^4(K)\int_0^{2\pi}\sin^{2j}(\alpha-\theta_0)K^4\{\kappa[1-\cos(\alpha-\theta_0)]\}f(\alpha)d\alpha\\ =&nc_\kappa^4(K)\int_0^{2\pi}\sin^{2j}(\varphi)K^4[\kappa(1-\cos\varphi)]f(\theta_0+\varphi)d\varphi.\end{aligned}$$

After the change of variables $r=\kappa(1-\cos\varphi)$ and with analogous derivations to the bias

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

calculations, one obtains

$$\begin{aligned}\text{Var}(\gamma_{n,j}) \leqslant & n\frac{c_\kappa^4(K)}{\kappa}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{2j-1}{2}} K^4(r)f[\theta_0+\arccos(1-r/\kappa)]dr \\ & - n\frac{c_\kappa^4(K)}{\kappa}\int_0^{2\kappa}\left[-\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{\frac{1}{2}}\right]^{2j-1} K^4(r)f[\theta_0-\arccos(1-r/\kappa)]dr \\ = & O\left(n\kappa^{-\frac{2j-3}{2}}\right).\end{aligned}$$

Lastly, it holds

$$\begin{aligned}\gamma_{n,j} &= \mathbb{E}(\gamma_{n,j}) + O_P\left(\sqrt{\text{Var}(\gamma_{n,j})}\right) \\ &= nf(\theta_0)2^{\frac{j-1}{2}}\kappa^{-\frac{j-1}{2}}\left[d_j^*(K)+o(1)\right] + O_P\left(n^{\frac{1}{2}}\kappa^{-\frac{2j-3}{4}}\right) \\ &= nf(\theta_0)2^{\frac{j-1}{2}}\kappa^{-\frac{j-1}{2}}\left[d_j^*(K)+o(1)+O_P\left(\frac{\kappa^{\frac{1}{4}}}{n^{\frac{1}{2}}}\right)\right] \\ &= nf(\theta_0)2^{\frac{j-1}{2}}\kappa^{-\frac{j-1}{2}}\left[d_j^*(K)+o_P(1)\right].\end{aligned}$$

□

Proof of Proposition 3.1

Proof of Proposition 3.1. The proof is based on the proof of Theorem 1 in Fan and Gijbels (1995), taking into account the results in Lemma C.2. By denoting $d_n = \text{tr}[\boldsymbol{W} - \boldsymbol{W}\boldsymbol{\Theta}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^T\boldsymbol{W}]$, the expression $\hat{\sigma}^2(\theta_0)$ can be written as

$$\begin{aligned}\hat{\sigma}^2(\theta_0) &= d_n^{-1}\sum_{i=1}^n (Y_i-\hat{Y}_i)^2 K_\kappa(\Theta_i-\theta_0) = d_n^{-1}(\boldsymbol{Y}-\boldsymbol{\Theta}\hat{\boldsymbol{\beta}})^T\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{\Theta}\hat{\boldsymbol{\beta}}) \\ &= d_n^{-1}\boldsymbol{Y}^T[\boldsymbol{W}-\boldsymbol{W}\boldsymbol{\Theta}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^T\boldsymbol{W}]\boldsymbol{Y}.\end{aligned}$$

Now, the continuity of $\sigma^2(\theta)$ implies

$$\mathbb{E}\left[\hat{\sigma}^2(\theta_0)|\Theta_1,\ldots,\Theta_n\right] = d_n^{-1}\boldsymbol{m}^T[\boldsymbol{W}-\boldsymbol{W}\boldsymbol{\Theta}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^T\boldsymbol{W}]\boldsymbol{m} + \sigma^2(\theta_0), \quad \text{(C.12)}$$

where $\boldsymbol{m} = [m(\Theta_1),\ldots,m(\Theta_n)]^T$. Further, $\boldsymbol{r} = \boldsymbol{m} - \boldsymbol{\Theta}\boldsymbol{\beta}$ can be approximated by the vector with elements

$$r(\Theta_i) = m(\Theta_i) - \sum_{j=0}^p \beta_j \sin^j(\Theta_i-\theta_0) = \beta_{p+1}\sin^{p+1}(\Theta_i-\theta_0) + O_P\left(\kappa^{-\frac{p+2}{2}}\right). \quad \text{(C.13)}$$

In addition,

$$\begin{aligned} d_n &= \text{tr}[\boldsymbol{W} - \boldsymbol{W}\boldsymbol{\Theta}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^T\boldsymbol{W}] \\ &= s_{n,0} - \text{tr}[(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^T\boldsymbol{W}^2\boldsymbol{\Theta}] = nf(\theta_0) + O_P\left(\kappa^{\frac{1}{2}}\right), \end{aligned} \tag{C.14}$$

where Lemma C.2 was used. Then, starting from (C.12) and applying (C.13) and (C.14) yields

$$\begin{aligned} \mathbb{E}\left[\hat{\sigma}^2(\theta_0)|\Theta_1,\ldots,\Theta_n\right] &= d_n^{-1}\left(s_{n,2p+2} - \boldsymbol{c}_n^T\boldsymbol{S}_n^{-1}\boldsymbol{c}_n\right)\beta_{p+1}^2 + \sigma^2(\theta_0) + o_P\left(\kappa^{-(p+1)}\right) \\ &= C_p\beta_{p+1}^2 2^{p+1}\kappa^{-(p+1)} + \sigma^2(\theta_0) + o_P\left(\kappa^{-(p+1)}\right), \end{aligned} \tag{C.15}$$

where $\boldsymbol{c}_n = (s_{n,p+1},\ldots,s_{n,2p+1})^T$. On the other hand, it holds that

$$N^{-1} = \boldsymbol{e}_1^T(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^T\boldsymbol{W}^2\boldsymbol{\Theta}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{e}_1 = \boldsymbol{e}_1^T\boldsymbol{S}_n^{-1}\boldsymbol{\Gamma}_n\boldsymbol{S}_n^{-1}\boldsymbol{e}_1.$$

By recalling Lemma C.2, it is easy to see that

$$\boldsymbol{S}_n = nf(\theta_0)[\boldsymbol{L}\boldsymbol{B}\boldsymbol{L} + o_P(\boldsymbol{L}\boldsymbol{1}\boldsymbol{L})], \quad \text{and} \quad \boldsymbol{\Gamma}_n = 2^{-\frac{1}{2}}nf(\theta_0)\kappa^{\frac{1}{2}}[\boldsymbol{L}\boldsymbol{D}\boldsymbol{L} + o_P(\boldsymbol{L}\boldsymbol{1}\boldsymbol{L})], \tag{C.16}$$

where $\boldsymbol{L} = \text{diag}\left\{1, 2^{\frac{1}{2}}\kappa^{-\frac{1}{2}},\ldots,2^{\frac{p}{2}}\kappa^{-\frac{p}{2}}\right\}$ and $\boldsymbol{1}$ denotes the $(p+1)\times(p+1)$ matrix with all elements equal to 1. Thus,

$$\begin{aligned} N^{-1} &= \frac{\kappa^{\frac{1}{2}}}{2^{\frac{1}{2}}nf(\theta_0)}\boldsymbol{e}_1^T\boldsymbol{L}^{-1}\boldsymbol{B}^{-1}\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{L}^{-1}\boldsymbol{e}_1 + o_P\left(n^{-1}\kappa^{\frac{1}{2}}\right) \\ &= \frac{\kappa^{\frac{1}{2}}}{2^{\frac{1}{2}}nf(\theta_0)}\boldsymbol{e}_1^T\boldsymbol{B}^{-1}\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{e}_1 + o_P\left(n^{-1}\kappa^{\frac{1}{2}}\right) = \frac{a_0\kappa^{\frac{1}{2}}}{2^{\frac{1}{2}}nf(\theta_0)} + o_P\left(n^{-1}\kappa^{\frac{1}{2}}\right). \end{aligned} \tag{C.17}$$

The combination of (C.15) and (C.17) leads to

$$\begin{aligned} \mathbb{E}[\text{CRSC}(\theta_0;\kappa)|\Theta_1,\cdots,\Theta_n] &= C_p\beta_{p+1}^2 2^{p+1}\kappa^{-(p+1)} + \sigma^2(\theta_0) + (p+1)\frac{\sigma^2(\theta_0)a_0\kappa^{\frac{1}{2}}}{2^{\frac{1}{2}}nf(\theta_0)} \\ &\quad + o_P\left(\frac{1}{\kappa^{p+1}} + \frac{\kappa^{\frac{1}{2}}}{n}\right). \end{aligned}$$

□

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Derivation of the optimal smoothing parameter in the least-squares case

In this section, the expression of the optimal smoothing parameter in the least-squares case, given in equation (3.16), is derived. First, the bias of the estimator in equation (2.5) is computed:

$$\text{Bias}(\hat{\boldsymbol{\beta}}|\Theta_1,\ldots,\Theta_n) = (\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{r},$$

with $\boldsymbol{r}$ being the vector

$$\boldsymbol{r} = \left\{\beta_{p+1}\sin^{p+1}(\Theta_i-\theta_0) + o_P\left[\sin^{p+1}(\Theta_i-\theta_0)\right]\right\}_{i=1,\ldots,n}.$$

Then, it holds

$$\text{Bias}(\hat{\boldsymbol{\beta}}|\Theta_1,\ldots,\Theta_n) = \boldsymbol{S}_n^{-1}\left[\beta_{p+1}\boldsymbol{c}_n + o_P\left(n\kappa^{-\frac{p+1}{2}}\right)\right],$$

where $\boldsymbol{c}_n = (s_{n,p+1},\ldots,s_{n,2p+1})^T$. Now, using the expression of $\boldsymbol{S}_n$ in equation (C.16) and Lemma C.2 yields

$$\text{Bias}(\hat{\boldsymbol{\beta}}|\Theta_1,\ldots,\Theta_n) = \beta_{p+1}\boldsymbol{L}^{-1}\boldsymbol{B}^{-1}\boldsymbol{c}_p 2^{\frac{p+1}{2}}\kappa^{-\frac{p+1}{2}}[1+o_P(1)].$$

Thus, since $\hat{g}^{(\nu)}(\theta_0) = \nu!\boldsymbol{e}_{\nu+1}^T\hat{\boldsymbol{\beta}}$, it holds that

$$\text{Bias}[\hat{g}^{(\nu)}(\theta_0)|\Theta_1,\ldots,\Theta_n] = \nu!\beta_{p+1}\boldsymbol{e}_{\nu+1}^T\boldsymbol{B}^{-1}\boldsymbol{c}_p 2^{\frac{p+1-\nu}{2}}\kappa^{-\frac{p+1-\nu}{2}} + o_P\left(\kappa^{-\frac{p+1-\nu}{2}}\right).$$

On the other hand, the variance of $\hat{\boldsymbol{\beta}}$ is given by

$$\text{Var}(\hat{\boldsymbol{\beta}}|\Theta_1,\ldots,\Theta_n) = (\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^T\boldsymbol{W}\text{Var}(\boldsymbol{Y})\boldsymbol{W}\boldsymbol{\Theta}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta})^{-1} = \boldsymbol{S}_n^{-1}\boldsymbol{\Theta}^T\boldsymbol{\Sigma}\boldsymbol{\Theta}\boldsymbol{S}_n^{-1},$$

where $\boldsymbol{\Sigma} = \text{diag}\{K_\kappa^2(\Theta_i-\theta_0)\sigma^2(\Theta_i)\}_{i=1}^n$. Note that the $(i,j)th$ element of $\boldsymbol{\Theta}^T\boldsymbol{\Sigma}\boldsymbol{\Theta}$ is given by $\delta_{n,i+j-2}$ where

$$\delta_{n,j} = \sum_{i=0}^{n}\sin^j(\Theta_i-\theta_0)K_\kappa^2(\Theta_i-\theta_0)\sigma^2(\Theta_i).$$

Using arguments analogue to the proof of statement b) in Lemma C.2, $\delta_{n,j}$ can be expressed as

$$\delta_{n,j} = nf(\theta_0)\sigma^2(\theta_0)2^{\frac{j-1}{2}}\kappa^{-\frac{j-1}{2}}[d_j^*(K)+o_P(1)]$$

and, thus,

$$\boldsymbol{\Theta}^T\boldsymbol{\Sigma}\boldsymbol{\Theta} = nf(\theta_0)\sigma^2(\theta_0)2^{-\frac{1}{2}}\kappa^{\frac{1}{2}}[\boldsymbol{L}\boldsymbol{D}\boldsymbol{L}+o_P(\boldsymbol{L}\boldsymbol{1}\boldsymbol{L})].$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The previous equation in addition to the expression of $\boldsymbol{S}_n$ in equation (C.15) implies

$$\text{Var}(\hat{\boldsymbol{\beta}}|\Theta_1,\ldots,\Theta_n) = \frac{\sigma^2(\theta_0)2^{-\frac{1}{2}}\kappa^{\frac{1}{2}}}{nf(\theta_0)}[\boldsymbol{L}^{-1}\boldsymbol{B}^{-1}\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{L}^{-1}+o_P(\boldsymbol{L}^{-1}\boldsymbol{1}\boldsymbol{L}^{-1})].$$

Now, recalling that $\hat{g}^{(\nu)}(\theta_0) = \nu! \boldsymbol{e}_{\nu+1}^T \hat{\boldsymbol{\beta}}$, it can be derived that

$$\begin{aligned}
&\text{Var}[\hat{g}^{(\nu)}(\theta_0)|\Theta_1, \ldots, \Theta_n] \\
&= \frac{\nu!^2 \sigma^2(\theta_0) 2^{-\frac{1+2\nu}{2}} \kappa^{\frac{1+2\nu}{2}}}{n f(\theta_0)} \boldsymbol{e}_{\nu+1}^T \boldsymbol{L}^{-1} \boldsymbol{B}^{-1} \boldsymbol{D} \boldsymbol{B}^{-1} \boldsymbol{L}^{-1} \boldsymbol{e}_{\nu+1} + o_P\left(n^{-1} \kappa^{\frac{1+2\nu}{2}}\right) \\
&= \frac{\nu!^2 \sigma^2(\theta_0) 2^{-\frac{1+2\nu}{2}} \kappa^{\frac{1+2\nu}{2}}}{n f(\theta_0)} a_\nu + o_P\left(n^{-1} \kappa^{\frac{1+2\nu}{2}}\right).
\end{aligned}$$

Consequently, the Mean Squared Error of the estimator in equation (2.5) can be expressed as

$$\begin{aligned}
\text{MSE}[\hat{g}^{(\nu)}(\theta_0)|\Theta_1, \ldots, \Theta_n] &= \nu!^2 \beta_{p+1}^2 (\boldsymbol{e}_{\nu+1}^T \boldsymbol{B}^{-1} \boldsymbol{c}_p)^2 2^{p+1-\nu} \kappa^{-(p+1-\nu)} \\
&+ \frac{\nu!^2 \sigma^2(\theta_0) 2^{-\frac{1+2\nu}{2}} \kappa^{\frac{1+2\nu}{2}}}{n f(\theta_0)} a_\nu + o_P\left(\kappa^{-\frac{p+1-\nu}{2}} + n^{-1} \kappa^{\frac{1+2\nu}{2}}\right).
\end{aligned}$$

Finally, the concentration which minimizes the asymptotic version of the MSE is given by the expression in equation (3.16). □

C.3 Assumptions and proofs of Chapter 5

In this section, the assumptions and proofs corresponding to the propositions of Chapter 5 are presented. In the first place, some preliminary results are given. Secondly, the assumptions for Propositions 5.1, 5.2 and 5.3 are stated. Next, six lemmas, necessary to complete the proofs of the aforementioned results, are enunciated and proved. Finally, the proofs of Propositions 5.1, 5.2 and 5.3 are given.

Preliminary tools

In what follows, some preliminary definitions and results, which will be used within the proofs of this section, will be introduced.

In the first place, for the next proofs it is necessary to define the classes of functions of Vapnik-Cervonenkis (VC) type. Let $(S, \mathcal{S})$ be a measurable space and $\mathcal{F}$ a uniformly bounded collection of measurable functions on it. $\mathcal{F}$ is a bounded measurable VC class of functions if the class $\mathcal{F}$ is separable or is image admissible Suslin (Dudley, 1999, Chap. 5) and if there exist $A, u > 0$ such that $\forall\ 0 < \varepsilon < 1$,

$$N\left(\mathcal{F}, L_2(P), \varepsilon ||F||_{L_2(P)}\right) \leqslant \left(\frac{A}{\varepsilon}\right)^u, \tag{C.18}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

where $N(T, d, \varepsilon)$ is the ε-covering number for a metric space (T, d) and P is any probability measure. The quantities A and u for which the inequality holds for all probability measures on $(S, \mathcal{S})$ are known as the set of VC characteristics of the class $\mathcal{F}$. In what follows, the

notation $||G||_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |G(f)|$ will be used.

Secondly, a result given in Giné and Guillou (2002) is enunciated. This result will be employed in the proofs of Lemmas C.6, C.7 and C.8.

Theorem C.1 (Corollary 2.2 in Giné and Guillou (2002))**.** *Let $\mathcal{F}$ be a measurable uniformly bounded VC class of functions and let σ^2 and U be any numbers such that $\sigma^2 \geqslant \sup_{f\in\mathcal{F}} Var_P f$, $U \geqslant \sup_{f\in\mathcal{F}} ||f||_\infty$, $0 < \sigma < U/2$ and $\sqrt{n}\sigma \geqslant U[\log(U/\sigma)]^{1/2}$. Then, there exist $B, C > 0$ depending only on the VC characteristics of $\mathcal{F}$ such that for all $\lambda \geqslant C$ and t satisfying*

$$C\sqrt{n}\sigma\sqrt{\log\frac{U}{\sigma}} \leqslant t \leqslant \lambda\frac{n\sigma^2}{U},$$

$$\mathbb{P}\left\{\left|\left|\sum_{i=1}^{n}[f(\chi_i) - \mathbb{E}[f(\chi_i)]]\right|\right|_{\mathcal{F}} > t\right\} \leqslant B\exp\left(-\frac{1}{B}\frac{\log[1+\lambda/(4B)]}{\lambda}\frac{t^2}{n\sigma^2}\right).$$

In particular, taking

$$t = C_2\sqrt{n}\sigma\sqrt{\log\frac{U}{\sigma}} \quad \textit{with } C_2 \geqslant C, \textit{ then}$$

$$\mathbb{P}\left\{\left|\left|\sum_{i=1}^{n}\{f(\chi_i)-\mathbb{E}[f(\chi_i)]\}\right|\right|_{\mathcal{F}} > C_2\sqrt{n}\sigma\sqrt{\log\frac{U}{\sigma}}\right\} \leqslant L\exp\left(-\frac{C_2\log[1+C_2/(4B)]}{B}\log\frac{U}{\sigma}\right). \tag{C.19}$$

Lastly, recall the Montgomery-Smith's maximal inequality (Montgomery-Smith, 1993), which will be also used in the proofs of Lemmas C.6, C.7 and C.8. The Montgomery-Smith's maximal inequality states that

$$\mathbb{P}\left\{\max_{k\leqslant n}\left|\left|\sum_{i=1}^{n}\{f(\chi_i) - \mathbb{E}[f(\chi_i)]\}\right|\right|_{\mathcal{F}} > t\right\} \leqslant 9\mathbb{P}\left\{\left|\left|\sum_{i=1}^{n}\{f(\chi_i) - \mathbb{E}[f(\chi_i)]\}\right|\right|_{\mathcal{F}} > \frac{t}{30}\right\}. \tag{C.20}$$

Assumptions

The assumptions needed for the asymptotic results stated in Section 5.2 will be summarized next. For that purpose, recall that Θ and X represent, respectively, circular and real-valued covariates and that Δ denotes a general predictor variables. Supp(Δ) is used to denote the support of the predictor variable, which can be $\mathbb{T}$ or $\mathbb{R}$. In addition, Y and Φ denote, respectively, real-valued and circular response variables. It will be assumed that:

(A1) The joint density function $f(\theta, y)$ is at least four times continuously differentiable and its partial derivatives with respect to y are bounded in absolute value by some positive constant.

(A2) The joint density function $f(\delta, \phi)$ is at least four times continuously differentiable and its partial derivatives with respect to ϕ are are bounded in absolute value by some positive constant.

(A3) The linear kernel $L(\cdot)$ is a symmetric around zero linear density and

(A3-1)

$$0 < \int_{\mathbb{R}} L^2(u)du < \infty, \quad 0 < \int_{\mathbb{R}} u^2 L^2(u)du < \infty.$$

(A3-2)

$$0 < \int_{\mathbb{R}} {L^{(q)}}^2(u)du < \infty, \quad 0 < \int_{\mathbb{R}} u^2 {L^{(q)}}^2(u)du < \infty, \quad \text{for} \quad q = 1, 2.$$

(A4) The circular kernel $K_\kappa(\cdot)$ satisfies $K_\kappa(\cdot) = c_\kappa K[\kappa(1 - \cos(\cdot))]$ where c_κ is a normalization constant, $K : [0, \infty) \to [0, \infty)$ with

(A4-1)

$$0 < \int_0^\infty r^{\frac{j-1}{2}} K^l(r)dr < \infty, \quad \text{for } l = 1, 2,\ j = 0, 2.$$

(A4-2)

$$0 < \int_0^\infty r^{\frac{j-1}{2}} {K^{(q)}}^2(r)dr < \infty \quad \text{and} \quad 0 < \int_0^\infty r^{\frac{j-1}{2}} K''(r)K'(r)dr < \infty,$$

for $l = 1, 2,\ \ q = 1, 2,\ j = 0, 2, 4.$

Note that conditions **(A1)** and **(A2)** are standard smoothing conditions on the true density required in order to obtain the convergence rates of the partial derivatives of the joint density estimators. Condition **(A3)** is also a usual condition on kernel density estimation literature and is satisfied by many kernels in practice, such as the normal kernel. The last condition, **(A4)**, was required by Zhang and Chen (2021a) in order to obtain the asymptotic convergence rates of the derivatives of the kernel density estimator in the unit hypersphere and, as noted by the authors, integrating by parts shows that the von Mises kernel satisfies **(A4)**. For the different results, when referring to condition **(A3)**, both sub-conditions **(A3-1)** and **(A3-2)** will be alluded to. For some results, not so much regularity is needed, and only the first condition will be mentioned. The same holds for assumption **(A4)**.

For the next block of assumptions, a function class will be defined for each regression setting (circular-linear, linear-circular or circular circular). The defined function classes are the following:

- Circular-linear setting

$$\mathcal{KL} = \left\{ K\{\kappa[1 - \cos(\cdot - \theta)]\} \frac{\partial^q}{\partial y^q} L\left(\frac{\cdot - y}{h}\right) : (\theta, y) \in \mathbb{T} \times \mathbb{R}, \kappa, h > 0,\ q = 0, 1, 2 \right\},$$

- Linear-circular setting

$$\mathcal{LK} = \left\{ L\left(\frac{\cdot - x}{h}\right) \frac{\partial^q}{\partial \phi^q} K\{\kappa[1 - \cos(\cdot - \phi)]\} : (x, \phi) \in \mathbb{R} \times \mathbb{T}, h, \kappa > 0,\ q = 0, 1, 2 \right\},$$

- Circular-circular setting

$$\mathcal{KK} = \left\{ K\{\upsilon[1-\cos(\cdot-\theta)]\}\frac{\partial^q}{\partial\phi^q}K\{\kappa[1-\cos(\cdot-\phi)]\} : (\theta,\phi)\in\mathbb{T}\times\mathbb{T}, \upsilon,\kappa>0,\ q=0,1,2 \right\}.$$

It will be assumed that the following conditions are satisfied:

(B1) The class of functions $\mathcal{LK}$ is a VC-type class, *i.e.*, it satisfies (C.18) for some $A, u > 0$ and all probability measures P.

(B2) The class of functions $\mathcal{KL}$ is a VC-type class, *i.e.*, it satisfies (C.18) for some $A, u > 0$ and all probability measures P.

(B3) The class of functions $\mathcal{KK}$ is a VC-type class, *i.e.*, it satisfies (C.18) for some $A, u > 0$ and all probability measures P.

Conditions **(B1)**, **(B2)** and **(B3)** correspond to standard conditions on the kernels in order to obtain the uniform convergence rates (Giné and Guillou, 2002; Einmahl and Mason, 2005). Such conditions can be satisfied in practice by considering, for example, the Gaussian kernel in the linear case and the von Mises kernel in the circular case.

With the objective of obtaining the asymptotic error rates for the multimodal estimators, it is also necessary to assume the following conditions on the joint densities:

(C1) There exists $\lambda_1 > 0$ such that for any $(\theta, y) \in \mathbb{T}\times\mathbb{R}$ it holds $\frac{\partial}{\partial y}f(\theta,y) = 0$ and $\left|\frac{\partial^2}{\partial y^2}f(\theta,y)\right| < \lambda_1$. In addition, for each $\theta\in\mathbb{T}$ there exist $B_1,\rho_1>0$ such that

$$\left\{ y\in\Omega : \max_{y\in\Omega}\left|\frac{\partial}{\partial y}f(\theta,y)\right| \leqslant B_1, \frac{\partial^2}{\partial^2 y}f(\theta,y)\leqslant -\lambda_1 < 0 \right\} \subset M(\theta)\oplus\rho_1,$$

where $A\oplus r = \{x\in\mathbb{R} : \inf_{z\in A}|z-x| < r\}$.

(C2) There exists $\lambda_2 > 0$ such that for any $(\delta,\phi)\in \mathrm{Supp}(\Delta)\times\mathbb{T}$ it holds $\frac{\partial}{\partial\phi}f(\delta,\phi) = 0$, $\left|\frac{\partial^2}{\partial\phi^2}f(\delta,\phi)\right| < \lambda_2$. In addition, for each $\delta\in\mathrm{Supp}(\Delta)$ there exist $B_2,\rho_2>0$ such that

$$\left\{ \phi\in\mathbb{T} : \max_{\phi\in\mathbb{T}}\left|\frac{\partial}{\partial\phi}f(\delta,\phi)\right| \leqslant B_2, \frac{\partial^2}{\partial^2\phi}f(\delta,\phi)\leqslant -\lambda_2 < 0 \right\} \subset M(\delta)\oplus\rho_2,$$

where $A\oplus r = \{\gamma\in\mathbb{T} : \inf_{z\in A}(1-\cos(z-\gamma)) < r\}$.

Conditions **(C1)** and **(C2)** are imposed so that every conditional local mode is isolated from other critical values and so that points with near zero first partial derivative and negative second partial derivative are close to a conditional local mode. This assumptions eliminate the cases where the *branches* of the regression multifunction cross each other (Chen et al., 2016a).

Auxiliary lemmas

The proofs of Propositions 5.1, 5.2 and 5.3 require the pointwise and uniform consistency of the joint kernel density estimators and their first two partial derivatives (with respect to the response variable). Here, six lemmas are enunciated and proved. Lemmas C.3, C.4 and C.5 state the pointwise consistency of the joint kernel density estimators (and their partial derivatives) for, respectively, the circular-linear, linear-circular and circular-circular settings. In addition, Lemmas C.6, C.7 and C.8 state the uniform consistency of the same estimators.

Lemma C.3. *Under conditions **(A1)**,**(A3)** and **(A4-1)**, for fixed* $(\theta, y) \in \mathbb{T} \times \mathbb{R}$,

$$\frac{\partial^q}{\partial y^q}\hat{f}(\theta, y) - \frac{\partial^q}{\partial y^q} f(\theta, y) = O(\kappa^{-1} + h^2) + O_P\left(\sqrt{\frac{\kappa^{1/2}}{nh^{1+2q}}}\right) \quad \text{as } nh^{1+2q}\kappa^{-1/2} \to \infty,$$

for $q = 0, 1, 2$.

Proof. The derivatives of the cylindrical kernel density estimator with respect to the real-valued components are given by

$$\begin{aligned}\frac{\partial^q}{\partial y^q}\hat{f}(\theta, y) &= \frac{(-1)^q}{nh^q}\sum_{i=1}^{n} L_h^{(q)}(Y_i - y)K_\kappa(\Theta_i - \theta_0) \\ &= (-1)^q \frac{c_\kappa}{nh^{q+1}}\sum_{i=1}^{n} L^{(q)}\left(\frac{Y_i - y}{h}\right) K\{\kappa[1 - \cos(\Theta_i - \theta)]\},\end{aligned}$$

where the form of the circular kernel in **(A4)** was assumed. In what follows, the bias and variance of $\frac{\partial^q}{\partial y^q}\hat{f}(\theta, y)$ will be derived. Note that the case $q = 0$ was already proved by García-Portugués et al. (2013) in the context of hyperspherical data.

For the bias, it holds that, for $q = 0, 1, 2$,

$$\begin{aligned}&\text{Bias}\left[\frac{\partial^q}{\partial y^q}\hat{f}(\theta, y)\right] = \mathbb{E}\left[\frac{\partial^q}{\partial y^q}\hat{f}(\theta, y)\right] - \frac{\partial^q}{\partial y^q} f(\theta, y) \\ &= \mathbb{E}\left[(-1)^q \frac{c_\kappa}{nh^{q+1}}\sum_{i=1}^{n} L^{(q)}\left(\frac{Y_i - y}{h}\right) K\{\kappa[1 - \cos(\Theta_i - \theta)]\}\right] - \frac{\partial^q}{\partial y^q} f(\theta, y) \\ &= (-1)^q \frac{c_\kappa}{h^{q+1}}\int_0^{2\pi}\int_{\mathbb{R}} L^{(q)}\left(\frac{z - y}{h}\right) K\{\kappa[1 - \cos(\alpha - \theta)]\} f(\alpha, z)dzd\alpha - \frac{\partial^q}{\partial y^q} f(\theta, y) \\ &= (-1)^q c_\kappa \int_0^{2\pi} K\{\kappa[1 - \cos(\alpha - \theta)]\}\frac{1}{h^{q+1}}\int_{\mathbb{R}} L^{(q)}\left(\frac{z - y}{h}\right) f(\alpha, z)dzd\alpha - \frac{\partial^q}{\partial y^q} f(\theta, y).\end{aligned}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Note that integrating by parts yields

$$\frac{1}{h^{q+1}}\int_{\mathbb{R}} L^{(q)}\left(\frac{z - y}{h}\right) f(\alpha, z)dz = -\frac{1}{h^q}\int_{\mathbb{R}} L^{(q-1)}\left(\frac{z - y}{h}\right)\frac{\partial}{\partial y} f(\alpha, z)dz.$$

Consequently, integrating by parts q times yields

$$\begin{aligned}&\text{Bias}\left[\frac{\partial^q}{\partial y^q}\hat{f}(\theta,y)\right]\\&=\frac{c_\kappa}{h}\int_0^{2\pi}\int_{\mathbb{R}} L\left(\frac{z-y}{h}\right)K\{\kappa[1-\cos(\alpha-\theta)]\}\left[\frac{\partial^q}{\partial y^q}f(\alpha,z)-\frac{\partial^q}{\partial y^q}f(\theta,y)\right]dzd\alpha.\end{aligned}$$

After employing the changes of variables $u=(z-y)/h$ and $\varphi=\alpha-\theta$, it holds

$$\begin{aligned}&\text{Bias}\left[\frac{\partial^q}{\partial y^q}\hat{f}(\theta,y)\right]\\&=c_\kappa\int_0^{2\pi}\int_{\mathbb{R}} L(u)\,K[\kappa(1-\cos\varphi)]\left[\frac{\partial^q}{\partial y^q}f(\theta+\varphi,y+uh)-\frac{\partial^q}{\partial y^q}f(\theta,y)\right]dud\varphi\\&=c_\kappa\int_{\mathbb{R}}\int_0^{\pi} L(u)\,K[\kappa(1-\cos\varphi)]\left[\frac{\partial^q}{\partial y^q}f(\theta+\varphi,y+uh)-\frac{\partial^q}{\partial y^q}f(\theta,y)\right]dud\varphi\\&+c_\kappa\int_{\mathbb{R}}\int_{\pi}^{2\pi} L(u)\,K[\kappa(1-\cos\varphi)]\left[\frac{\partial^q}{\partial y^q}f(\theta+\varphi,y+uh)-\frac{\partial^q}{\partial y^q}f(\theta,y)\right]dud\varphi.\end{aligned}$$

Next, the change of variables $r=\kappa(1-\cos\varphi)$ is employed. For the first integral, $\varphi\in[0,\pi]$, which gives $\varphi=\arccos(1-r/\kappa)$ and $d\varphi=\kappa^{-1}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}dr$. On the other hand, for the second integral it holds that $\varphi\in[\pi,2\pi]$, and thus, $\varphi=-\arccos(1-r/\kappa)$ and $d\varphi=-\kappa^{-1}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}dr$. Consequently, the bias can be expressed as

$$\begin{aligned}&\text{Bias}\left[\frac{\partial^q}{\partial y^q}\hat{f}(\theta,y)\right]\\&=\frac{c_\kappa}{\kappa}\int_{\mathbb{R}}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2} L(u)\,K(r)\left[\frac{\partial^q}{\partial y^q}f[\theta+\arccos(1-r/\kappa),y+uh]\right.\\&\left.-\frac{\partial^q}{\partial y^q}f(\theta,y)\right]dudr\\&+\frac{c_\kappa}{\kappa}\int_{\mathbb{R}}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2} L(u)\,K(r)\left[\frac{\partial^q}{\partial y^q}f[\theta-\arccos(1-r/\kappa),y+uh]\right.\\&\left.-\frac{\partial^q}{\partial y^q}f(\theta,y)\right]dudr.\end{aligned}\tag{C.21}$$

In what follows, the next Taylor expansion is used, where the sine-approximation given by Di Marzio et al. (2009) for the circular variable is employed:

$$\begin{aligned}&\frac{\partial^q}{\partial y^q}f[\theta\pm\arccos(1-r/\kappa),y+uh]-\frac{\partial^q}{\partial y^q}f(\theta,y)\\&=uh\frac{\partial^{q+1}}{\partial y^{q+1}}f(\theta,y)\pm\sin[\arccos(1-r/\kappa)]\frac{\partial^{q+1}}{\partial y^q\partial\theta}f(\theta,y)+\frac{1}{2}u^2h^2\frac{\partial^{q+2}}{\partial y^{q+2}}f(\theta,y)\end{aligned}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$
\begin{aligned}
&+ \frac{1}{2}\sin^2[\arccos(1-r/\kappa)]\frac{\partial^{q+2}}{\partial y^q \partial\theta^2}f(\theta,y) \pm uh\sin[\arccos(1-r/\kappa)]\frac{\partial^{q+2}}{\partial y^{q+1}\partial\theta}f(\theta,y)\\
&+ o(u^2h^2) + r\cdot o\{\sin^2[\arccos(1-r/\kappa)]\}\\
&= uh\frac{\partial^{q+1}}{\partial y^{q+1}}f(\theta,y) \pm \left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}\frac{\partial^{q+1}}{\partial y^q\partial\theta}f(\theta,y) + \frac{1}{2}u^2h^2\frac{\partial^{q+2}}{\partial y^{q+2}}f(\theta,y)\\
&+ \frac{1}{2}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)\frac{\partial^{q+2}}{\partial y^q\partial\theta^2}f(\theta,y) \pm uh\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}\frac{\partial^{q+2}}{\partial y^{q+1}\partial\theta}f(\theta,y)\\
&+ u^2\cdot o(h^2) + r\cdot o(\kappa^{-1}).
\end{aligned}
$$

Plugging these expansions into (C.21) yields

$$
\begin{aligned}
&\text{Bias}\left[\frac{\partial^q}{\partial y^q}\hat{f}(\theta,y)\right]\\
&= h\frac{c_\kappa}{\kappa}\frac{\partial^{q+1}}{\partial y^{q+1}}f(\theta,y)\underbrace{\int_{\mathbb{R}} uL(u)du}_{=0}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}K(r)dr\\
&+ \frac{c_\kappa}{\kappa}\frac{\partial^{q+1}}{\partial y^q\partial\theta}f(\theta,y)\underbrace{\int_{\mathbb{R}} L(u)du}_{=1}\int_0^{2\kappa}K(r)dr\\
&+ h^2\frac{c_\kappa}{2\kappa}\frac{\partial^{q+2}}{\partial y^{q+2}}f(\theta,y)\underbrace{\int_{\mathbb{R}} u^2L(u)du}_{=\mu_2(L)}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}K(r)dr\\
&+ \frac{c_\kappa}{2\kappa}\frac{\partial^{q+2}}{\partial y^q\partial\theta^2}f(\theta,y)\underbrace{\int_{\mathbb{R}} L(u)du}_{=1}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}K(r)dr\\
&+ h\frac{c_\kappa}{2\kappa}\frac{\partial^{q+2}}{\partial y^{q+1}\partial\theta}f(\theta,y)\underbrace{\int_{\mathbb{R}} uL(u)du}_{=0}\int_0^{2\kappa}K(r)dr\\
&+ h\frac{c_\kappa}{\kappa}\frac{\partial^{q+1}}{\partial y^{q+1}}f(\theta,y)\underbrace{\int_{\mathbb{R}} uL(u)du}_{=0}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}K(r)dr\\
&- \frac{c_\kappa}{\kappa}\frac{\partial^{q+1}}{\partial y^q\partial\theta}f(\theta,y)\underbrace{\int_{\mathbb{R}} L(u)du}_{=1}\int_0^{2\kappa}K(r)dr\\
&+ h^2\frac{c_\kappa}{2\kappa}\frac{\partial^{q+2}}{\partial y^{q+2}}f(\theta,y)\underbrace{\int_{\mathbb{R}} u^2L(u)du}_{=\mu_2(L)}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}K(r)dr
\end{aligned}
$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$
\begin{aligned}
&+ \frac{c_\kappa}{2\kappa}\frac{\partial^{q+2}}{\partial y^q \partial\theta^2} f(\theta,y) \underbrace{\int_{\mathbb{R}} L(u)du}_{=1} \int_0^{2\kappa} \left(\frac{2r}{\kappa} - \frac{r^2}{\kappa^2}\right)^{1/2} K(r)dr \\
&- h\frac{c_\kappa}{2\kappa}\frac{\partial^{q+2}}{\partial y^{q+1} \partial\theta} f(\theta,y) \underbrace{\int_{\mathbb{R}} uL(u)du}_{=0} \int_0^{2\kappa} K(r)dr, \\
&+ \frac{c_\kappa}{\kappa} \underbrace{\int_{\mathbb{R}} u^2 L(u)du}_{=\mu_2(L)} \int_0^{2\kappa} \left(\frac{2r}{\kappa} - \frac{r^2}{\kappa^2}\right)^{-1/2} K(r)dr \cdot o(h^2) \\
&+ \frac{c_\kappa}{\kappa} \underbrace{\int_{\mathbb{R}} L(u)du}_{=1} \int_0^{2\kappa} r \left(\frac{2r}{\kappa} - \frac{r^2}{\kappa^2}\right)^{-1/2} K(r)dr \cdot o(\kappa^{-1}),
\end{aligned}
$$

where some of the addends are zero due to the symmetry of $L(\cdot)$. Note that the second and seventh terms cancel each other out and the third and eighth and fourth and ninth terms are the same. Thus, the bias can be expressed as

$$
\begin{aligned}
&\text{Bias}\left[\frac{\partial^q}{\partial y^q}\hat{f}(\theta,y)\right] \\
&= h^2 \frac{c_\kappa}{\kappa}\mu_2(L)\frac{\partial^{q+2}}{\partial y^{q+2}} f(\theta,y) \int_0^{2\kappa} \left(\frac{2r}{\kappa} - \frac{r^2}{\kappa^2}\right)^{-1/2} K(r)dr \\
&+ \frac{c_\kappa}{\kappa}\frac{\partial^{q+2}}{\partial y^q \partial\theta^2} f(\theta,y) \int_0^{2\kappa} \left(\frac{2r}{\kappa} - \frac{r^2}{\kappa^2}\right)^{1/2} K(r)dr \\
&+ \frac{c_\kappa}{\kappa}\mu_2(L) \int_0^{2\kappa} \left(\frac{2r}{\kappa} - \frac{r^2}{\kappa^2}\right)^{-1/2} K(r)dr \cdot o(h^2) \\
&+ \frac{c_\kappa}{\kappa} \int_0^{2\kappa} r \left(\frac{2r}{\kappa} - \frac{r^2}{\kappa^2}\right)^{-1/2} K(r)dr \cdot o(\kappa^{-1}) \\
&= h^2 \frac{c_\kappa}{\kappa^{1/2}}\mu_2(L)\frac{\partial^{q+2}}{\partial y^{q+2}} f(\theta,y) \int_0^{2\kappa} r^{-1/2} (2 - r/\kappa)^{-1/2} K(r)dr \\
&+ \frac{c_\kappa}{\kappa^{3/2}}\frac{\partial^{q+2}}{\partial y^q \partial\theta^2} f(\theta,y) \int_0^{2\kappa} r^{1/2} (2 - r/\kappa)^{1/2} K(r)dr \\
&+ \frac{c_\kappa}{\kappa^{1/2}}\mu_2(L) \int_0^{2\kappa} r^{-1/2} (2 - r/\kappa)^{-1/2} K(r)dr \cdot o(h^2) \\
&+ \frac{c_\kappa}{\kappa^{1/2}} \int_0^{2\kappa} r^{1/2} (2 - r/\kappa)^{-1/2} K(r)dr \cdot o(\kappa^{-1}).
\end{aligned}
$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Then, taking into account (C.6) and Lemma C.1, the bias can be expressed as

$$\begin{aligned}
&\text{Bias}\left[\frac{\partial^q}{\partial y^q}\hat{f}(\theta,y)\right]\\
&=\frac{h^2}{\kappa^{1/2}}\mu_2(L)\frac{\partial^{q+2}}{\partial y^{q+2}}f(\theta,y)\left(\frac{\kappa^{1/2}2^{-1/2}\int_0^\infty r^{-1/2}K(r)dr}{2^{1/2}\int_0^\infty r^{-1/2}K(r)dr}+o(1)\right)\\
&+\frac{1}{\kappa^{3/2}}\frac{\partial^{q+2}}{\partial y^q\partial\theta^2}f(\theta,y)\left(\frac{\kappa^{1/2}2^{1/2}\int_0^\infty r^{1/2}K(r)dr}{2^{1/2}\int_0^\infty r^{-1/2}K(r)dr}+o(1)\right)\\
&+\frac{1}{\kappa^{1/2}}\mu_2(L)\left(\frac{\kappa^{1/2}2^{-1/2}\int_0^\infty r^{-1/2}K(r)dr}{2^{1/2}\int_0^\infty r^{-1/2}K(r)dr}+o(1)\right)\cdot o(h^2)\\
&+\frac{1}{\kappa^{1/2}}\left(\frac{\kappa^{1/2}2^{-1/2}\int_0^\infty r^{1/2}K(r)dr}{2^{1/2}\int_0^\infty r^{-1/2}K(r)dr}+o(1)\right)\cdot o(\kappa^{-1})\\
&=\frac{h^2}{2}\mu_2(L)\frac{\partial^{q+2}}{\partial y^{q+2}}f(\theta,y)+\frac{1}{\kappa}b_2(K)\frac{\partial^{q+2}}{\partial y^q\partial\theta^2}f(\theta,y)+o(h^2)+o(\kappa^{-1})\\
&=O(h^2+\kappa^{-1}),
\end{aligned}$$

where, as in (C.8),

$$b_j(K)=\frac{\int_0^\infty r^{\frac{j-1}{2}}K(r)dr}{\int_0^\infty r^{-1/2}K(r)dr}.$$

Now the focus is set to the derivation of the variance. It holds

$$\begin{aligned}
\text{Var}\left[\frac{\partial^q}{\partial y^q}\hat{f}(\theta,y)\right]&=\frac{1}{nh^{2+2q}}\text{Var}\left[L^{(q)}\left(\frac{Y_1-y}{h}\right)K_\kappa(\Theta_1-\theta)\right]\\
&=\underbrace{\frac{1}{nh^{2+2q}}\mathbb{E}\left[L^{(q)^2}\left(\frac{Y_1-y}{h}\right)K_\kappa^2(\Theta_1-\theta)\right]}_{=(A)}-\underbrace{\frac{1}{n}\left\{\mathbb{E}\left[\frac{\partial^q}{\partial y^q}\hat{f}(\theta,y)\right]\right\}^2}_{=(B)}. \qquad \text{(C.22)}
\end{aligned}$$

By applying the same changes of variables as in the bias calculations, *i.e.*, $u=(z-y)/h$, $\varphi=\alpha-\theta$ first and $r=\kappa(1-\cos\varphi)$ later, it can be obtained that

$$\begin{aligned}
(A)&=\frac{c_\kappa^2}{nh^{2+2q}}\int_0^{2\pi}\int_{\mathbb{R}}L^{(q)^2}\left(\frac{z-y}{h}\right)K^2\{\kappa[1-\cos(\alpha-\theta)]\}f(\alpha,z)dzd\alpha\\
&=\frac{c_\kappa^2}{nh^{1+2q}}\int_0^{\pi}\int_{\mathbb{R}}L^{(q)^2}(u)\,K^2[\kappa(1-\cos\varphi)]f(\theta+\varphi,y+uh)dud\varphi\\
&+\frac{c_\kappa^2}{nh^{1+2q}}\int_{\pi}^{2\pi}\int_{\mathbb{R}}L^{(q)^2}(u)\,K^2[\kappa(1-\cos\varphi)]f(\theta+\varphi,y+uh)dud\varphi\\
&=\frac{c_\kappa^2}{nh^{1+2q}\kappa}\int_0^{2\kappa}\int_{\mathbb{R}}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}L^{(q)^2}(u)\,K^2(r)f[\theta+\arccos(1-r/\kappa),y+uh]dudr
\end{aligned}$$

USC DE COMPOSTELA

$$+\frac{c_\kappa^2}{nh^{1+2q}\kappa}\int_0^{2\kappa}\int_{\mathbb{R}}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}{L^{(q)}}^2(u)\,K^2(r)f[\theta-\arccos(1-r/\kappa),y+uh]dudr.$$

The following Taylor expansion is used:

$$\begin{aligned}f[\theta\pm\arccos(1-r/\kappa),y+uh] &= f(\theta,y)+uh\frac{\partial}{\partial y}f(\theta,y)\pm\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}\frac{\partial}{\partial\theta}f(\theta,y)\\ &+\frac{1}{2}u^2h^2\frac{\partial^2}{\partial y^2}f(\theta,y)+\frac{1}{2}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)\frac{\partial^2}{\partial\theta^2}f(\theta,y)\pm uh\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}\frac{\partial^2}{\partial y\partial\theta}f(\theta,y)\\ &+u^2o(h^2)+ro(\kappa^{-1}).\end{aligned}$$

Plugging these expansions into the expression of (A), it is obtained, after some rearrangement, that

$$\begin{aligned}(A) &= \frac{2c_\kappa^2}{nh^{1+2q}\kappa}f(\theta,y)\underbrace{\int_{\mathbb{R}}{L^{(q)}}^2(u)\,du}_{=R\left(L^{(q)}\right)}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}K^2(r)dr\\ &+\frac{c_\kappa^2}{nh^{2q-1}\kappa}\frac{\partial^2}{\partial y^2}f(\theta,y)\underbrace{\int_{\mathbb{R}}u^2{L^{(q)}}^2(u)\,du}_{=\mu_2\left(L^{(q)}\right)}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}K^2(r)dr\\ &+\frac{c_\kappa^2}{nh^{1+2q}\kappa}\frac{\partial^2}{\partial\theta^2}f(\theta,y)\underbrace{\int_{\mathbb{R}}{L^{(q)}}^2(u)\,du}_{=R\left(L^{(q)}\right)}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}K^2(r)dr\\ &+o(h^2)+o(\kappa^{-1})\\ &=\frac{2c_\kappa^2}{nh^{1+2q}\kappa^{1/2}}R\left(L^{(q)}\right)f(\theta,y)\int_0^{2\kappa}r^{-1/2}\left(2-r/\kappa\right)^{-1/2}K^2(r)dr\\ &+\frac{c_\kappa^2}{nh^{2q-1}\kappa^{1/2}}\mu_2\left(L^{(q)}\right)\frac{\partial^2}{\partial y^2}f(\theta,y)\int_0^{2\kappa}r^{-1/2}\left(2-r/\kappa\right)^{-1/2}K^2(r)dr\\ &+\frac{c_\kappa^2}{nh^{1+2q}\kappa^{3/2}}R\left(L^{(q)}\right)\frac{\partial^2}{\partial\theta^2}f(\theta,y)\int_0^{2\kappa}r^{1/2}\left(2-r/\kappa\right)^{1/2}K^2(r)dr\\ &+o(h^2)+o(\kappa^{-1}).\end{aligned}$$

By employing (C.6) and Lemma C.1, it holds

$$(A)=\frac{2c_\kappa}{nh^{1+2q}\kappa^{1/2}}R\left(L^{(q)}\right)f(\theta,y)\left(\frac{\kappa^{1/2}2^{-1/2}\int_0^\infty r^{-1/2}K^2(r)dr}{2^{1/2}\int_0^\infty r^{-1/2}K(r)dr}+o(1)\right)$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$
\begin{aligned}
&+ \frac{c_\kappa}{nh^{2q-1}\kappa^{1/2}}\mu_2\left(L^{(q)}\right)\frac{\partial^2}{\partial y^2}f(\theta,y)\left(\frac{\kappa^{1/2}2^{-1/2}\int_0^\infty r^{-1/2}K^2(r)dr}{2^{1/2}\int_0^\infty r^{-1/2}K(r)dr}+o(1)\right)\\
&+ \frac{c_\kappa}{nh^{1+2q}\kappa^{3/2}}R\left(L^{(q)}\right)\frac{\partial^2}{\partial\theta^2}f(\theta,y)\left(\frac{\kappa^{1/2}2^{1/2}\int_0^\infty r^{1/2}K^2(r)dr}{2^{1/2}\int_0^\infty r^{-1/2}K(r)dr}+o(1)\right)\\
&+ o(h^2)+o(\kappa^{-1}).\\
&= \frac{c_\kappa}{nh^{1+2q}}R\left(L^{(q)}\right)e_0(K)f(\theta,y)+\frac{c_\kappa}{2nh^{2q-1}}\mu_2\left(L^{(q)}\right)e_0(K)\frac{\partial^2}{\partial y^2}f(\theta,y)\\
&+ \frac{c_\kappa}{nh^{1+2q}\kappa}R\left(L^{(q)}\right)e_2(K)\frac{\partial^2}{\partial\theta^2}f(\theta,y)+o(h^2)+o(\kappa^{-1})\\
&= \frac{c_\kappa}{nh^{1+2q}}R\left(L^{(q)}\right)e_0(K)f(\theta,y)+o\left(\frac{\kappa^{1/2}}{nh^{1+2q}}\right),
\end{aligned}
$$

where

$$
e_j(K)=\frac{\int_0^\infty r^{\frac{j-1}{2}}K^2(r)dr}{\int_0^\infty r^{-1/2}K(r)dr}.
$$

From the bias calculations, it is known that the term (B) in (C.22) yields $o[n^{-1}(h^2+\kappa^{-1})]$. Consequently,

$$
\mathrm{Var}\left[\frac{\partial^q}{\partial y^q}\hat{f}(\theta,y)\right]=\frac{c_\kappa}{nh^{1+2q}}R\left(L^{(q)}\right)e_0(K)f(\theta,y)+o\left(\frac{\kappa^{1/2}}{nh^{1+2q}}\right)=O\left(\frac{\kappa^{1/2}}{nh^{1+2q}}\right).
$$

Finally, because of the Central Limit Theorem, it holds

$$
\frac{\partial^q}{\partial y^q}\hat{f}(\theta,y)-\frac{\partial^q}{\partial y^q}f(\theta,y)=O(\kappa^{-1}+h^2)+O_P\left(\sqrt{\frac{\kappa^{1/2}}{nh^{1+2q}}}\right).
$$

□

Lemma C.4. *Under conditions* ***(A2)***, ***(A3-1)*** *and* ***(A4)***, *for fixed* $(x,\phi)\in\mathbb{R}\times\mathbb{T}$,

$$
\frac{\partial^q}{\partial\phi^q}\hat{f}(x,\phi)-\frac{\partial^q}{\partial y^q}f(x,\phi)=O(\kappa^{-1}+h^2)+O_P\left(\sqrt{\frac{\kappa^{(1+2q)/2}}{nh}}\right)\quad \textit{as } nh\kappa^{\frac{-(1+2q)}{2}}\to\infty,
$$

for $q=0,1,2$.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Proof. Note that the case $q=0$ was already proven in Lemma C.3. The first two derivatives, with respect to the circular component, of the cylindrical kernel density estimator are given by

$$
\frac{\partial}{\partial\phi}\hat{f}(x,\phi)=-\frac{\kappa c_\kappa}{nh}\sum_{i=1}^n L\left(\frac{X_i-x}{h}\right)K'\{\kappa[1-\cos(\Phi_i-\phi)]\}\sin(\Phi_i-\phi)
$$

and

$$\frac{\partial^2}{\partial\phi^2}\hat{f}(x,\phi) = \frac{\kappa c_\kappa}{nh}\sum_{i=1}^{n} L\left(\frac{X_i - x}{h}\right)\left[\kappa K''\{\kappa[1-\cos(\Phi_i-\phi)]\}\sin^2(\Phi_i-\phi) + K'\{\kappa[1-\cos(\Phi_i-\phi)]\}\cos(\Phi_i-\phi)\right]$$

For $q = 1$, it holds that

$$\text{Bias}\left[\frac{\partial}{\partial\phi}\hat{f}(x,\phi)\right] = \mathbb{E}\left[\frac{\partial}{\partial\phi}\hat{f}(x,\phi)\right] - \frac{\partial}{\partial\phi}f(x,\phi)$$
$$= -\frac{\kappa c_\kappa}{h}\int_{\mathbb{R}}\int_0^{2\pi} L\left(\frac{z-x}{h}\right)K'\{\kappa[1-\cos(\alpha-\phi)]\}\sin(\alpha-\phi)f(z,\alpha)d\alpha dz - \frac{\partial}{\partial\phi}f(x,\phi).$$

Note that integrating by parts yields

$$\kappa\int_0^{2\pi} K'\{\kappa[1-\cos(\alpha-\phi)]\}\sin(\alpha-\phi)f(z,\alpha)d\alpha$$
$$= -\int_0^{2\pi} K\{\kappa[1-\cos(\alpha-\phi)]\}\frac{\partial}{\partial\phi}f(z,\alpha)d\alpha.$$

Consequently, for $q = 1$,

$$\text{Bias}\left[\frac{\partial}{\partial\phi}\hat{f}(x,\phi)\right]$$
$$= \frac{c_\kappa}{h}\int_{\mathbb{R}}\int_0^{2\pi} L\left(\frac{z-x}{h}\right)K\{\kappa[1-\cos(\alpha-\phi)]\}\frac{\partial}{\partial\phi}f(z,\alpha)d\alpha dz - \frac{\partial}{\partial\phi}f(x,\phi).$$

Furthermore, for $q = 2$,

$$\text{Bias}\left[\frac{\partial^2}{\partial\phi^2}\hat{f}(x,\phi)\right] = \mathbb{E}\left[\frac{\partial^2}{\partial\phi^2}\hat{f}(x,\phi)\right] - \frac{\partial^2}{\partial\phi^2}f(x,\phi)$$
$$= \frac{\kappa c_\kappa}{h}\int_{\mathbb{R}}\int_0^{2\pi} L\left(\frac{z-x}{h}\right)\left[\kappa K''\{\kappa[1-\cos(\alpha-\phi)]\}\sin^2(\alpha-\phi)\right.$$
$$\left.+K'\{\kappa[1-\cos(\alpha-\phi)]\}\cos(\alpha-\phi)\right]f(z,\alpha)d\alpha dz - \frac{\partial}{\partial\phi}f(x,\phi).$$

Integration by parts (twice) yields

$$\kappa\int_0^{2\pi}\left[\kappa K''\{\kappa[1-\cos(\alpha-\phi)]\}\sin^2(\alpha-\phi) + K'\{\kappa[1-\cos(\alpha-\phi)]\}\cos(\alpha-\phi)\right]f(z,\alpha)d\alpha$$
$$= -\kappa\int_0^{2\pi} K'\{\kappa[1-\cos(\alpha-\phi)]\}\sin(\alpha-\phi)\frac{\partial}{\partial\phi}f(z,\alpha)d\alpha$$
$$= \int_0^{2\pi} K\{\kappa[1-\cos(\alpha-\phi)]\}\frac{\partial^2}{\partial\phi^2}f(z,\alpha)d\alpha.$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Consequently, for $q = 1, 2$, the bias can be expressed as

$$
\begin{aligned}
&\text{Bias}\left[\frac{\partial^q}{\partial\phi^q}\hat{f}(x,\phi)\right] \\
&= c_\kappa \int_{\mathbb{R}}\int_0^{2\pi} L\left(\frac{z-x}{h}\right) K\{\kappa[1-\cos(\alpha-\phi)]\}\left[\frac{\partial^q}{\partial\phi^q}f(z,\alpha)-\frac{\partial^q}{\partial\phi^q}f(x,\phi)\right] d\alpha dz \\
&= c_\kappa \int_{\mathbb{R}}\int_0^{2\pi} L(u)\, K\{\kappa[1-\cos(\varphi)]\}\left[\frac{\partial^q}{\partial\phi^q}f(x+uh,\phi+\varphi)-\frac{\partial^q}{\partial\phi^q}f(x,\phi)\right] d\varphi du \\
&= \frac{c_\kappa}{\kappa}\int_{\mathbb{R}}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2} L(u)\, K(r)\left[\frac{\partial^q}{\partial\phi^q}f[x+uh,\phi+\arccos(1-r/\kappa)]\right. \\
&\left.-\frac{\partial^q}{\partial\phi^q}f(x,\phi)\right] dr du \\
&+\frac{c_\kappa}{\kappa}\int_{\mathbb{R}}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2} L(u)\, K(r)\left[\frac{\partial^q}{\partial\phi^q}f[x+uh,\phi-\arccos(1-r/\kappa)]\right. \\
&\left.-\frac{\partial^q}{\partial\phi^q}f(x,\phi)\right] dr du,
\end{aligned}
$$

where, in the second equality the changes of variables $u = (z-x)/h$ and $\varphi = \alpha - \phi$ were employed and, in the third equality, the change of variables $r = \kappa(1-\cos\varphi)$ was applied. The following Taylor expansion will be used:

$$
\begin{aligned}
&\frac{\partial^q}{\partial\phi^q}f\left[x+uh,\phi\pm\arccos(1-r/\kappa)\right]-\frac{\partial^q}{\partial\phi^q}f(x,\phi) \\
&= uh\frac{\partial^{q+1}}{\partial\phi^q\partial x}f(x,\phi)\pm\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}\frac{\partial^{q+1}}{\partial\phi^{q+1}}f(x,\phi)+\frac{1}{2}u^2h^2\frac{\partial^{q+2}}{\partial\phi^q\partial x^2}f(x,\phi) \\
&+\frac{1}{2}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)\frac{\partial^{q+2}}{\partial\phi^{q+2}}f(x,\phi)\pm uh\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}\frac{\partial^{q+2}}{\partial\phi^{q+1}\partial x}f(x,\phi)+u^2o(h^2)+ro(\kappa^{-1}).
\end{aligned}
$$

Plugging these expansions into the bias calculations, yields, after some rearrangements,

$$
\begin{aligned}
&\text{Bias}\left[\frac{\partial^q}{\partial\phi^q}\hat{f}(x,\phi)\right] \\
&= h^2\frac{c_\kappa}{\kappa}\frac{\partial^{q+2}}{\partial\phi^q\partial x^2}f(x,\phi)\underbrace{\int_{\mathbb{R}}u^2L(u)\,du}_{=\mu_2(L)}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}K(r)dr \\
&+\frac{c_\kappa}{\kappa}\frac{\partial^{q+2}}{\partial\phi^{q+2}}f(x,\phi)\underbrace{\int_{\mathbb{R}}L(u)\,du}_{=1}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}K(r)dr
\end{aligned}
$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$
\begin{aligned}
&+\frac{c_\kappa}{\kappa}\underbrace{\int_{\mathbb{R}} u^2 L(u)\,du}_{=\mu_2(L)}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}K(r)dr\cdot o(h^2)\\
&+\frac{c_\kappa}{\kappa}\underbrace{\int_{\mathbb{R}} L(u)\,du}_{=1}\int_0^{2\kappa} r\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}K(r)dr\cdot o(\kappa^{-1})\\
&=h^2\frac{c_\kappa}{\kappa^{1/2}}\mu_2(L)\frac{\partial^{q+2}}{\partial\phi^q\partial x^2}f(x,\phi)\int_0^{2\kappa} r^{-1/2}\left(2-r/\kappa\right)^{-1/2}K(r)dr\\
&+\frac{c_\kappa}{\kappa^{3/2}}\frac{\partial^{q+2}}{\partial\phi^{q+2}}f(x,\phi)\int_{\mathbb{R}}\mu_2(L)\int_0^{2\kappa} r^{1/2}(2-r/\kappa)^{1/2}K(r)dr\\
&+\frac{c_\kappa}{\kappa^{1/2}}\mu_2(L)\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}K(r)dr\cdot o(h^2)\\
&+\frac{c_\kappa}{\kappa^{1/2}}\int_0^{2\kappa} r^{1/2}\left(2-r/\kappa\right)^{-1/2}K(r)dr\cdot o(\kappa^{-1}).
\end{aligned}
$$

Now, applying (C.6) and Lemma C.1, it is finally obtained that

$$
\begin{aligned}
&\text{Bias}\left[\frac{\partial^q}{\partial\phi^q}\hat{f}(x,\phi)\right]\\
&=\frac{h^2}{\kappa^{1/2}}\mu_2(L)\frac{\partial^{q+2}}{\partial\phi^q\partial x^2}f(x,\phi)\left(\frac{\kappa^{1/2}2^{-1/2}\int_0^\infty r^{-1/2}K(r)dr}{2^{1/2}\int_0^\infty r^{-1/2}K(r)dr}+o(1)\right)\\
&+\frac{1}{\kappa^{3/2}}\frac{\partial^{q+2}}{\partial\phi^{q+2}}f(x,\phi)\left(\frac{\kappa^{1/2}2^{1/2}\int_0^\infty r^{1/2}K(r)dr}{2^{1/2}\int_0^\infty r^{-1/2}K(r)dr}+o(1)\right)\\
&+\frac{1}{\kappa^{1/2}}\mu_2(L)\left(\frac{\kappa^{1/2}2^{-1/2}\int_0^\infty r^{-1/2}K(r)dr}{2^{1/2}\int_0^\infty r^{-1/2}K(r)dr}+o(1)\right)\cdot o(h^2)\\
&+\frac{1}{\kappa^{1/2}}\left(\frac{\kappa^{1/2}2^{-1/2}\int_0^\infty r^{1/2}K(r)dr}{2^{1/2}\int_0^\infty r^{-1/2}K(r)dr}+o(1)\right)\cdot o(\kappa^{-1})\\
&=\frac{h^2}{2}\mu_2(L)\frac{\partial^{q+2}}{\partial\phi^q\partial x^2}f(x,\phi)+\frac{1}{\kappa}\frac{\partial^{q+2}}{\partial\phi^{q+2}}f(x,\phi)b_2(K)+o(h^2)+o(\kappa^{-1})\\
&=O(h^2+\kappa^{-1}).
\end{aligned}
$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Now, the variance calculations will be carried out. For $q=1$, it holds

$$
\text{Var}\left[\frac{\partial}{\partial\phi}\hat{f}(x,\phi)\right]=\frac{c_\kappa^2\kappa^2}{nh^2}\text{Var}\left[L\left(\frac{X_1-x}{h}\right)K'\{\kappa[1-\cos(\Phi_1-\phi)]\}\sin(\Phi_1-\phi)\right]
$$

$$= \underbrace{\frac{c_\kappa^2\kappa^2}{nh^2}\mathbb{E}\left[L^2\left(\frac{X_1-x}{h}\right)K'^2\{\kappa[1-\cos(\Phi_1-\phi)]\}\sin(\Phi_1-\phi)\right]}_{=(A)} - \underbrace{\frac{1}{n}\left\{\mathbb{E}\left[\frac{\partial}{\partial\phi}\hat{f}(x,\phi)\right]\right\}^2}_{=(B)}.$$

As in the bias calculations, the changes of variables $u=(z-x)/h$ and $\varphi=\alpha-\phi$ are applied first, and the change of variables $r=\kappa(1-\cos\varphi)$ is employed afterwards, to obtain

$$\begin{aligned}
(A) &= \frac{c_\kappa^2\kappa^2}{nh^2}\int_{\mathbb{R}}\int_0^{2\pi} L^2\left(\frac{z-x}{h}\right)K'^2\{\kappa[1-\cos(\alpha-\phi)]\}\sin^2(\alpha-\phi)f(z,\alpha)d\alpha dz\\
&= \frac{c_\kappa^2\kappa^2}{nh}\int_{\mathbb{R}}\int_0^{\pi} L^2(u)\,K'^2[\kappa(1-\cos\varphi)]\sin^2\varphi f(x+uh,\phi+\varphi)d\varphi du\\
&+ \frac{c_\kappa^2\kappa^2}{nh}\int_{\mathbb{R}}\int_\pi^{2\pi} L^2(u)\,K'^2[\kappa(1-\cos\varphi)]\sin^2\varphi f(x+uh,\phi+\varphi)d\varphi du\\
&= \frac{c_\kappa^2\kappa}{nh}\int_{\mathbb{R}}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2} L^2(u)\,K'^2(r)f[x+uh,\phi+\arccos(1-r/\kappa)]drdu\\
&+ \frac{c_\kappa^2\kappa}{nh}\int_{\mathbb{R}}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2} L^2(u)\,K'^2(r)f[x+uh,\phi-\arccos(1-r/\kappa)]drdu.
\end{aligned}$$

Employing the Taylor expansion

$$f[x+uh,\phi\pm\arccos(1-r/\kappa)] = f(x,\phi)+O(h+\kappa^{-1})$$

and plugging it into the previous expression gives

$$\begin{aligned}
(A) &= \frac{2c_\kappa^2\kappa}{nh}f(x,\phi)\underbrace{\int_{\mathbb{R}}L^2(u)du}_{=R(L)}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}K'^2(r)dr\\
&+ \frac{c_\kappa^2\kappa}{nh}\int_{\mathbb{R}}L^2(u)\,du\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}K'^2(r)dr\cdot O(h+\kappa^{-1})\\
&= \frac{2c_\kappa^2\kappa^{1/2}}{nh}R(L)f(x,\phi)\int_0^{2\kappa} r^{1/2}(2-r/\kappa)^{1/2}K'^2(r)dr\\
&+ \frac{c_\kappa^2\kappa^{1/2}}{nh}R(L)\int_0^{2\kappa} r^{1/2}(2-r/\kappa)^{1/2}K'^2(r)dr\cdot O(h+\kappa^{-1}),
\end{aligned}$$

where

$$R(L)=\int_{\mathbb{R}}L^2(u)du.$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Next, making use of (C.6) and Lemma C.1, leads to

$$(A) = \frac{2c_\kappa \kappa^{1/2}}{nh} R(L) f(x,\phi) \left(\frac{\kappa^{1/2} 2^{1/2} \int_0^\infty r^{1/2} K''^2(r) dr}{2^{1/2} \int_0^\infty r^{-1/2} K(r) dr} + o(1) \right)$$
$$+ \frac{c_\kappa}{nh} R(L) \left(\frac{\kappa^{1/2} 2^{1/2} \int_0^\infty r^{1/2} K''^2(r) dr}{2^{1/2} \int_0^\infty r^{-1/2} K(r) dr} + o(1) \right) \cdot O(h + \kappa^{-1}).$$

By denoting

$$e_{j,q}(K) = \frac{\int_0^\infty r^{\frac{j-1}{2}} {K^{(q)}}^2(r) dr}{\int_0^\infty r^{-1/2} K(r) dr},$$

it holds

$$(A) = \frac{2c_\kappa \kappa}{nh} R(L) f(x,\phi) e_{2,1}(K) + o\left(\frac{\kappa^{3/2}}{nh}\right).$$

From the bias calculations, the term (B) in (C.22) yields $o[n^{-1}(h^2 + \kappa^{-1})]$. Therefore,

$$\operatorname{Var}\left[\frac{\partial}{\partial\phi}\hat{f}(x,\phi)\right] = \frac{2c_\kappa \kappa}{nh} R(L) f(x,\phi) e_{2,1}(K) + o\left(\frac{\kappa^{3/2}}{nh}\right) + o[n^{-1}(h^2 + \kappa^{-1})]$$
$$= \frac{2c_\kappa \kappa}{nh} R(L) f(x,\phi) e_{2,1}(K) + o\left(\frac{\kappa^{3/2}}{nh}\right) = O\left(\frac{\kappa^{3/2}}{nh}\right).$$

Finally, the variance for $q = 2$ is computed. It holds

$$\begin{aligned}
&\operatorname{Var}\left[\frac{\partial^2}{\partial\phi^2}\hat{f}(x,\phi)\right] \\
&= \frac{c_\kappa^2 \kappa^2}{nh^2} \operatorname{Var}\left[L\left(\frac{X_1 - x}{h}\right) \left(\kappa K''^2\{\kappa[1-\cos(\Phi_1 - \phi)]\} \sin^2(\Phi_1 - \phi) \right.\right. \\
&\left.\left. + K'\{\kappa[1-\cos(\Phi_1 - \phi)]\} \cos(\Phi_1 - \phi)\right)\right] \\
&= \underbrace{\frac{c_\kappa^2 \kappa^2}{nh^2} \mathbb{E}\left[L^2\left(\frac{X_1 - x}{h}\right) \left(\kappa K''^2\{\kappa[1-\cos(\Phi_1 - \phi)]\} \sin^2(\Phi_1 - \phi) \right.\right.}_{=(A)} \\
&\underbrace{\left.\left. + K'\{\kappa[1-\cos(\Phi_1 - \phi)]\} \cos(\Phi_1 - \phi)\right)^2\right]} \\
&- \underbrace{\frac{1}{n}\left\{\mathbb{E}\left[\frac{\partial^2}{\partial^2\phi}\hat{f}(x,\phi)\right]\right\}^2}_{=(B)}.
\end{aligned} \tag{C.23}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The quantity (A) above can be expressed as

$$(A) = \frac{c_\kappa^2 \kappa^4}{nh^2} \int_{\mathbb{R}} \int_0^{2\pi} L^2\left(\frac{z - x}{h}\right) K''^2\{\kappa[1-\cos(\alpha - \phi)]\} \sin^4(\alpha - \phi) f(z,\alpha) d\alpha dz$$

$$
\begin{aligned}
&+\frac{c_\kappa^2\kappa^2}{nh^2}\int_{\mathbb{R}}\int_0^{2\pi} L^2\left(\frac{z-x}{h}\right)K'^2\{\kappa[1-\cos(\alpha-\phi)]\}\cos^2(\alpha-\phi)f(z,\alpha)d\alpha dz\\
&+\frac{2c_\kappa^2\kappa^3}{nh^2}\int_{\mathbb{R}}\int_0^{2\pi} L^2\left(\frac{z-x}{h}\right)K''\{\kappa[1-\cos(\alpha-\phi)]\}K'\{\kappa[1-\cos(\alpha-\phi)]\}\\
&\times\sin^2(\alpha-\phi)\cos(\alpha-\phi)f(z,\alpha)d\alpha dz.
\end{aligned}
$$

The changes of variables $u=(z-x)/h$ and $\varphi=\alpha-\phi$ are applied first, and $r=\kappa(1-\cos\varphi)$ is used afterwards, obtaining

$$
\begin{aligned}
(A)&=\frac{c_\kappa^2\kappa^4}{nh}\int_{\mathbb{R}}\int_0^{2\pi} L^2(u)\,K''^2[\kappa(1-\cos\varphi)]\sin^4\varphi f(x+uh,\phi+\varphi)d\varphi du\\
&+\frac{c_\kappa^2\kappa^2}{nh}\int_{\mathbb{R}}\int_0^{2\pi} L^2(u)\,K'^2[\kappa(1-\cos\varphi)]\cos^2\varphi f(x+uh,\phi+\varphi)d\varphi du\\
&+\frac{2c_\kappa^2\kappa^3}{nh}\int_{\mathbb{R}}\int_0^{2\pi} L^2(u)\,K''[\kappa(1-\cos\varphi)]K'[\kappa(1-\cos\varphi)]\\
&\times\sin^2\varphi\cos\varphi f(x+uh,\phi+\varphi)d\varphi du\\
&=\frac{c_\kappa^2\kappa^3}{nh}\int_{\mathbb{R}}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{3/2} L^2(u)\,K''^2(r)f[x+uh,\phi+\arccos(1-r/\kappa)]dr du\\
&+\frac{c_\kappa^2\kappa^3}{nh}\int_{\mathbb{R}}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{3/2} L^2(u)\,K''^2(r)f[x+uh,\phi-\arccos(1-r/\kappa)]dr du\\
&+\frac{c_\kappa^2\kappa}{nh}\int_{\mathbb{R}}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}\left[1-\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)\right] L^2(u)\,K'^2(r)\\
&\times f[x+uh,\phi+\arccos(1-r/\kappa)]dr du\\
&+\frac{c_\kappa^2\kappa}{nh}\int_{\mathbb{R}}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{-1/2}\left[1-\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)\right] L^2(u)\,K'^2(r)\\
&\times f[x+uh,\phi-\arccos(1-r/\kappa)]dr du\\
&+\frac{2c_\kappa^2\kappa^2}{nh}\int_{\mathbb{R}}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}\left(1-\frac{r}{\kappa}\right) L^2(u)\,K''(r)K'(r)\\
&\times f[x+uh,\phi+\arccos(1-r/\kappa)]dr du\\
&+\frac{2c_\kappa^2\kappa^2}{nh}\int_{\mathbb{R}}\int_0^{2\kappa}\left(\frac{2r}{\kappa}-\frac{r^2}{\kappa^2}\right)^{1/2}\left(1-\frac{r}{\kappa}\right) L^2(u)\,K''(r)K'(r)\\
&\times f[x+uh,\phi-\arccos(1-r/\kappa)]dr du.
\end{aligned}
$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Now, the Taylor expansion

$$
f[x+uh,\phi\pm\arccos(1-r/\kappa)]=f(x,\phi)+O(h+\kappa^{-1})
$$

is used. Plugging it into the expression of (A) implies

$$
\begin{aligned}
(A) = & \frac{2c_\kappa^2\kappa^3}{nh} f(x,\phi) \int_{\mathbb{R}} \int_0^{2\kappa} \left(\frac{2r}{\kappa} - \frac{r^2}{\kappa^2}\right)^{3/2} L^2(u)\, K''^2(r) dr du \\
& + \frac{2c_\kappa^2\kappa}{nh} f(x,\phi) \int_{\mathbb{R}} \int_0^{2\kappa} \left(\frac{2r}{\kappa} - \frac{r^2}{\kappa^2}\right)^{-1/2} \left[1 - \left(\frac{2r}{\kappa} - \frac{r^2}{\kappa^2}\right)\right] L^2(u)\, K'^2(r) dr du \\
& + \frac{4c_\kappa^2\kappa^2}{nh} f(x,\phi) \int_{\mathbb{R}} \int_0^{2\kappa} \left(\frac{2r}{\kappa} - \frac{r^2}{\kappa^2}\right)^{1/2} \left(1 - \frac{r}{\kappa}\right) L^2(u)\, K''(r) K'(r) dr du + o\left(\frac{\kappa^{5/2}}{nh}\right).
\end{aligned}
$$

Consequently, it holds that

$$
\begin{aligned}
(A) = & \frac{2c_\kappa^2\kappa^{3/2}}{nh} R(L) f(x,\phi) \int_0^{2\kappa} r^{3/2} (2 - r/\kappa)^{3/2} (u)\, K''^2(r) dr \\
& + \frac{2c_\kappa^2\kappa^{3/2}}{nh} R(L) f(x,\phi) \int_0^{2\kappa} r^{-1/2} (2 - r/\kappa)^{-1/2} K'^2(r) dr \\
& - \frac{2c_\kappa^2\kappa^{1/2}}{nh} R(L) f(x,\phi) \int_0^{2\kappa} r^{1/2} (2 - r/\kappa)^{1/2} K'^2(r) dr \\
& + \frac{4c_\kappa^2\kappa^{3/2}}{nh} R(L) f(x,\phi) \int_0^{2\kappa} r^{1/2} (2 - r/\kappa)^{1/2} K''(r) K'(r) dr \\
& - \frac{4c_\kappa^2\kappa^{1/2}}{nh} R(L) f(x,\phi) \int_0^{2\kappa} r^{3/2} (2 - r/\kappa)^{1/2} K''(r) K'(r) dr + o\left(\frac{\kappa^{5/2}}{nh}\right).
\end{aligned}
$$

Now, taking into account (C.6) and Lemma C.1, it can be obtained that

$$
\begin{aligned}
(A) = & \frac{2c_\kappa\kappa^{3/2}}{nh} R(L) f(x,\phi) \left(\frac{\kappa^{1/2} 2^{3/2} \int_0^\infty r^{3/2} K''^2(r) dr}{2^{1/2} \int_0^\infty r^{-1/2} K(r) dr} + o(1)\right) \\
& + \frac{2c_\kappa\kappa^{3/2}}{nh} R(L) f(x,\phi) \left(\frac{\kappa^{1/2} 2^{-1/2} \int_0^\infty r^{-1/2} K'^2(r) dr}{2^{1/2} \int_0^\infty r^{-1/2} K(r) dr} + o(1)\right) \\
& - \frac{2c_\kappa\kappa^{1/2}}{nh} R(L) f(x,\phi) \left(\frac{\kappa^{1/2} 2^{1/2} \int_0^\infty r^{1/2} K'^2(r) dr}{2^{1/2} \int_0^\infty r^{-1/2} K(r) dr} + o(1)\right) \\
& + \frac{4c_\kappa\kappa^{3/2}}{nh} R(L) f(x,\phi) \left(\frac{\kappa^{1/2} 2^{1/2} \int_0^\infty r^{1/2} K''(r) K'(r) dr}{2^{1/2} \int_0^\infty r^{-1/2} K(r) dr} + o(1)\right) \\
& - \frac{4c_\kappa\kappa^{1/2}}{nh} R(L) f(x,\phi) \left(\frac{\kappa^{1/2} 2^{1/2} \int_0^\infty r^{3/2} K''(r) K'(r) dr}{2^{1/2} \int_0^\infty r^{-1/2} K(r) dr} + o(1)\right) + o\left(\frac{\kappa^{5/2}}{nh}\right).
\end{aligned}
$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

By denoting

$$z_j(K) = \frac{\int_0^\infty r^{\frac{j-1}{2}} K''(r)K'(r)dr}{\int_0^\infty r^{-1/2}K(r)dr},$$

it is finally obtained that

$$\begin{aligned}(A) &= \frac{4c_\kappa\kappa^2}{nh}R(L)f(x,\phi)e_{4,2}(K) + \frac{c_\kappa\kappa^2}{nh}R(L)f(x,\phi)e_{0,1}(K) - \frac{2c_\kappa\kappa}{nh}R(L)f(x,\phi)e_{2,1}(K)\\ &+ \frac{4c_\kappa\kappa^2}{nh}R(L)f(x,\phi)z_2(K) + \frac{4c_\kappa\kappa}{nh}R(L)f(x,\phi)z_4(K) + o\left(\frac{\kappa^{5/2}}{nh}\right)\\ &= \frac{c_\kappa\kappa^2}{nh}R(L)f(x,\phi)[4e_{4,2}(K) + e_{0,1}(K) + 4z_2(K)] + o\left(\frac{\kappa^{5/2}}{nh}\right)\\ &= O\left(\frac{\kappa^{5/2}}{nh}\right).\end{aligned}$$

From the calculus of the bias, the term (B) in (C.23) is $o(n^{-1}(h^2+\kappa^{-1}))$. Consequently,

$$\text{Var}\left[\frac{\partial^q}{\partial\phi^q}\hat{f}(x,\phi)\right] = O\left(\frac{\kappa^{(1+2q)/2}}{nh}\right), \quad \text{for } q = 0,1,2.$$

Finally, because of the Central Limit Theorem, it holds that, for $q = 0,1,2$,

$$\frac{\partial^q}{\partial\phi^q}\hat{f}(x,\phi) - \frac{\partial^q}{\partial\phi^q}f(x,\phi) = O(\kappa^{-1}+h^2) + O_P\left(\sqrt{\frac{\kappa^{(1+2q)/2}}{nh}}\right).$$

□

Lemma C.5. *Under conditions **(A2)**, **(A3-1)** and **(A4)**, for fixed $(\theta,\phi) \in \mathbb{T}\times\mathbb{T}$,*

$$\frac{\partial^q}{\partial\phi^q}\hat{f}(\theta,\phi) - \frac{\partial^q}{\partial\phi^q}f(\theta,\phi) = O(\kappa^{-1}+\upsilon^{-1}) + O_P\left(\sqrt{\frac{\upsilon^{1/2}\kappa^{(1+2q)/2}}{n}}\right)$$

as $n\upsilon^{-1/2}\kappa^{-(1+2q)/2} \to \infty$, for $q = 0,1,2$.

Proof. The arguments of the proof are analogous to those of Lemma C.3 and Lemma C.4. □

Lemma C.6. *Under assumptions **(A1)**, **(A3)**, **(A4-1)** and **(B1)**, the uniform convergence rate of $\frac{\partial^q}{\partial y^q}\hat{f}(\theta,y)$ is given by*

$$\begin{aligned}\left\|\frac{\partial^q}{\partial y^q}\hat{f} - \frac{\partial^q}{\partial y^q}f\right\|_\infty &:= \sup_{(\theta,y)}\left|\frac{\partial^q}{\partial y^q}\hat{f}(\theta,y) - \frac{\partial^q}{\partial y^q}f(\theta,y)\right|\\ &= O(\kappa^{-1}+h^2) + O_P\left(\sqrt{\frac{\kappa^{1/2}|\log(h\kappa^{-1/2})|}{nh^{1+2q}}}\right),\end{aligned}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

as $h \to 0$, $\kappa \to \infty$, $nh^{1+2q}/[\kappa^{1/2}|\log(h\kappa^{-1/2})|] \to \infty$ *and* $|\log(h\kappa^{-1/2})|/(\log\log n) \to \infty$, *for* $q = 0, 1, 2$.

Proof. It holds that

$$\left\| \frac{\partial^q}{\partial y^q}\hat{f} - \frac{\partial^q}{\partial y^q} f \right\|_\infty \leqslant \left\| \frac{\partial^q}{\partial y^q}\hat{f} - \mathbb{E}\left(\frac{\partial^q}{\partial y^q}\hat{f}\right) \right\|_\infty + \left\| \mathbb{E}\left(\frac{\partial^q}{\partial y^q}\hat{f}\right) - \frac{\partial^q}{\partial y^q} f \right\|_\infty,$$

where the second term is deterministic and, because of Lemma C.3, its value is $O(\kappa^{-1} + h^2)$. Consequently, for what remains it will be proved that

$$\limsup_{n\to\infty} \sqrt{\frac{nh^{1+2q}}{\kappa^{1/2}\log(h^{-1}\kappa^{1/2})}} \left\| \frac{\partial^q}{\partial y^q}\hat{f} - \mathbb{E}\left(\frac{\partial^q}{\partial y^q}\hat{f}\right) \right\|_\infty = C \ a.s., \tag{C.24}$$

where $C^2 \leqslant M^2||f||_\infty||L^{(q)}||_2^2 v(K)$, M only depends on the VC characteristics of $\mathcal{KL}$ and $v(K) = \int_0^\infty r^{-1/2}K^2(r)dr$, which is finite because of assumption **(A4-1)**.

In order to prove (C.24), the proof of Theorem 2.3 in Giné and Guillou (2002) will be followed, with the necessary modifications. In the rest of the proof, the notation $h \equiv h_n$ and $\kappa \equiv \kappa_n$ will be used to make the dependence of the smoothing parameters on n explicit. Recall that the normalization constant of the circular kernel, $c_{\kappa_n} = \kappa_n^{1/2}\lambda_{\kappa_n}^{-1}(K)$, with $\lambda_{\kappa_n}(K) = 2^{1/2}\int_0^{2\kappa_n} r^{-1/2}(2 - r/\kappa_n)^{-1/2}K(r)dr$. For $k \in \mathbb{N}$ and for any $\lambda > 0$, it holds

$$\begin{aligned}
&\mathbb{P}\left\{ \max_{2^{k-1}<n\leqslant 2^k} \sqrt{\frac{nh_n^{1+2q}}{\kappa_n^{1/2}\log(h_n^{-1}\kappa_n^{1/2})}} \left\| \frac{\partial^q}{\partial y^q}\hat{f} - \mathbb{E}\left(\frac{\partial^q}{\partial y^q}\hat{f}\right) \right\|_\infty > \lambda \right\} \\
&\leqslant 9\mathbb{P}\left\{ \sqrt{\frac{\kappa_{2^k}^{1/2}}{2^{k-1}h_{2^k}\lambda_{\kappa_{2^k}}^2(K)\log(h_{2^k}^{-1}\kappa_{2^k}^{1/2})}} \sup_{\substack{(\theta,y)\in\mathbb{T}\times\mathbb{R} \\ h_{2^k}\leqslant h\leqslant h_{2^{k-1}} \\ \kappa_{2^{k-1}}\leqslant\kappa\leqslant\kappa_{2^k}}} \left| \sum_{i=1}^{2^k} \left[K\{\kappa[1-\cos(\Theta_i-\theta)]\} \right.\right.\right. \\
&\left.\left.\left. \times L^{(q)}\left(\frac{Y_i - y}{h}\right) - \mathbb{E}\left(K\{\kappa[1-\cos(\Theta_i-\theta)]\}L^{(q)}\left(\frac{Y_i-y}{h}\right)\right)\right] \right| > \frac{\lambda}{30} \right\},
\end{aligned} \tag{C.25}$$

where the Montgomery-Smith's maximal inequality in (C.20) was employed. Now, Theorem C.1 is applied to the subclass

$$\mathcal{LK}^k = \left\{ K\{\kappa[1-\cos(\cdot-\theta)]\}L^{(q)}\left(\tfrac{\cdot - y}{h}\right) : (\theta, y) \in \mathbb{T}\times\mathbb{R}, h_{2^k} \leqslant h \leqslant h_{2^{k-1}}, \kappa_{2^{k-1}} \leqslant \kappa \leqslant \kappa_{2^k} \right\},$$

using the same B and C_2 for all k. Now, note that

$$\mathrm{Var}\left[K\{\kappa[1-\cos(\Theta_1-\theta)]\}L^{(q)}\left(\tfrac{Y_1-y}{h}\right)\right] \leqslant \mathbb{E}\left[K^2\{\kappa[1-\cos(\Theta_1-\theta)]\}{L^{(q)}}^2\left(\tfrac{Y_1-y}{h}\right)\right]$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

and

$$\begin{aligned}
&\mathbb{E}\left[K^2\{\kappa[1-\cos(\Theta_1-\theta)]\}L^{(q)^2}\left(\frac{Y_1-y}{h}\right)\right]\\
&=\int_0^{2\pi}\int_{\mathbb{R}}K^2\{\kappa[1-\cos(\alpha-\theta)]\}L^{(q)^2}\left(\frac{z-y}{h}\right)f(\alpha,z)dzd\alpha\\
&=\frac{h}{\kappa^{1/2}}\int_0^{2\kappa}\int_{\mathbb{R}}r^{-1/2}(2-r/\kappa)^{-1/2}K^2(r)L^{(q)^2}(u)\,f[\theta+\arccos(1-r/\kappa),y+uh]dudr\\
&+\frac{h}{\kappa^{1/2}}\int_0^{2\kappa}\int_{\mathbb{R}}r^{-1/2}(2-r/\kappa)^{-1/2}K^2(r)L^{(q)^2}(u)\,f[\theta-\arccos(1-r/\kappa),y+uh]dudr\\
&\leqslant 2^{1/2}h\kappa^{-1/2}||f||_\infty||L^{(q)}||_2^2v(K),
\end{aligned}$$

where the same three changes of variables than in the proof of Lemma C.3 have been used. Consequently, it can be taken that

$$U_k=||L^{(q)}||_\infty||K||_\infty\quad\text{and}\quad\sigma_k^2=2^{1/2}h_{2^k}\kappa_{2^k}^{-1/2}||f||_\infty||L^{(q)}||_2^2v(K).$$

Since $h\kappa^{-1/2}\to 0$ and $nh/[\kappa^{1/2}|\log(h\kappa^{-1/2})|]\to\infty$, there exists $k_0<\infty$ such that, for all $k\geqslant k_0$,

$$\sigma_k<U_k/2\quad\text{and}\quad\sqrt{2^k}\sigma_k\geqslant U_k\sqrt{\log\frac{U_k}{\sigma_k}}.$$

Thus, the conditions to apply inequality (C.19) from Theorem C.1 (Giné and Guillou, 2002) are satisfied. In addition, there exists $k_1<\infty$ such that for all $k\geqslant k_1$,

$$\sigma_k\sqrt{2^k}\sqrt{\log\frac{U_k}{\sigma_k}}\leqslant\sqrt{2^{3/2}h_{2^k}\kappa_{2^k}^{-1/2}2^{k-1}\lambda_{\kappa_{2^k}}^2(K)||f||_\infty||L^{(q)}||_2^2v(K)\log\left(h_{2^k}^{-1}\kappa_{2^k}^{1/2}\right)}.$$

Therefore, it can be taken that $\lambda=30C_2\sqrt{2^{3/2}||f||_\infty||L^{(q)}||_2^2v(K)}$ in (C.25), where C_2 corresponds to the one in (C.19) and, thus, it is obtained that

$$\begin{aligned}
&\mathbb{P}\left\{\max_{2^{k-1}<n\leqslant 2^k}\sqrt{\frac{nh_n^{1+2q}}{\kappa_n^{1/2}\log(h_n^{-1}\kappa_n^{1/2})}}\left|\left|\frac{\partial^q}{\partial y^q}\hat{f}-\mathbb{E}\left(\frac{\partial^q}{\partial y^q}\hat{f}\right)\right|\right|_\infty>30C_2\sqrt{2^{3/2}||f||_\infty||L^{(q)}||_2^2v(K)}\right\}\\
&\leqslant 9L\exp\left(-\frac{D}{B}\log\frac{U_k}{\sigma_k}\right),
\end{aligned}\tag{C.26}$$

with $D=C_2\log[1+C_2/(4B)]$. Now, since

$$\frac{\log(U_k/\sigma_k)}{\log k}=\frac{\log\left[||L^{(q)}||_\infty^2||K||_\infty^2/\left(||L^{(q)}||_2^2v(K)||f||_\infty 2^{1/2}h_{2^k}\kappa_{2^k}^{-1/2}\right)\right]}{2\log k}\to\infty$$

because $|\log(h\kappa^{-1/2})|/(\log\log n)\to\infty$, the probabilities in (C.26) are summable, and Borel-Cantelli can be applied to obtain (C.24). □

Lemma C.7. *Under assumptions **(A2)**, **(A3-1)**, **(A4)** and **(B2)**, the uniform convergence rate of $\frac{\partial^q}{\partial y^q}\hat{f}(x,\phi)$ is given by*

$$\left|\left|\frac{\partial^q}{\partial\phi^q}\hat{f}-\frac{\partial^q}{\partial\phi^q}f\right|\right|_\infty := \sup_{(x,\phi)}\left|\frac{\partial^q}{\partial\phi^q}\hat{f}(x,\phi)-\frac{\partial^q}{\partial\phi^q}f(x,\phi)\right|$$
$$= O(\kappa^{-1}+h^2)+O_P\left(\sqrt{\frac{\kappa^{(1+2q)/2}|\log(h\kappa^{-1/2})|}{nh}}\right),$$

as $h\to 0$, $\kappa\to\infty$, $nh/[\kappa^{(1+2q)/2}|\log(h\kappa^{-1/2})|]\to\infty$ and $|\log(h\kappa^{-1/2})|/(\log\log n)\to\infty$, for $q=0,1,2$.

Proof. For $q=0$, the result was obtained in Lemma C.6. For $q=1,2$, the proof is analogous to that of Lemma C.6, but taking into account the following remarks.

For $q=1$, the constant C in (C.24) is, in this case, such that $C^2\leqslant M^2||f||_\infty||L||_2^2\tilde{v}(K)$, M only depends on the VC characteristics of $\mathcal{LK}$ and $\tilde{v}(K)=\int_0^\infty r^{1/2}K'^2(r)dr<\infty$.

On the other hand, inequality (C.25) is replaced by

$$\mathbb{P}\left\{\max_{2^{k-1}<n\leqslant 2^k}\sqrt{\frac{nh_n}{\kappa_n^{3/2}\log(h_n^{-1}\kappa_n^{1/2})}}\left|\left|\frac{\partial}{\partial\phi}\hat{f}-\mathbb{E}\left(\frac{\partial}{\partial\phi}\hat{f}\right)\right|\right|_\infty>\lambda\right\}$$
$$\leqslant 9\mathbb{P}\left\{\sqrt{\frac{\kappa_{2^k}^{3/2}}{2^{k-1}h_{2^k}\lambda_{\kappa_{2^k}}^2(K)\log(h_{2^k}^{-1}\kappa_{2^k}^{1/2})}}\sup_{\substack{(x,\phi)\in\mathbb{R}\times\mathbb{T}\\ h_{2^k}\leqslant h\leqslant h_{2^{k-1}}\\ \kappa_{2^{k-1}}\leqslant\kappa\leqslant\kappa_{2^k}}}\left|\sum_{i=1}^{2^k}\left[L\left(\frac{X_i-x}{h}\right)K'\{\kappa[1-\cos(\Phi_i-\phi)]\}\right.\right.\right.$$
$$\left.\left.\left.\times\sin(\Phi_i-\phi)-\mathbb{E}\left(L\left(\frac{X_i-x}{h}\right)K'\{\kappa[1-\cos(\Phi_i-\phi)]\}\sin(\Phi_i-\phi)\right)\right]\right|>\frac{\lambda}{30}\right\} \tag{C.27}$$

and it holds

$$\mathbb{E}\left[L^2\left(\frac{X_1-x}{h}\right)K'^2\{\kappa[1-\cos(\Phi_1-\phi)]\}\sin^2(\Phi_1-\phi)\right]$$
$$=\int_0^{2\pi}\int_{\mathbb{R}}L^2\left(\frac{z-x}{h}\right)K'^2\{\kappa[1-\cos(\alpha-\phi)]\}\sin^2(\alpha-\phi)f(z,\alpha)dzd\alpha$$
$$=\frac{h}{\kappa^{3/2}}\int_0^{2\kappa}\int_{\mathbb{R}}r^{1/2}(2-r/\kappa)^{1/2}L^2(u)K'^2(r)f[x+uh,\phi+\arccos(1-r/\kappa)]dudr$$
$$+\frac{h}{\kappa^{3/2}}\int_0^{2\kappa}\int_{\mathbb{R}}r^{1/2}(2-r/\kappa)^{1/2}L^2(u)K'^2(r)f[x+uh,\phi-\arccos(1-r/\kappa)]dudr$$
$$\leqslant 2^{3/2}h\kappa^{-3/2}||f||_\infty||L||_2^2\tilde{v}(K),$$

where the changes of variables $u=(z-y)/h$, $\varphi=\alpha-\theta$ and $r=\kappa(1-\cos\varphi)$ have been employed as in the previous proof. Therefore, it can be taken that

$$U_k=||L||_\infty||K'||_\infty \quad\text{and}\quad \sigma_k^2=2^{3/2}h_{2^k}\kappa_{2^k}^{-3/2}||f||_\infty||L||_2^2\tilde{v}(K)$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

to apply inequality (C.19) and $\lambda = 30C_2\sqrt{2^{5/2}||f||_\infty||L||_2^2\tilde{v}(K)}$ in (C.27), while the rest of the proof is analogous to that of Lemma C.6.

Finally, for $q = 2$, the constant C in (C.24) is, in this case, such that

$$C^2 \leqslant M^2||f||_\infty||L||_2^2\tilde{t}(K),$$

with

$$\tilde{t}(K) = \max\left\{\int_0^\infty r^{3/2}K''^2(r)dr, \int_0^\infty r^{1/2}K'^2(r)dr, \int_0^\infty r^{1/2}K''(r)K'(r)dr\right\}.$$

Note that $\tilde{t}(K) < \infty$ because of assumption **(A4)**.

In this case, inequality (C.25) in the proof of Lemma (C.24) is replaced by

$$\begin{aligned}
&\mathbb{P}\left\{\max_{2^{k-1}<n\leqslant 2^k}\sqrt{\frac{nh_n}{\kappa_n^{5/2}\log(h_n^{-1}\kappa_n^{1/2})}}\left|\left|\frac{\partial^2}{\partial\phi^2}\hat{f} - \mathbb{E}\left(\frac{\partial^2}{\partial\phi^2}\hat{f}\right)\right|\right|_\infty > \lambda\right\}\\
&\leqslant 9\mathbb{P}\left\{\sqrt{\frac{\kappa_{2^k}^{1/2}}{2^{k-1}h_{2^k}\lambda_{\kappa_{2^k}}^2(K)\log(h_{2^k}^{-1}\kappa_{2^k}^{1/2})}}\sup_{\substack{(x,\phi)\in\mathbb{R}\times\mathbb{T}\\ h_{2^k}\leqslant h\leqslant h_{2^{k-1}}\\ \kappa_{2^{k-1}}\leqslant\kappa\leqslant\kappa_{2^k}}}\left|\sum_{i=1}^{2^k}\left[L\left(\frac{X_i-x}{h}\right)\right.\right.\right.\\
&\times\left(\kappa K''\{\kappa[1-\cos(\Phi_i-\phi)]\}\sin^2(\Phi_i-\phi) + K'\{\kappa[1-\cos(\Phi_i-\phi)]\}\cos(\Phi_i-\phi)\right)\\
&-\mathbb{E}\left(L\left(\frac{X_i-x}{h}\right)\left(\kappa K''\{\kappa[1-\cos(\Phi_i-\phi)]\}\sin^2(\Phi_i-\phi)\right.\right.\\
&\left.\left.\left.\left.+K'\{\kappa[1-\cos(\Phi_i-\phi)]\}\cos(\Phi_i-\phi)\right)\right]\right| > \frac{\lambda}{30}\right\}.
\end{aligned}$$

In addition, by employing the same three changes of variables as before, it holds that

$$\begin{aligned}
&\mathbb{E}\left[\kappa^2L^2\left(\frac{X_1-x}{h}\right)K''^2\{\kappa[1-\cos(\Phi_1-\phi)]\}\sin^4(\Phi_1-\phi)\right]\\
&+\mathbb{E}\left[L^2\left(\frac{X_1-x}{h}\right)K'^2\{\kappa[1-\cos(\Phi_1-\phi)]\}\cos^2(\Phi_1-\phi)\right]\\
&+\mathbb{E}\left[2\kappa L^2\left(\frac{X_1-x}{h}\right)K'^2\{\kappa[1-\cos(\Phi_1-\phi)]\}\sin^2(\Phi_1-\phi)\cos(\Phi_1-\phi)\right]\\
&=\kappa^2\int_0^{2\pi}\int_{\mathbb{R}}L^2\left(\frac{z-x}{h}\right)K''^2\{\kappa[1-\cos(\alpha-\phi)]\}\sin^4(\alpha-\phi)f(z,\alpha)dzd\alpha\\
&+\int_0^{2\pi}\int_{\mathbb{R}}L^2\left(\frac{z-x}{h}\right)K'^2\{\kappa[1-\cos(\alpha-\phi)]\}\cos^2(\alpha-\phi)f(z,\alpha)dzd\alpha\\
&+2\kappa\int_0^{2\pi}\int_{\mathbb{R}}L^2\left(\frac{z-x}{h}\right)K''\{\kappa[1-\cos(\alpha-\phi)]\}K'\{\kappa[1-\cos(\alpha-\phi)]\}
\end{aligned}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$\times \sin^4(\alpha-\phi)\cos^2(\alpha-\phi)f(z,\alpha)dzd\alpha$$
$$\leqslant 2^{1/2}h\kappa^{-1/2}||L||_2^2||f||_\infty\left(2\int_0^\infty r^{3/2}K''^2(r)dr+\int_0^\infty r^{1/2}K'^2(r)dr\right.$$
$$\left.+2\int_0^\infty r^{1/2}K''(r)K'(r)dr\right)$$
$$\leqslant 2^{1/2}\cdot 5h\kappa^{-1/2}||L||_2^2||f||_\infty\tilde{t}(K).$$

Consequently, by taking

$$U_k=||L||_\infty||K''||_\infty \quad \text{and} \quad \sigma_k^2=2^{1/2}\cdot 5h_{2^k}\kappa_{2^k}^{-1/2}||L||_2^2||f||_\infty\tilde{t}(K)$$

and $\lambda=30C_2\sqrt{2^{1/2}\cdot 5||f||_\infty||L||_2^2\tilde{t}(K)}$, the rest of the proof is analogous to that of Lemma C.6. □

Lemma C.8. *Under assumptions* ***(A2)***, ***(A3-1)***, ***(A4)*** *and* ***(B2)***, *the uniform convergence rate of* $\frac{\partial^q}{\partial\phi^q}\hat{f}(\theta,\phi)$ *is given by*

$$\left|\left|\frac{\partial^q}{\partial\phi^q}\hat{f}-\frac{\partial^q}{\partial\phi^q}f\right|\right|_\infty:=\sup_{(\theta,\phi)}\left|\frac{\partial^q}{\partial\phi^q}\hat{f}(\theta,\phi)-\frac{\partial^q}{\partial y^q}f(\theta,\phi)\right|$$
$$=O(\nu^{-1}\kappa^{-1})+O_P\left(\sqrt{\frac{\nu^{1/2}\kappa^{(1+2q)/2}|\log(\nu^{-1/2}\kappa^{-1/2})|}{n}}\right),$$

as $\nu\to\infty$, $\kappa\to\infty$, $n/\left[\nu^{1/2}\kappa^{(1+2q)/2}|\log(\nu^{-1/2}\kappa^{-1/2})|\right]\to\infty$ *and* $|\log(\nu^{-1/2}\kappa^{-1/2})|/(\log\log n)\to\infty$, *for* $q=0,1,2$.

Proof. The arguments of the proof are analogous to those of Lemma C.6 and Lemma C.7 and, thus, the proof is omitted. □

Proofs of the main results

In this section, the proofs of the main results of Chapter 5 are given. With a slight abuse of notation, for the proofs in this section, K will denote the number of local modes of the conditional density for a given value of the predictor variable.

Proof of Proposition 5.1.

Proof. In order to obtain the proof, it is first necessary to note that the regression multifunction in (5.1) and its estimator in (5.4) are, respectively, equivalent to

$$\left\{y\in\mathbb{R}:\frac{\partial}{\partial y}f(\theta,y)=0,\frac{\partial^2}{\partial y^2}f(\theta,y)<0\right\}$$

and

$$\left\{ y \in \mathbb{R} : \frac{\partial}{\partial y} \hat{f}(\theta, y) = 0, \frac{\partial^2}{\partial y^2} \hat{f}(\theta, y) < 0 \right\}.$$

Hence, it is enough to work with the joint kernel density estimator and its partial derivatives (with respect to the response variable). Chen et al. (2016b) showed that, if assumption **(C1)** is satisfied, then when

$$\left\| \frac{\partial^q}{\partial y^q} \hat{f} - \frac{\partial^q}{\partial y^q} f \right\|_\infty := \sup_{(\theta,y)} \left| \frac{\partial^q}{\partial y^q} \hat{f}(\theta, y) - \frac{\partial^q}{\partial y^q} f(\theta, y) \right|$$

is small enough for $q = 0, 1, 2$, the number of local modes of $f(y|\theta)$ and $\hat{f}(y|\theta)$ is the same and each local mode is close to a unique estimated local mode, in the sense that there is a one-to-one correspondence between them.

Let $\{y_1, \ldots, y_K\} \subset M(\theta)$ be the set of local modes and $\{\hat{y}_1, \ldots, \hat{y}_{\hat{K}}\} \subset \widehat{M}(\theta)$ the set of estimated local modes. Under conditions **(A1)**, **(A3)**, **(A4-1)**, **(B1)** and **(C1)**, because of Lemma C.6 and Theorem 1 in Chen et al. (2016b), it can be assumed that for n sufficiently large, $K = \hat{K}$ and for each $\theta \in \mathbb{T}$, it holds $\widehat{M}(\theta) \subset M(\theta) \oplus \rho_1$. In addition, there is a unique estimated local mode $\hat{y}_k$ within $y_k \oplus \rho_1$.

The focus is now set to one local mode, y_k, and the rate of convergence of $|\hat{y}_k - y_k|$ will be derived, to subsequently generalize it to all the conditional local modes. By definition,

$$\frac{\partial}{\partial y} f(\theta, y_k) = 0 \quad \text{and} \quad \frac{\partial}{\partial y} \hat{f}(\theta, \hat{y}_k) = 0.$$

Since by assumption **(A1)**, $\frac{\partial^3}{\partial y^3} f(\theta, y)$ is bounded in absolute value, by applying a Taylor expansion it holds that

$$\frac{\partial}{\partial y} f(\theta, \hat{y}_k) - \underbrace{\frac{\partial}{\partial y} f(\theta, y_k)}_{=0} = \frac{\partial^2}{\partial y^2} f(\theta, y_k)(\hat{y}_k - y_k) + o[(\hat{y}_k - y_k)].$$

Under assumption **(C1)**, $\frac{\partial^2}{\partial y^2} f(\theta, y_k)$ is non-zero. Then,

$$|\hat{y}_k - y_k| = \left| \left[\frac{\partial^2}{\partial y^2} f(\theta, y_k) \right]^{-1} \right| \left| \frac{\partial}{\partial y} f(\theta, \hat{y}_k) \right| + o(|\hat{y}_k - y_k|).$$

In addition, under **(C1)**, when y is close to a local mode, $\frac{\partial^2}{\partial y^2} f(\theta, y)$ is bounded. Consequently, $|\hat{y}_k - y_k|$ shares the rate of $\left|\frac{\partial}{\partial y} f(\theta, \hat{y}_k)\right|$ and

$$\left| \frac{\partial}{\partial y} f(\theta, \hat{y}_k) \right| = \left| \frac{\partial}{\partial y} f(\theta, \hat{y}_k) - \underbrace{\frac{\partial}{\partial y} \hat{f}(\theta, \hat{y}_k)}_{=0} \right| = O(\kappa^{-1} + h^2) + O_P\left(\sqrt{\frac{\kappa^{1/2}}{nh^3}} \right), \tag{C.28}$$

where Lemma C.3 was applied. Now,

$$\Lambda(\theta) = \text{Haus}\left[\widehat{M}(\theta), M(\theta)\right] = \max_{k=1,\dots,K} |\hat{y}_k - y_k|,$$

and the rate in (C.28) holds for every mode. Then,

$$\Lambda(\theta) = O(\kappa^{-1} + h^2) + O_P\left(\sqrt{\frac{\kappa^{1/2}}{nh^3}}\right).$$

In addition, for the second assertation of Proposition 5.1, it holds

$$\mathbb{E}\left[\Lambda^2(\theta)\right] = O(\kappa^{-2} + h^4) + O\left(\frac{\kappa^{1/2}}{nh^3}\right)$$

and the integrated bias and variance yield the same rate of convergence (see Chacón et al., 2011; Chacón and Duong, 2013). □

Proof of Proposition 5.2.

Proof. The proofs follows the same steps as the previous one. First, note that $M(x)$ is equivalent to

$$\left\{\phi \in \mathbb{T} : \frac{\partial}{\partial\phi} f(x,\phi) = 0, \frac{\partial^2}{\partial\phi^2} f(x,\phi) < 0\right\}$$

and $\widehat{M}(x)$ is equivalent to

$$\left\{\phi \in \mathbb{T} : \frac{\partial}{\partial\phi} \hat{f}(x,\phi) = 0, \frac{\partial^2}{\partial\phi^2} \hat{f}(x,\phi) < 0\right\}.$$

Now, using Theorem 6 in Zhang and Chen (2021a), Lemma C.7 and assumptions **(A2)**, **(A3)**, **(A4)**, **(B2)** and **(C2)**, it can be stated that for a large enough n, the set of local modes $\{\phi_1, \dots, \phi_K\} \subset M(x)$ and the set of estimated local modes $\{\hat{\phi}_1, \dots, \hat{\phi}_{\hat{K}}\} \subset \widehat{M}(x)$ have the same cardinal, *i.e.*, $K = \hat{K}$. In addition, $\widehat{M}(x) \subset M(x) \oplus \rho_2$ and there is a unique estimated local mode $\hat{\phi}_k$ within $\phi_k \oplus \rho_2$.

Now, consider the local modes ϕ_k and $\hat{\phi}_k$. Trivially, it holds that

$$\frac{\partial}{\partial\phi} f(x,\phi_k) = 0 \quad \text{and} \quad \frac{\partial}{\partial\phi} \hat{f}(x,\hat{\phi}_k) = 0.$$

By **(A2)**, $\frac{\partial^3}{\partial\phi^3} f(x,\phi)$ is bounded in absolute value. Then, by employing a Taylor expansion with the sine approximation, it holds that

$$\frac{\partial}{\partial\phi} f(x,\hat{\phi}_k) - \underbrace{\frac{\partial}{\partial\phi} f(x,\phi_k)}_{=0} = \frac{\partial^2}{\partial\phi^2} f(x,\phi_k)\sin(\hat{\phi}_k - \phi_k) + o[\sin(\hat{\phi}_k - \phi_k)].$$

Under assumption **(C2)**, $\frac{\partial^2}{\partial\phi^2}f(x,\phi_k)$ is non-zero. Then,

$$|\sin(\hat{\phi}_k-\phi_k)| = \left|\left[\frac{\partial^2}{\partial\phi^2}f(x,\phi_k)\right]^{-1}\right| \left|\frac{\partial}{\partial\phi}f(x,\hat{\phi}_k)\right| + o(|\sin(\hat{\phi}_k-\phi_k)|).$$

In addition, under **(C2)**, when ϕ is close to a local mode, $\frac{\partial^2}{\partial\phi^2}f(x,\phi)$ is bounded and, since $\hat{\phi}_k \in \phi_k \oplus \rho_2$, the value of ρ_2 can be chosen small enough so that $1-\cos(\hat{\phi}_k-\phi_k) < 1$. Then,

$$1-\cos(\hat{\phi}_k-\phi_k) \leqslant |\sin(\hat{\phi}_k-\phi_k)| = \left|\left[\frac{\partial^2}{\partial\phi^2}f(x,\phi_k)\right]^{-1}\right| \left|\frac{\partial}{\partial\phi}f(x,\hat{\phi}_k)\right| + o(|\sin(\hat{\phi}_k-\phi_k)|).$$

Consequently, $1-\cos(\hat{\phi}_k-\phi_k)$ shares the rate of $\left|\frac{\partial}{\partial\phi}f(x,\hat{\phi}_k)\right|$ and

$$\left|\frac{\partial}{\partial\phi}f(x,\hat{\phi}_k)\right| = \left|\frac{\partial}{\partial\phi}f(x,\hat{\phi}_k) - \underbrace{\frac{\partial}{\partial\phi}\hat{f}(x,\hat{\phi}_k)}_{=0}\right| = O(h^2+\kappa^{-1}) + O_P\left(\sqrt{\frac{\kappa^{3/2}}{nh}}\right), \quad \text{(C.29)}$$

where Lemma C.4 was applied. Now,

$$\text{Haus}\left[\widehat{M}(x), M(x)\right] = \max_{k=1,\ldots,K}[1-\cos(\hat{\phi}_k-\phi_k)],$$

and the rate in (C.29) holds for every mode. Then,

$$\tilde{\Lambda}(x) = O(h^2+\kappa^{-1}) + O_P\left(\sqrt{\frac{\kappa^{3/2}}{nh}}\right).$$

For the second assertation,

$$\mathbb{E}\left[\tilde{\Lambda}(x)\right] = O(h^2+\kappa^{-1}) + O\left(\sqrt{\frac{\kappa^{3/2}}{nh}}\right)$$

and the integrated expectation yields the same rate of convergence. □

Proof of Proposition 5.3.

Proof. Under conditions **(A2)**, **(A3-1)**, **(A4)**, and **(B2)**, the pointwise and uniform consistency of the toroidal kernel density estimator and its first two partial derivatives with respect to the response are guaranteed because of Lemmas C.5 and C.8. The rest of the proof is analogous to the proofs of Proposition 5.1 and 5.2. □

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Appendix D

Software: update of the NPCirc package

D.1 The NPCirc package

As it has been discussed throughout this dissertation, circular data often arise in several fields. It seems of great importance, then, that the methodologies designed to analyze circular data are available to interested users. With this objective in mind, a variety of packages dealing with the exploration of circular data have been made available in R (R Core Team, 2021) in recent years.

The package NPCirc (Oliveira et al., 2014b), initially released in 2012, deals with the nonparametric estimation of curves involving circular data. Among other tools, it contains functions to perform circular kernel density estimation (including methods for selecting the concentration parameter) or to compute the nonparametric estimation of regression functions with the methods described in Section 2.1.

Given that the contents of this dissertation deal with nonparametric methods for regression involving circular data, functions which perform the methodologies presented in this work have been coded and included in NPCirc package. Version 3.0.1 was released in July, 2021 with the functions corresponding to Chapter 2. The new version 3.1.1 also contains the functions with the methods of Chapters 3, 4 and 5, as well as the data sets employed throughout the dissertation to illustrate the proposed methodologies.

In comparison with the latest version available previous to the inclusion of the methods proposed in this thesis (2.0.1), the version 3.1.1 of NPCirc contains eight new functions and six new data sets. The documentation relative to the new functions and data sets is provided in the next section.

D.2 Documentation for the new functions

The documentation corresponding to the NPCirc package is shown in this section, focusing only on the new updated functions which are related to this dissertation. Therefore, the new functionalities allow for significance tests in the context of mean regression with circular variables; equality and parallelism tests for circular ANCOVA models; local likelihood estimation and smoothing parameter selection in models with circular covariates; joint estimation of mean and dispersion in double Poisson models with circular covariates; and circular multimodal regression, including the selection of the smoothing parameters. In addition, six of

the data sets described in Appendix A were added to the package, and the corresponding documentation is also provided.

Package NPCirc

May 25, 2022

Title NPCirc

Version 3.1.1

Date 2022-05-25

Author Maria Alonso-Pena [aut, cre]

Maria Oliveira [aut, cre],

Jose Ameijeiras-Alonso [aut],

Rosa M. Crujeiras [aut],

Irene Gijbels [aut],

Alberto Rodriguez-Casal [aut],

Felicita Scapini [dtc]

Maintainer Maria Alonso-Pena <mariaalonso.pena@usc.es>

Description

Nonparametric smoothing methods for density and regression estimation involving circular data, including the estimation of the mean regression function and other conditional characteristics.

Depends R(⩾ 3.5.0), circular

License GPL (>= 2)

Imports Rcpp (>= 1.0.8.3), misc3d, movMF, plotrix, rgl, shape, Bolstad2

LinkingTo Rcpp, RcppArmadillo

URL https://www.jstatsoft.org/v61/i09/

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

NeedsCompilation yes

ExperimentalWindowsRuntime ucrt

Archs x64

R topics documented:

NPCirc-package 183
ancova.circ.lin 186
bw.circ.local.lik 189
bw.joint.dpcirc 191
bw.modalreg.circ.lin 193
circ.local.lik 195
flywheels 198
HumanMotorResonance 199
kern.dpreg.circ 200
modalreg.circ.lin 202
noeffect.circ.lin 205
pm10 208
sandhoppers 209
spikes 210
zebrafish 211

`NPCirc-package` *Nonparametric circular methods.*

Description

Nonparametric smoothing methods for density and regression estimation involving circular data, including methods for estimating circular densities and mean regression functions, described in Oliveira et al. (2014), testing proposals for circular mean regression described in Alonso-Pena et al. (2021), estimation of conditional characteristics (Alonso-Pena et al. 2022a,2022b) and estimation of conditional modes (Alonso-Pena and Crujeiras, 2022).

Details

Package:	NPCirc
Type:	Package
Version:	3.1.1
Date:	2022-05-25
License:	GPL-2
LazyLoad:	yes

This package incorporates the function `kern.den.circ` which computes the circular kernel density estimator. For choosing the smoothing parameter different functions are available: `bw.rt`, `bw.CV`, `bw.pi`, and `bw.boot`. For mean regression involving circular variables, the package includes the functions: `kern.reg.circ.lin` for a circular covariate and linear response; `kern.reg.circ.circ` for a circular covariate and a circular response; `kern.reg.lin.circ` for a linear covariate and a circular response. The three functions compute Nadaraya-Watson and Local-Linear smoothers. The functions `bw.reg.circ.lin`, `bw.reg.circ.circ` and `bw.reg.circ.lin` implement cross–validation rules for selecting the smoothing parameter. Functions `circsizer.density` and `circsizer.re- gression` provide CircSiZer maps for kernel density estimation and regression estimation, respectively. Functions `noeffect.circ.lin`, `noeffect.circ.circ` and `noeffect.lin.circ` compute the test of no effect to assess the significance of the predictor variable. Additionally, functions `ancova.circ.lin`, `ancova.circ.circ` and `ancova.lin.circ` implement hypothesis testing tools to assess the equality and parallelism of regression curves across different groups of observations.

Function `circ.local.lik` implements the estimation of different functions of interest (transformations of the mean function) in contexts where the predictor variable is circular and the conditional distribution is a Gaussian, Bernoulli, Poisson or gamma. Function `bw.circ.local.lik` computes smoothing parameters for the estimation in the previously described cases, allowing for three different rules. Function `kern.dpreg.circ` implements the joint kernel estimation of the mean and dispersion functions in cases where the predictor is circular and the conditional distribution is a double Poisson, a particular case of the double exponential family. Smoothing parameters in this context can be computed with function `bw.joint.dpcirc`, using a two-step cross-validation algorithm.

Functions `modalreg.circ.circ`, `modalreg.lin.circ` and `modalreg.circ.lin` implement the estimation of the modal regression multifunction (the conditional local modes) in the three circular regression scenarios. Smoothing parameters for these contexts can by computed by modal cross-validation employin functions `bw.modalreg.circ.circ`, `bw.modalreg.lin.circ` and `bw.modalreg.circ.lin`.

Functions `dcircmix` and `rcircmix` compute the density function and generate random samples of a circular distribution or a mixture of circular distributions, allowing for different components such as the circular uniform, von Mises, cardioid, wrapped Cauchy, wrapped normal and wrapped skew-normal. Finally, some data sets are provided. Missing data are allowed. Registries with missing data are simply removed.

For a complete list of functions, use `library(help="NPCirc")`.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Acknowledgements

This work has been supported by Project MTM2008-03010 from the Spanish Ministry of Science; Project and MTM201676969-P from the AEI co-funded by the European Re-

gional Development Fund (ERDF), the Competitive Reference Groups 2017-2020 (ED431 C 2017/38) from the Xunta de Galicia through the ERDF; and Innovation IAP network (Developing crucial Statistical methods for Understanding major complex Dynamic Systems in natural, biomedical and social sciences (StUDyS)) from Belgian Science Policy. Work of Maria Alonso-Pena was supported by grant ED481A-2019/139 from Xunta de Galicia. Work of Jose Ameijeiras-Alonso was supported by the FWO research project G.0826.15N (Flemish Science Foundation); and GOA/12/014 project (Research Fund KU Leuven). The authors want to acknowledge Prof. Arthur Pewsey for facilitating data examples and for his comments.

Author(s)

Maria Oliveira, Maria Alonso-Pena, Jose Ameijeiras-Alonso, Rosa M. Crujeiras, Alberto Rodriguez–Casal and Irene Gijbels.

Maintainer: Maria Alonso-Pena <mariaalonso.pena@usc.es>

References

Oliveira, M., Crujeiras, R.M. and Rodriguez–Casal, A. (2012) A plug–in rule for bandwidth selection in circular density. *Computational Statistics and Data Analysis*, **56**, 3898–3908.

Oliveira, M., Crujeiras R.M. and Rodriguez–Casal, A. (2013) Nonparametric circular methods for exploring environmental data. *Environmental and Ecological Statistics*, **20**, 1–17.

Oliveira, M., Crujeiras, R.M. and Rodriguez–Casal (2014) CircSiZer: an exploratory tool for circular data. *Environmental and Ecological Statistics*, **21**, 143–159.

Oliveira, M., Crujeiras R.M. and Rodriguez–Casal, A. (2014) NPCirc: an R package for nonparametric circular methods. *Journal of Statistical Software*, **61**(9), 1–26. https://www.jstatsoft.org/v61/i09/

Alonso-Pena, M., Ameijeiras-Alonso, J. and Crujeiras, R.M. (2021) Nonparametric tests for circular regression. *Journal of Statistical Computation and Simulation*, **91**(3), 477–500.

Alonso-Pena, M., Gijbels, I. and Crujeiras, R.M. (2022). A general framework for circular local likelihood regression. *Under review*.

Alonso-Pena, M., Gijbels, I. and Crujeiras, R.M. (2022). Flexible joint modeling of mean and dispersion for the directional tuning of neuronal spike counts. *Under review*.

Alonso-Pena, M. and Crujeiras, R. M. (2022). Analizing animal escape data with circular nonparametric multimodal regression. Annals of Applied Statistics. (To appear).

`ancova.circ.lin` *Nonparametric analysis of covariance tests for circular regression*

Description

Function `ancova.circ.lin` computes nonparametric ANCOVA tests to compare regression curves with a circular predictor variable and a real-valued response variable. The null hypothesis may be either equality or parallelism of the regression curves, as described in Alonso-Pena et al. (2021). It uses the nonparametric Nadaraya-Watson estimator or the Local-Linear estimator for circular-linear data described in Di Marzio et al. (2009) and Oliveira et al. (2013).

Function `ancova.lin.circ` computes nonparametric ANCOVA tests to compare regression curves with a real-valued predictor variable and a circular response variable. The null hypothesis may be either equality or parallelism of the regression curves, as described in Alonso-Pena et al. (2021). It uses the nonparametric Nadaraya-Watson estimator or the Local-Linear estimator for linear-circular data described in Di Marzio et al. (2012).

Function `ancova.circ.circ` computes nonparametric ANCOVA tests to compare regression curves with a circular predictor variable and a circular response variable. The null hypothesis may be either equality or parallelism of the regression curves, as described in Alonso-Pena et al. (2021). It uses the nonparametric Nadaraya-Watson estimator or the Local-Linear estimator for circular-circular data described in Di Marzio et al. (2012).

Usage

```
ancova.circ.lin(x, y, g, bw, bw1, test = "eq", method = "LL",calib = "chisq",
n_boot = 500)
ancova.lin.circ(x, y, g, bw, bw1, test = "eq", method = "LL", n_boot = 500)
ancova.circ.circ(x, y, g, bw, bw1, test = "eq", method = "LL", n_boot = 500)
```

Arguments

`x`	Vector of data for the independent variable. The object is coerced to class `circular` when using functions `ancova.circ.lin` and `ancova.circ.circ`.
`y`	Vector of data for the dependent variable. This must be same length as `x`. The object is coerced to class `circular` when using functions `ancova.lin.circ` and `ancova.circ.circ`.
`g`	Vector of group indicators.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

bw	Smoothing parameter to be used. If not provided it selects the parameter obtained by cross-validation.
bw1	Preliminary smoothing parameter for the parallelism test.
test	Character string giving the type of test to be performed. Must be one of "eq" for the test of equality or "paral" for the test of parallelism. Default is test="eq".
method	Character string giving the estimator to be used. This must be one of "LL" for Local-Linear estimator or "NW" for Nadaraya-Watson estimator. Default method="LL".
calib	Character string giving the calibration method to be used in ancova.circ.lin function. This must be one of "chisq" for the chi-squared approximation or "boot" for the bootstrap calibration.
n_boot	Number of bootstrap resamples. Default is n_boot=500. In function ancova.circ.lin, only if calib="boot".

Details

See Alonso-Pena et al. (2021). The NAs will be automatically removed.

Value

A list with class "htest" containing the following components:

statistic	observed value of the statistic.
bw	Smoothing parameter used.
p.value	p-value for the test.
data.name	a character string giving the name(s) of the data.
alternative	a character string describing the alternative hypothesis.

Author(s)

Maria Alonso-Pena, Jose Ameijeiras-Alonso and Rosa M. Crujeiras

References

USC DE SANTIAGO DE COMPOSTELA

Alonso-Pena, M., Ameijeiras-Alonso, J. and Crujeiras, R.M. (2021) Nonparametric tests for circular regression. *Journal of Statistical Computation and Simulation*, **91**(3), 477–500.

Di Marzio, M., Panzera A. and Taylor, C. C. (2009) Local polynomial regression for circular predictors. *Statistics and Probability Letters*, **79**, 2066–2075.

Di Marzio, M., Panzera A. and Taylor, C. C. (2012) Non–parametric regression for circular responses. *Scandinavian Journal of Statistics*, **40**, 228–255.

Oliveira, M., Crujeiras R.M. and Rodriguez-Casal, A. (2013) Nonparametric circular methods for exploring environmental data. *Environmental and Ecological Statistics*, **20**, 1–17.

See Also

kern.reg.circ.lin, kern.reg.lin.circ, kern.reg.circ.circ

Examples

```
# ANCOVA circ-lin
set.seed(2025)
x1 <- rcircularuniform(100)
x2 <- rcircularuniform(100)
x <- c(x1, x2)
y1 <- 2*sin(as.numeric(x1)) + rnorm(100, sd=2)
y2 <- 4 + 2*sin(as.numeric(x2)) + rnorm(100, sd=2)
y <- c(y1, y2)
g <- c(rep(0,100), rep(1,100))
ancova.circ.lin(x, y, g,  test = "eq")
ancova.circ.lin(x, y, g,  test = "paral")


# ANCOVA lin-circ
set.seed(2025)
x1 <- runif(100)
x2 <- runif(100)
y1 <- 3*pi*x1^2 + rvonmises(100, mu = 0, kappa = 6)
y2 <- 2*pi/8 + 3*pi*x2^2 + rvonmises(100, mu = 0, kappa = 6)
x <- c(x1, x2)
y <- c(y1, y2)
g<-c(rep(0, 100), rep(1, 100))
ancova.lin.circ(x, y, g,  test = "eq")
ancova.lin.circ(x, y, g,  test = "paral")


# ANCOVA circ-circ
set.seed(2025)
x1 <- rcircularuniform(100)
x2 <- rcircularuniform(100)
y1 <- 2*sin(2*x1) + rvonmises(100, mu = 0, kappa = 8 )
```

```
y2 <- pi/8 + 2*sin(2*x2) + rvonmises(100, mu = 0, kappa = 8 )
x <- c(x1, x2)
y <- c(y1, y2)
g<-c(rep(0, 100), rep(1, 100))
ancova.circ.circ(x, y, g,  test = "eq")
ancova.circ.circ(x, y, g,  test = "paral")
```

`bw.circ.local.lik` *Smoothing parameter selection for circular local likelihood regression*

Description

Function `bw.circ.local.lik` computes values of the smoothing (concentration) parameter for local likelihood regression when the predictor is circular and the conditional density is either gaussian, Bernoulli, Poisson or gamma. The smoothing parameter can be selected by the refined rule, the CRSC rule or cross-validation, as described in Alonso-Pena et al. (2022).

Usage

```
bw.circ.local.lik(x, y, t, rule = NULL, p, family, startv = NULL,
lower = 0, upper = 50, lower_ast = 0, upper_ast = 15, tol = 0.00001,
maxit = 300, from = circular(0), to = circular(2 * pi), len = 250)
```

Arguments

`x`	Vector of data for the independent variable. The object is coerced to class `circular`.
`y`	Vector of data for the dependent variable. This must be same length as `x`.
`t`	Points where the regression function is estimated. If `NULL`, equally spaced points are used according to the parameters `from`, `to` and `len`.
`rule`	Character string giving the rule to be used to select the smoothing (concentration) parameter. This must be one of `"refined"` (only for `p = 1`), `"CRSC"` or `"cv"`.
`p`	Degree of the local sine-polynomial to be used in the estimation process. It must be 1 or 3.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

`family` Character string indicating the conditional density to be used. It must be one of `"gaussian"`, `"bernoulli"`, `"poisson"` or `"gamma"`.

`startv` Vector containing the initial values for the estimation algorithm if family is set as `"bernoulli"`, `"poisson"` or `"gamma"`. The vector must be of length 2 if `p = 1` and length 4 if `p = 3`. If `NULL`, the initial parameters are the ones corresponding to the global mean of the responses.

`lower, upper` Lower and upper boundary of the interval to be used in the search for the value of the smoothing parameter. Default `lower = 0` and `upper = 50`.

`lower_ast, upper_ast` Lower and upper boundary of the interval to be used in the search for the value of the pilot smoothing parameter in the refined rule. Default `lower_ast = 0` and `upper_ast = 15`.

`tol` Tolerance parameter for convergence in the numerical estimation. Only needed if family is one of `"bernoulli"`, `"poisson"` or `"gamma"`. Default is `tol = 0.00001`.

`maxit` Maximum number of iterations in the numerical estimation. Only needed if family is one of `"bernoulli"`, `"poisson"` or `"gamma"`. Default is `maxit = 300`.

`from, to` Left and right-most points of the grid at which the density is to be estimated. The objects are coerced to class `circular`.

`len` Number of equally spaced points at which the density is to be estimated.

Details

For the `refined` rule, which is only available for `p = 1`, first a pilot concentration parameter is selected with the (E)CRSC rule using a sine-polynomial of degree 3, where the search is conducted between the values `lower_ast` and `upper_ast`. With the pilot smoothing parameter, the MISE of the estimator is approximated and the refined rule selects the parameter which minimizes the approximated MISE.

The `CRSC` rule selects the parameter minimizing the Circular Residual Squares Criterion if `family = "gaussian"` and the Extended Circular Residual Squares Criterion in the other cases.

The `cv` rule performs a cross-validation search.

See Alonso-Pena et al. (2022) for more details.

Value

Value of the smoothing parameter.

Author(s)

Maria Alonso-Pena, Irene Gijbels and Rosa M. Crujeiras.

References

Alonso-Pena, M., Gijbels, I. and Crujeiras, R.M. (2022). A general framework for circular local likelihood regression. *Under review*.

See Also

`circ.local.lik`

Examples

```
data(spikes)
direction<-circular(spikes$direction,units="degrees")
counts<-spikes$counts
bw.circ.local.lik(direction, counts, rule = "refined", p=1, family="poisson")
```

---

bw.joint.dpcirc *Smoothing parameter selection for circular double Poisson regression*

Description

Function `bw.joint.dpcirc` provides the smoothing parameters for the nonparametric joint estimator of the mean and dispersion functions when the conditional density is a double Poisson. It performs a joint cross-validation search.

Usage

```
bw.joint.dpcirc(x, y, startvmu = NULL, startvgam = NULL,
lower=c(0.05,0.05), upper=c(50,7),tol = 0.00001, maxit = 300)
```

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Arguments

x	Vector of data for the independent variable. The object is coerced to class circular.
y	Vector of data for the dependent variable. This must be same length as x and should contain counts.

`startvmu`	Vector of length two containing the initial values for the parameters corresponding to the estimation of the mean.
`startvgam`	Vector of length two containing the initial values for the parameters corresponding to the estimation of the dispersion.
`lower, upper`	Vectors of length two with the `lower` and `upper` boundaries of the intervals to be used in the search for the values of the smoothing parameters. The first component of each corresponds to the parameter associated to the estimation of the mean, while the second component corresponds to the estimation of the dispersion.
`tol`	Tolerance parameter for convergence in the numerical estimation.
`maxit`	Maximum number of iterations in the numerical estimation.

Details

See Alonso-Pena et al. (2022) for details.

The NAs will be automatically removed.

Value

A vector of length two with the first component being the value of the smoothing parameter associated to the mean estimation and with the second component being the value of the smoothing parameter associated to the dispersion estimation.

Author(s)

Maria Alonso-Pena, Irene Gijbels and Rosa M. Crujeiras

References

Alonso-Pena, M., Gijbels, I. and Crujeiras, R.M. (2022). Flexible joint modeling of mean and dispersion for the directional tuning of neuronal spike counts. *Under review*.

See Also

`kern.dpreg.circ`

Examples

```
data(spikes)
direction<-circular(spikes$direction,units="degrees")
counts<-spikes$counts
bw.joint.dpcirc(direction, counts, lower=c(0.5,0.5), upper=c(50,7))
```

bw.modalreg.circ.lin *Smoothing parameter selection for circular multimdodal regression*

Description

Function bw.modalreg.circ.lin provides the modal cross-validation smoothing parameters for the multimodal regression estimator when the covariate is circular and the response variable is linear.

Function bw.modalreg.circ.circ provides the modal cross-validation smoothing parameters for the multimodal regression estimator when the covariate and the response variable are circular.

Function bw.modalreg.lin.circ provides the modal cross-validation smoothing parameters for the multimodal regression estimator when the covariate is linear and the response variable is circular.

Usage

```
bw.modalreg.circ.lin(x, y, lower = NULL, upper = NULL, maxit = 500,
tol = 0.00001)
bw.modalreg.circ.circ(x, y, lower = NULL, upper = NULL, maxit = 500,
tol = 0.00001)
bw.modalreg.lin.circ(x, y, lower = NULL, upper = NULL, maxit = 500,
tol = 0.00001)
```

Arguments

x	Vector of data for the independent variable. The object is coerced to class circular when using functions bw.modalreg.circ.lin and bw.modalreg.circ.circ.
y	Vector of data for the dependent variable. This must be same length as x. The object is coerced to class circular when using functions bw.modalreg.circ.circ and bw.modalreg.lin.circ.
lower, upper	Vectors of length two with the lower and upper boundaries of the intervals to be used in the search for the values of the smoothing parameters.
maxit	Maximum number of iterations in the estimation through the conditional (circular) mean shift.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

`tol` Tolerance parameter for convergence in the estimation through the conditional (circular) mean shift.

Details

See Alonso-Pena and Crujeiras (2022) for details.

The NAs will be automatically removed.

Value

A vector of length two with the first component being the value of the smoothing parameter associated to the predictor variable and with the second component being the value of the smoothing parameter associated to the response variable.

Author(s)

Maria Alonso-Pena and Rosa M. Crujeiras.

References

Alonso-Pena, M. and Crujeiras, R. M. (2022). Analizing animal escape data with circular nonparametric multimodal regression. Annals of Applied Statistics. (To appear).

See Also

`modalreg.circ.lin`, `modalreg.circ.circ`, `modalreg.lin.circ`

Examples

```
# Circ-lin
set.seed(8833)
n1<-100
n2<-100
gamma<-8
sigma<-1.5
theta1<-rcircularuniform(n1)
theta2<-rcircularuniform(n2)
theta<-c(theta1,theta2)
y1<-2*sin(2*theta1)+rnorm(n1,sd=sigma)
y2<-gamma+2*sin(2*theta2)+rnorm(n2,sd=sigma)
y<-as.numeric(c(y1,y2))
bw<-bw.modalreg.circ.lin(theta, y)

# Lin-circ
```

```
n1<-100
n2<-100
con<-8
set.seed(8833)
x1<-runif(n1)
x2<-runif(n2)
phi1<-(6*atan(2.5*x1-3)+rvonmises(n1,m=0,k=con))
phi2<-(pi+6*atan(2.5*x2-3)+rvonmises(n2,m=0,k=con))
x<-c(x1,x2)
phi<-c(phi1,phi2)
bw<-bw.modalreg.lin.circ(x, phi)

# Circ-circ
n1<-100
n2<-100
con<-10
set.seed(8833)
theta1<-rcircularuniform(n1)
theta2<-rcircularuniform(n2)
phi1<-(2*cos(theta1)+rvonmises(n1,m=0,k=con))
phi2<-(3*pi/4+2*cos(theta2)+rvonmises(n2,m=0,k=con))
theta=c(theta1,theta2)
phi=c(phi1,phi2)
bw<-bw.modalreg.lin.circ(theta, phi)
```

`circ.local.lik` *Local likelihood estimation for regression with circular covariate*

Description

Function `circ.local.lik` computes a nonparametric estimation of a curve of interest, regarded as a transformation of the mean regression function, when the predictor is circular and the conditional density is either gaussian, Bernoulli, Poisson or gamma. It also computes the derivatives of the function of interest. It uses the method described in Alonso-Pena et al. (2022).

Usage

```
circ.local.lik(x, y, t = NULL, bw = NULL, family, p = 1,
```

```
startv = NULL, tol = 0.00001, maxit = 300, from = circular(0),
to = circular(2 * pi),len = 250)
```

Arguments

x	Vector of data for the independent variable. The object is coerced to class circular.
y	Vector of data for the dependent variable. This must be same length as x.
t	Points where the regression function is estimated. If NULL, equally spaced points are used according to the parameters from, to and len.
bw	Value of the smoothing (concentration) parameter used. The value of the smoothing parameter can be chosen by using bw.circ.local.lik.
family	Character string indicating the conditional density to be used. It must be one of "gaussian", "bernoulli", "poisson" or "gamma". When family = "gaussian", the conditional mean is estimated; when family = "bernoulli", the logit function is estimated and when family = "poisson" or family = "gamma", the function of interest is the logarithm of the conditional mean.
p	Degree of the local sine-polynomial to be used in the estimation process. It must be 1 or 3.
startv	Vector containing the initial values for the estimation algorithm if family is set as "bernoulli", "poisson" or "gamma". The vector must be of length 2 if p = 1 and length 4 if p = 3. If NULL, the initial parameters are the ones corresponding to the global mean of the responses.
tol	Tolerance parameter for convergence in the numerical estimation. Only needed if family is one of "bernoulli", "poisson" or "gamma". Default is tol = 0.00001.
maxit	Maximum number of iterations in the numerical estimation. Only needed if family is one of "bernoulli", "poisson" or "gamma". Default is maxit = 300.
from, to	Left and right-most points of the grid at which the density is to be estimated. The objects are coerced to class circular.
len	Number of equally spaced points at which the density is to be estimated.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Details

See Alonso-Pena et al. (2022).

The NAs will be automatically removed.

Value

A list containing the following components:

`datax, datay`	Original dataset.
`x`	The n coordinates of the points where the regression function and its derivatives are estimated.
`y`	A list containing the estimated values of the function of interest and its derivatives up to order p.
`bw`	The smoothing parameter used.
`n`	The sample size after elimination of missing values.
`call`	The call which produced the result.
`data.name`	The deparsed name of the x argument.
`has.na`	Logical, for compatibility (always FALSE).

Author(s)

Maria Alonso-Pena, Irene Gijbels and Rosa M. Crujeiras.

References

Alonso-Pena, M., Gijbels, I. and Crujeiras, R.M. (2022). A general framework for circular local likelihood regression. *Under review*.

See Also

`bw.circ.local.lik`

Examples

```
data(spikes)
direction<-circular(spikes$direction,units="degrees")
counts<-spikes$counts
circ.local.lik(direction, counts, bw=8, p=1, family="poisson")
```

`flywheels` *Flywheel measurements*

Description

The data consists on measurements on mechanical flywheels, which are tools to regulate an engine's rotation in the automotive industry. The dataset contains 60 observations, containing the flywheels' angle of imbalance and the magnitude required to correct them. Four different types of metals were employed in the production process.

Usage

```
data("flywheels")
```

Format

A data frame with 60 observations on the following 3 variables.

`angle` A numeric vector containing the angles of imbalance (in radians)

`weight` A numeric vector containing the magnitudes of correction (in inch-ounces).

`group` A vector with numbers from 1 to 4 indicating the type of metal.

Details

The data were analyzed in Anderson-Cook (1999) and Alonso-Pena et al. (2021).

Source

The dataset was obtained from Anderson-Cook (1999).

References

Anderson-Cook, C.M. (1999). A tutorial on one-way analysis of circular-linear data. *Journal of Quality Technology*, **31**(1), 109–119.

Alonso-Pena, M., Ameijeiras-Alonso, J. and Crujeiras, R.M. (2021) Nonparametric tests for circular regression. *Journal of Statistical Computation and Simulation*, **91**(3), 477–500.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Examples

```
data(flywheels)
```

`HumanMotorResonance` *Human motor resonance data*

Description

This dataset, given by Puglisi et al. (2017), contains measurements of motor resonance responses in humans, who were requested to observe a movement of a rythmic hand flexion-extension. The dataset is composed of 70 observations, including the angular position of the hand and the resonance response.

Usage

```
data("HumanMotorResonance")
```

Format

A data frame with 70 observations on the following 2 variables.

`Angular.position` A numeric vector containing the angular position of the hand (in radians).

`Reflex.Amplitude` A numeric vector containing the resonance response.

Details

See Puglisi et al. (2017) for details.

Source

The dataset was obtained by Puglisi et al. (2017).

References

Puglisi, G., Leonetti, A., Landau, A., Fornia, L., Cerri, G. and Borroni, P. (2017). The role of attention in human motor resonance. *PLOS ONE*, **12**(5), e0177457.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Examples

```
data(HumanMotorResonance)
```

`kern.dpreg.circ` *Joint kernel estimation of mean and dispersion functions in circular double Poisson models*

Description

Function `kern.dpreg.circ` implements the nonparametric joint estimator of the mean and dispersion functions when the covariate is circular and the conditional distribution is a double Poisson, a particular case of the double exponential family. It takes the von Mises distribution as the kernel employed for both the estimation of the mean and the dispersion. The employed estimator is a local-linear type.

Usage

```
kern.dpreg.circ(x, y, t = NULL, bw, startvmu = NULL, startvgam = NULL,
tol= 0.000001, maxit = 300, from =  circular(0),to = circular(2 * pi),
len = 250)
```

Arguments

`x`	Vector of data for the independent variable. The object is coerced to class `circular`.
`y`	Vector of data for the dependent variable. This must be same length as `x` and should contain counts.
`t`	Points where the regression function is estimated. If `NULL`, equally spaced points are used according to the parameters `from`, `to` and `len`.
`bw`	Numeric vector of length two containing the smoothing (concentration) parameters used. The first component is the parameter used for the estimation of the mean, while the second component is used for estimating the dispersion. The value of the smoothing parameters can be chosen by using `bw.joint.dpcirc`.
`startvmu`	Vector of length two containing the initial values for the parameters corresponding to the estimation of the mean.
`startvgam`	Vector of length two containing the initial values for the parameters corresponding to the estimation of the dispersion.
`tol`	Tolerance parameter for convergence in the numerical estimation.
`maxit`	Maximum number of iterations in the numerical estimation.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

`from, to`	Left and right-most points of the grid at which the density is to be estimated. The objects are coerced to class `circular`.
`len`	Number of equally spaced points at which the density is to be estimated.

Details

See Alonso-Pena et al. (2022) for details.

The NAs will be automatically removed.

Value

A list containing the following components:

`datax, datay`	Original dataset.
`x`	The n coordinates of the points where the regression function and its derivatives are estimated.
`estim`	A list containing the estimated values of the logarithm of the mean function (first component) and the logarithm of the dispersion function (second component).
`bw`	The vector of smoothing parameters used.
`n`	The sample size after elimination of missing values.
`call`	The call which produced the result.
`data.name`	The deparsed name of the x argument.
`has.na`	Logical, for compatibility (always FALSE).

Author(s)

Maria Alonso-Pena, Irene Gijbels and Rosa M. Crujeiras

References

Alonso-Pena, M., Gijbels, I. and Crujeiras, R.M. (2022). Flexible joint modeling of mean and dispersion for the directional tuning of neuronal spike counts. *Under review*.

See Also

`bw.joint.dpcirc`

Examples

```
data(spikes)
direction<-circular(spikes$direction,units="degrees")
counts<-spikes$counts
kern.dpreg.circ(direction, counts, bw=c(8,5))
```

`modalreg.circ.lin` *Circular multimodal regression estimation*

Description

Function `modalreg.circ.lin` implements the nonparametric multimodal regression estimator for a circular covariate and a real-valued response, as described in Alonso-Pena and Crujeiras (2022). It takes the von Mises distribution as the kernel associated to the predictor variable and the normal distribution as the kernel associated to the response variable.

Function `modalreg.circ.circ` implements the nonparametric multimodal regression estimator for a circular covariate and a circular response, as described in Alonso-Pena and Crujeiras (2022). It takes the von Mises distribution as the kernel associated to the predictor variable and the response variables.

Function `modalreg.lin.circ` implements the nonparametric multimodal regression estimator for a real-valued covariate and a circular response, as described in Alonso-Pena and Crujeiras (2022). It takes the normal distribution as the kernel associated to the predictor variable and the von Mises distribution as the kernel associated to the response variable.

Usage

```
modalreg.circ.lin(x, y, t=NULL, bw=NULL, tol = 0.0001, maxit = 500,
from = circular(0),to = circular(2 * pi), len = 300)

modalreg.circ.circ(x, y, t=NULL, bw=NULL, tol = 0.00001,
maxit = 500, from = circular(0), to = circular(2 * pi), len = 300)

modalreg.lin.circ(x, y, t=NULL, bw=NULL, tol = 0.0001, maxit = 500,
len=300)
```

USC UNIVERSIDADE DE COMPOSTELA

Arguments

x	Vector of data for the independent variable. The object is coerced to class circular when using functions modalreg.circ.lin and modalreg.circ.circ.
y	Vector of data for the dependent variable. This must be same length as x. The object is coerced to class circular when using functions modalreg.circ.circ and modalreg.lin.circ.
t	Points where the regression function is estimated. If NULL, equally spaced points are used according to the parameters from, to and len.
bw	Vector of length two with the values of the smoothing parameters to be used. The first component corresponds to the smoothing parameter associated to the predictor variable and the second component is the parameter associated to the response variable. If NULL, the parameters are selected via modal cross-validation.
tol	Tolerance parameter for convergence in the estimation through the conditional (circular) mean shift.
maxit	Maximum number of iterations in the estimation through the conditional (circular) mean shift.
from, to	Left and right-most points of the grid at which the regression multifunction is to be estimated. The objects are coerced to class circular.
len	Number of equally spaced points at which the regression multifunction is to be estimated.

Details

See Alonso-Pena and Crujeiras (2022) for details.

The NAs will be automatically removed.

Value

A list containing the following components:

datax, datay	Original dataset.
x	The n coordinates of the points where the regression multifunction is estimated.
y	A list with dimension the length of the number of evaluation points containing the estimated values of the multidunfunction for each evaluation point.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

bw	A verctor of lenght two with the smoothing parameters used.
n	The sample size after elimination of missing values.
call	The call which produced the result.
data.name	The deparsed name of the x argument.
has.na	Logical, for compatibility (always FALSE).

Author(s)

Maria Alonso-Pena and Rosa M. Crujeiras.

References

Alonso-Pena, M. and Crujeiras, R. M. (2022). Analizing animal escape data with circular nonparametric multimodal regression. Annals of Applied Statistics. (To appear).

See Also

bw.modalreg.circ.lin, bw.modalreg.circ.circ, bw.modalreg.lin.circ

Examples

```
# Circ-lin
set.seed(8833)
n1<-100
n2<-100
gamma<-8
sigma<-1.5
theta1<-rcircularuniform(n1)
theta2<-rcircularuniform(n2)
theta<-c(theta1,theta2)
y1<-2*sin(2*theta1)+rnorm(n1,sd=sigma)
y2<-gamma+2*sin(2*theta2)+rnorm(n2,sd=sigma)
y<-as.numeric(c(y1,y2))
fit<-modalreg.circ.lin(theta,y,bw=c(10,1.3))

# Lin-circ
n1<-100
n2<-100
con<-8
set.seed(8833)
x1<-runif(n1)
x2<-runif(n2)
```

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

```
phi1<-(6*atan(2.5*x1-3)+rvonmises(n1,m=0,k=con))
phi2<-(pi+6*atan(2.5*x2-3)+rvonmises(n2,m=0,k=con))
x<-c(x1,x2)
phi<-c(phi1,phi2)
fit<-modalreg.lin.circ(x, phi, bw=c(0.1,2.5))

# Circ-circ
n1<-100
n2<-100
con<-10
set.seed(8833)
theta1<-rcircularuniform(n1)
theta2<-rcircularuniform(n2)
phi1<-(2*cos(theta1)+rvonmises(n1,m=0,k=con))
phi2<-(3*pi/4+2*cos(theta2)+rvonmises(n2,m=0,k=con))
theta=c(theta1,theta2)
phi=c(phi1,phi2)
fit<-modalreg.circ.circ(theta, phi, bw=c(30,3))
```

`noeffect.circ.lin` *No-effect test for regression with circular data*

Description

Function `noeffect.circ.lin` computes the no-effect test for a circular predictor variable and a real-valued response variable as described in Alonso-Pena et al. (2021). It uses the nonparametric Nadaraya-Watson estimator or the Local-Linear estimator for circular-linear data described in Di Marzio et al. (2009) and Oliveira et al. (2013).

Function `noeffect.lin.circ` computes the no-effect test for a real-valued predictor variable and a circular response variable as described in Alonso-Pena et al. (2021). It uses the nonparametric Nadaraya-Watson estimator or the Local-Linear estimator for linear-circular data described in Di Marzio et al. (2012).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Function `noeffect.circ.circ` computes the no-effect test for a circular predictor variable and a circular response variable as described in Alonso-Pena et al. (2021). It uses the nonparametric Nadaraya-Watson estimator or the Local-Linear estimator for circular-circular data described in Di Marzio et al. (2012).

Usage

```
noeffect.circ.lin(x, y, bw, method = "LL", calib = "chisq", n_boot = 500)
noeffect.lin.circ(x, y, bw, method = "LL", n_boot = 500)
noeffect.circ.circ(x, y, bw, method = "LL", n_boot = 500)
```

Arguments

x	Vector of data for the independent variable. The object is coerced to class circular when using functions noeffect.circ.lin and noeffect.circ.circ.
y	Vector of data for the dependent variable. This must be same length as x. The object is coerced to class circular when using functions noeffect.lin.circ and noeffect.circ.circ.
bw	Smoothing parameter to be used. If not provided, functions noeffect.circ.lin and noeffect.circ.circ select $4cv$ and function noeffect.lin.circ selects $cv/4$, where cv is the parameter selected by cross-validation.
method	Character string giving the estimator to be used. This must be one of "LL" for Local-Linear estimator or "NW" for Nadaraya-Watson estimator. Default method="LL".
calib	Character string giving the calibration method to be used in noeffect.circ.lin function. This must be one of "chisq" for the chi-squared approximation or "boot" for the bootstrap calibration.
n_boot	Number of bootstrap resamples. Default is n_boot=500. In function noeffect.circ.lin, only if calib="boot".

Details

See Alonso-Pena et al. (2021). The NAs will be automatically removed.

Value

A list with class "htest" containing the following components:

statistic	observed value of the statistic.
bw	Smoothing parameter used.
p.value	p-value for the test.
data.name	a character string giving the name(s) of the data.
alternative	a character string describing the alternative hypothesis.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Author(s)

Maria Alonso-Pena, Jose Ameijeiras-Alonso and Rosa M. Crujeiras

References

Alonso-Pena, M., Ameijeiras-Alonso, J. and Crujeiras, R.M. (2021) Nonparametric tests for circular regression. *Journal of Statistical Computation and Simulation*, **91**(3), 477–500.

Di Marzio, M., Panzera A. and Taylor, C. C. (2009) Local polynomial regression for circular predictors. *Statistics and Probability Letters*, **79**, 2066–2075.

Di Marzio, M., Panzera A. and Taylor, C. C. (2012) Non–parametric regression for circular responses. *Scandinavian Journal of Statistics*, **40**, 228–255.

Oliveira, M., Crujeiras R.M. and Rodriguez-Casal, A. (2013) Nonparametric circular methods for exploring environmental data. *Environmental and Ecological Statistics*, **20**, 1–17.

See Also

kern.reg.circ.lin, kern.reg.lin.circ, kern.reg.circ.circ

Examples

```
# No-effect circ-lin
set.seed(2025)
x <- rcircularuniform(200)
y <- 2*sin(as.numeric(x)) + rnorm(200, sd=2)
noeffect.circ.lin(x, y)

# No-effect lin-circ
set.seed(2025)
x <- runif(200)
y<- pi/8 + rvonmises(200, mu = 0, kappa = 0.75)
noeffect.lin.circ(x, y)


# No-effect circ-circ
set.seed(2025)
x <- rcircularuniform(200)
y <- atan2(sin(2*x),cos(2*x)) + rvonmises(200, mu = 0, kappa = 2)
noeffect.circ.circ(x, y)
```

`pm10` *Pm10 particles in Pontevedra, Spain*

Description

This dataset contains measurements on pm10 particles concentration in the city of Pontevedra, Spain, as well as wind speed and wind direction recordings. The measurements were taken in a meteorological station in Pontevedra, during the year 2019, every six hours.

Usage

```
data("pm10")
```

Format

A data frame with 1156 observations on the following 3 variables.

`pm10` A numeric vector containing the concentration of pm10 particles (in micrograms per cubic meter).

`direction` A numeric vector containing the wind direction in degrees between 0 and 360.

`speed` A numeric vector containing the wind speed measured in kilometers per hour.

Source

Meteogalicia https://www.meteogalicia.gal/.

Examples

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

```
data(pm10)
```

`sandhoppers` *Behavioral plasticity of Talitrus saltator and Talorchestia brito*

Description

Orientation measured under natural conditions and other variables of interest for analyzing the behavioral plasticity of two sympatric sandhoppers species, Talitrus saltator and Talorchestia brito. The experiment was carried out on the exposed nontidal sand of Zouara beach located in the Tunisian northwestern coast. More details can be found in Marchetti and Scapini (2003) or Scapini et al. (2002).

Usage

```
data("sandhoppers")
```

Format

A data frame with 1828 observations on the following 12 variables.

`angle` Numeric vector containing the orientation angles in radians between 0 and 2π.

`date` A factor where each level indicates the date when angles were measured.

`month` A factor with two levels indicating the month when angles were measured. Experiments were performed in two different periods, `April` and `October`, which were chosen for the abundance of the populations, as well as for their non-extreme and changing climatic conditions.

`time` A factor with levels `afternoon`, `morning` and `noon`.

`azim` A numeric vector indicating the sun azimuth. The sun position was confounded with the time of the day (morning: 100-150, noon: az=151-210 and afternoon: az=211-260 experiments).

`hour` A factor with hours when angles were measured.

`species` A factor with three levels (`brito`, `salt`, `ND`) indicating the specie (brito, saltator, not determined).

`sex` A factor with three levels (`F`, `M`, `J`) indicating the sex (female, male, J).

`temp` A numeric vector indicating the temperature (degrees centigrade).

`humid` A numeric vector indicating the air relative humidity (%).

`land` A factor with two levels (`no`, `yes`) indicating landscape view was either permitted or screened.

`trap` A numeric vector containing the traps identifier used for capturing the sandhoppers.

Details

See Marchetti and Scapini (2003) and Scapini et al. (2002).

Source

Authors thank Prof. Felicita Scapini for providing the sandhoppers data (collected under the support of the European Project ERB ICI8-CT98-0270).

References

Marchetti, G. M. and Scapini, F. (2003). Use of multiple regression models in the study of sandhopper orientation under natural conditions. *Estuarine, Coastal and Shelf Science*, **58**, 207-215.

Scapini, F., Aloia, A., Bouslama, M. F., Chelazzi, L., Colombini, I., ElGtari, M., Fallaci, M. and Marchetti, G. M. (2002). Multiple regression analysis of the sources of variation in orientation of two sympatric sandhoppers, Talitrus saltator and Talorchestia brito, from an exposed Mediterranean beach. *Behavioral Ecology and Sociobiology*, **51**(5), 403–414.

Examples

```
data(sandhoppers)
```

`spikes` *Neuronal spikes in a macaque monkey*

Description

This dataset, given by Kohn and Movshon (2003), contains observations from an experimental study on macaque monkeys, where a macaque received a visual stimulus. The dataset is composed of 68 observations, consisting on the angle of stimulus and the number of spikes produced in a MT/V5 neuron.

Usage

```
data("spikes")
```

Format

A data frame with 68 observations on the following 2 variables.

`direction` A numeric vector containing the direction of the visual stimulus (in degrees).

`counts` A numeric vector containing the number of spikes.

Details

See Kohn and Movshon (2003) for details. This dataset was also studied in Alonso-Pena et al (2022).

Source

The dataset was obtained by Kohn and Movshon (2003).

References

Kohn, A., and Movshon, J.A. (2003). Neuronal adaptation to visual motion in area MT of the macaque. *Neuron*, **39**(4), 681–691.

Alonso-Pena, M., Gijbels, I. and Crujeiras, R.M. (2022). Flexible joint modeling of mean and dispersion for the directional tuning of neuronal spike counts. *Under review*.

Examples

```
data(spikes)
```

`zebrafish` *Zebrafish*

Description

The data consists on measurements from an experimental study on larval zebrafish, which were startled by an imitating predator consisting on a robot disguised as an adult zebrafish. The dataset includes 502 observations corresponding to the escape directions of each fish and the angles in which the zebrafish perceive the threat.

Usage

```
data("zebrafish")
```

Format

A data frame with 502 observations on the following 2 variables.

`stimulus` A numeric vector containing the angles of stimulus (in radians)

`res_angle` A numeric vector containing the directions of escape (in radians)

Details

Analysis of the zebrafish data with circular regression methods can be seen in Alonso-Pena et al. (2022).

Source

The data were obtained from the Dryad Digital Repository

`doi:https://doi.org/10.5061/dryad.47mq9.`

and first analyzed in Nair et al. (2017).

References

Nair, A., Changsing, K., Stewart, W.J. and McHenry, M.J. (2017). Fish prey change strategy with the direction of a threat. *Proceedings of the Royal Society B: Biological Sciences*, **284**, 20170393.

Alonso-Pena, M. and Crujeiras, R. M. (2022). Analizing animal escape data with circular nonparametric multimodal regression. Annals of Applied Statistics. (To appear).

Examples

```
data(zebrafish)
```

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Resumo en galego

Unha das tarefas máis importantes no campo da estatística é o estudo da relación de dependencia entre dúas ou máis variables aleatorias. Esta dependencia pódese modelar de formas moi diversas, sendo un dos enfoques máis empregados o uso de modelos de regresión onde, no marco clásico, se adoita tratar de explicar a esperanza condicional dunha variable resposta Y mediante unha función da variable explicativa X. Xeralmente, asúmese que a función de regresión ten unha forma explícita, que depende duns parámetros que é necesario estimar e sobre os que é posible realizar inferencia. A forma paramétrica máis simple é unha liña recta, o que dá lugar ao modelo de regresión lineal.

Porén, na práctica, asumir que a relación entre dúas variables é lineal pode resultar moi restritivo e non reflectir fielmente a relación existente entre as variables. Aínda que é posible expresar a media condicional da variable resposta como unha función máis complexa, dependente dun vector de parámetros, a miúdo o uso de modelos demasiado complicados pode resultar en problemas de especificación ou de *sobreaxustamento*. É neste contexto onde entran en xogo os modelos de regresión non paramétricos (Wand e Jones, 1995; Fan e Gijbels, 1996), nos que non se asume ningunha forma específica para a media condicional, senón que esta se estima directamente a partir dos propios datos. Na maior parte da literatura relativa á regresión non paramétrica, enténdese a función de regresión como a media ou esperanza condicional, pero existen outro tipo de técnicas onde a función de interese non é necesariamente a media, senón outra característica condicional da variable resposta, dado o valor da variable explicativa.

Máis aló do interese pola xeración de coñecemento neste ámbito e a comprensión matemática dos procesos de relación que involucran variables de natureza circular, en moitos campos aplicados como a bioloxía, a psicoloxía ou a meteoroloxía, existen medidas de interese que son direccións ou ángulos, polo que o seu soporte non é o espazo euclídeo. Este tipo de datos deben ser tratados con técnicas estatísticas que teñan en conta a natureza periódica das variables, para o que é necesario definir unha dirección inicial e un sentido de rotación. Algunhas referencias básicas no contexto da estatística circular (ou a súa extensión a outras variedades direccionais) son Mardia (1972), Fisher (1993), Mardia e Jupp (2000), Jammalamadaka e SenGupta (2001) ou Ley e Verdebout (2017). Nestes volumes recóllense ferramentas inferenciais, maioritariamente paramétricas e centradas no estudo das distribucións de probabilidade.

As observacións circulares poden vir acompañadas doutras medicións, que poden ser

tanto datos escalares como circulares, e en moitos casos resulta necesario estudar a relación de dependencia entre estas variables. Aínda que existen algúns modelos de regresión paramétricos que contemplan a existencia de variables circulares, a literatura de modelos de regresión circulares non paramétricos é bastante escasa, en especial a relacionada coa estimación doutras características condicionais distintas da media. Polo tanto, o obxectivo principal desta tese é a proposta e estudo de modelos de regresión non paramétricos con variables circulares, considerando distintas características condicionais como funcións de interese.

No que segue, exponse un resumo dos capítulos que conforman esta tese de doutoramento, destacando as principais achegas e avances recollidas en cada un deles, así como un breve resumo dos apéndices da memoria.

Capítulo 1: Introdución. Neste capítulo preséntanse distintos conceptos que serán abordados ao longo da memoria e expóñense, de xeito detallado, os principais obxectivos da tese.

En primeiro lugar, proporciónase unha introdución ás técnicas tipo núcleo para a estimación da función de regresión no contexto escalar, considerando a estimación de diferentes características condicionais. A función de regresión clásica defínese nun punto dado x_0 como

$$m(x_0) = \mathbb{E}(Y|X = x_0),$$

é dicir, a esperanza da variable resposta condicionada ao valor da variable explicativa. A función $m(\cdot)$ pódese estimar de xeito non paramétrico con estimadores tipo núcleo, como poden ser o estimador de Nadaraya-Watson (Nadaraya, 1964; Watson, 1964) ou o estimador polinómico local (Wand e Jones, 1995; Fan e Gijbels, 1996) onde, para levar a cabo o proceso de estimación, é necesario seleccionar un parámetro fiestra, que determina a suavidade das curvas de regresión estimadas. Unha cuestión importante na estimación da función de regresión é a determinación da significación do efecto da variable explicativa no valor esperado da resposta. Con fin de responder a esta problemática, Bowman e Azzalini (1997) propuxeron un contraste de hipóteses baseado na utilización de estimadores tipo núcleo que permite avaliar a significación da covariable. O modelo de regresión baseado na modelización da media pódese estender a contextos máis complexos, onde a existencia dunha variable aleatoria categórica determina a división da mostra en distintos grupos de observacións. Nestes casos, pódense empregar os chamados modelos de análise da covarianza non paramétricos (Young e Bowman, 1995). Estes modelos permiten, ademais, levar a cabo contrastes de hipóteses para determinar a igualdade ou paralelismo das curvas de regresión entre os distintos grupos de observacións.

Aínda que os métodos de estimación tipo núcleo proporcionan unha forma flexible para modelar a función de regresión sen especificar unha forma concreta para esta, este enfoque resulta de utilidade cando a esperanza do erro asociado ao modelo de regresión é cero, suposición que non sempre se cumpre na práctica. Unha clase máis ampla de estimadores tipo núcleo, axeitados para variables resposta de tipo xenérico, foi introducida por Staniswalis (1989) e Fan et al. (1995), que estenderon a idea da verosimilitude local proposta por Tib-

shirani e Hastie (1987). Neste contexto, a distribución condicional das variables asúmese coñecida, e considérase como función de regresión unha característica condicional desta. O proceso de estimación consiste na maximización da log-verosimilitude local, na que as contribucións individuais de cada observación se ponderan empregando unha función tipo núcleo, dependente dun parámetro fiestra. Este método permite a estimación de características condicionais en contextos onde a variable resposta é tanto continua como discreta. Ademais, Fan et al. (1998) propuxeron unha metodoloxía para aproximar o nesgo e a varianza dos estimadores, así como un método xeral para seleccionar o parámetro de suavizado.

A estimación tipo núcleo de certas características condicionais tamén se pode levar a cabo sen especificar a distribución condicional. Por exemplo, a estimación da mediana ou dos cuantís condicionais foi abordada por Chaudhuri (1991a) e Chaudhuri (1991b). Menos coñecida é a estimación da moda condicional (Yao et al., 2012), que dá lugar a unha forma de estimación máis robusta que as anteriormente mencionadas. Tanto a regresión á mediana como a regresión modal resultan de utilidade cando a distribución condicional, aínda que descoñecida, é unimodal. Non obstante, se a distribución condicional exhibe múltiples modas, a regresión multimodal preséntase como unha alternativa axeitada. A idea da estimación das modas locais condicionais foi brevemente abordada por Scott (1992), e estudada en máis profundidade por Einbeck e Tutz (2006) e Chen et al. (2016a). Neste contexto, a función de interese non é realmente unha función, senón unha multifunción ou función multiavaliada. O proceso de estimación é indirecto, obtendo en primeiro lugar un estimador tipo núcleo da densidade condicional, para despois calcular as modas locais de dito estimador mediante unha versión condicional do algoritmo *mean shift*. Revisións desta temática poden consultarse en Chen (2018) e Alonso-Pena (2020).

A segunda parte do Capítulo 1 introduce as ferramentas clásicas no contexto da estatística circular, na que se parte de observacións con soporte a circunferencia unidade $\mathbb{T} = [0, 2\pi)$. Introdúcese o concepto de densidade dunha variable aleatoria circular, que é análogo ao dunha función de densidade usual coa condición extra da periodicidade da función. Esta frecuencia con período 2π fai que as medidas usuais de localización e dispersión dunha variable aleatoria circular teñan que ser redefinidas, así como os seus análogos mostrais. Introdúcense tamén as principais funcións de densidade circulares, nas que destaca a densidade de von Mises, dada por

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \mu)], \quad \theta \in \mathbb{T},$$

onde μ representa o parámetro de localización, κ é o parámetro de concentración e $I_0(\cdot)$ é a función de Bessel de clase un e orde cero, que actúa como constante de normalización. Aínda que hai funcións de densidade especificamente enfocadas ao modelado de variables circulares, existen métodos para construír densidades circulares a partir de funcións de densidade para variables aleatorias escalares como, por exemplo, as densidades enroladas ou proxectadas. Non obstante, ao igual que sucede no contexto da regresión, estas formas paramétricas adoitan ser demasiado restritivas na práctica, polo que, como ferramenta exploratoria e infe-

rencial, pódese recorrer ao estimador tipo núcleo da densidade circular, dado por

$$\hat{f}(\theta_0) = \frac{1}{n}\sum_{i=1}^{n} K_\kappa(\Theta_i - \theta_0),$$

onde $K_\kappa(\cdot)$ é unha función tipo núcleo circular con parámetro de concentración κ, que actúa como parámetro de suavizado.

Por último, preséntanse os escenarios de regresión con variables circulares, contexto onde se enmarca esta tese de doutoramento, e os principais modelos paramétricos. Pódense distinguir tres escenarios diferentes no marco da regresión circular: o caso cunha variable explicativa circular Θ e unha resposta escalar Y, que se coñece como regresión circular-lineal; a situación contraria, na que a variable explicativa X é escalar e a resposta, Φ, ten natureza circular, coñecida como regresión lineal-circular; e o escenario onde ambas variables, Θ e Φ, son circulares, *i.e.*, regresión circular-circular. No contexto da regresión circular-lineal, adóitase modelar a esperanza de Y condicionada ao valor de X mediante unha combinación lineal do seno e coseno de X, aínda que modelos máis flexibles, con máis termos trigonométricos, son plausibles. No referente á regresión lineal-circular, os modelos máis empregados son a proposta de Fisher e Lee (1992) e o modelo proxectado de Presnell et al. (1998). A literatura correspondente á regresión circular-circular é máis escasa, destacando o modelo presentado por Jammalamadaka e SenGupta (2001), onde o tratamento da variable circular realízase considerando marxinalmente a súas compoñentes seno e coseno.

Capítulo 2: Inferencia non paramétrica para regresión circular. Neste capítulo preséntanse, en primeiro lugar, os estimadores tipo núcleo da regresión en media para variables circulares. No contexto da regresión con explicativa circular e resposta escalar, o modelo de regresión en media vén dado por

$$Y = m(\Theta) + \varepsilon,$$

onde ε é un erro aleatorio de media cero. Di Marzio et al. (2009) propuxeron un estimador tipo núcleo de $m(\cdot)$ baseado nunha modificación da expansión de Taylor, na que a función identidade é aproximada pola función seno para valores cercanos ao cero. Por outra banda, cando a variable resposta é circular, tense o seguinte modelo de regresión:

$$\Phi = [m(\Delta) + \varepsilon] (\text{mod } 2\pi),$$

onde Δ é unha variable aleatoria que pode ser tanto circular como escalar, e ε é un erro aleatorio circular con dirección media cero. Neste escenario, Di Marzio et al. (2013) propuxeron un estimador de $m(\cdot)$ mediante a aplicación do operador atan2$(\cdot,\cdot)$ aos estimadores tipo núcleo da media condicional do seno e coseno da resposta.

Co obxectivo de comprobar o efecto das variables explicativas nos modelos anteriores, propuxéronse nesta tese contrastes do efecto da covariable nos tres escenarios de regresión. No contexto circular-lineal, o estatístico de contraste constrúese medindo a distancia cadrática

entre os estimadores do modelo baixo a hipótese nula e a alternativa, mentres que nos escenarios lineal-circular e circular-circular, a distancia cadrática substitúese pola distancia en coseno. Plantexáronse tamén modelos de análise da covarianza (ANCOVA) non paramétricos, nos que se considera unha variable explicativa categórica adicional, determinando I grupos de observacións. En cada un dos tres contextos de regresión circular, propuxéronse contrastes de hipóteses para determinar tanto a igualdade como o paralelismo das curvas de regresión entre os I grupos, seguindo as ideas de Young e Bowman (1995).

Para poder analizar o funcionamento das distintas propostas de contraste, realizouse un estudo de simulación no que se consideraron diferentes modelos, tamaños mostrais e niveis de significación. Nos test de non-efecto, explorouse a influencia do parámetro de suavizado, determinando que un parámetro axeitado para a estimación dá lugar a un incorrecto calibrado. Determinouse que este efecto vén dado polo nesgo do estimador non paramétrico baixo a hipótese alternativa, e que a infrasuavización do estimador da curva de regresión conleva un correcto calibrado baixo a hipótese nula e unha potencia axeitada. Por outra banda, o problema do nesgo desaparece nos contrastes relativos aos modelos ANCOVA cando o mesmo parámetro de suavizado é empregado baixo cada unha das hipóteses, cancelando así o nesgo dos estimadores. Concluiuse que cando o mesmo parámetro de suavizado é empregado, os test de igualdade e paralelismo presentan un correcto calibrado baixo a hipótesis nula e unha potencia tamén adecuada.

As propostas deste capítulo (test de non-efecto e contrastes de igualdade e paralelismo) foron ilustradas con datos reais (ver Apéndice A) nos tres contextos de regresión circular. No escenario circular-lineal, estudouse o efecto do ángulo de desequilibrio na magnitude correctora en volantes de inercia nos sistemas de transmisión de vehículos, onde o tipo de metal empregado na construción dos mesmos determina catro grupos de observacións. No caso lineal-circular, explorouse o efecto da temperatura na dirección de escape de pulgas de praia, onde o tipo de experimento actúa como variable explicativa categórica. Por último, no contexto circular-circular, analizouse a dirección de escape de grilos en función do ángulo de estímulo, consistente nunha corrente de aire, onde o tipo de resposta de escape (correndo ou saltando) conforma dous grupos de observacións. Os contidos deste capítulo están publicados en Alonso-Pena et al. (2021).

Capítulo 3: Un marco xeral para regresión circular. A estimación de características condicionais en contextos onde a variable explicativa Θ é circular e a variable resposta Y é xeral, con distribución tanto continua como discreta, abordouse neste capítulo. Unha vez especificada a distribución condicional de $Y|\Theta$, dependente dunha característica condicional de interese $g(\theta)$, constrúese a función de log-verosimilitude local, dada por

$$\mathcal{L}_p(\boldsymbol{\beta}; \kappa, \theta_0) = \sum_{i=1}^{n} \ell(\boldsymbol{\Theta}_i^T \boldsymbol{\beta}, Y_i) K_\kappa(\Theta_i - \theta_0),$$

onde $K_\kappa(\cdot)$ é unha función tipo núcleo circular con parámetro de suavizado κ, $\boldsymbol{\beta}$ é un vector de parámetros locais, $\boldsymbol{\Theta_i} = [1, \sin(\Theta_i - \theta_0), \ldots, \sin^p(\Theta_i - \theta_0)]^T$ con p dado, e $\ell(\cdot, \cdot)$ é a contribución de (Θ_i, Y_i) á log-verosimilitude, coa función de interese, $g(\Theta_i)$, approximada mediante o modelo local $\boldsymbol{\Theta}_i^T\boldsymbol{\beta}$. A estimación da función $g(\cdot)$ e das súas derivadas realízase mediante a maximización da log-verosimilitude local con respecto ao vector de parámetros locais.

Co obxectivo de realizar inferencia sobre a función de interese, desenvolvéronse métodos xerais para a estimación do nesgo e da varianza dos estimadores, baseados na selección dun paramétro de suavizado (concentración) piloto, e un axuste local previo de grao superior. Estas estimacións do nesgo e da varianza permitiron a realización de tarefas inferenciais como é a construción de bandas de confianza puntuais. Ademais, estas estimacións dán lugar a un método natural para a selección do parámetro de concentración na práctica, chamado regra refinada, xa que depende da selección previa dun parámetro de suavizado piloto. Propúxose tamén unha regra para seleccionar o parámetro piloto, mediante a construción e estudo teórico da función CRSC, deseñada para facer un balance entre o nesgo e a varianza dos estimadores.

Para examinar o comportamento das propostas en mostras finitas, explorouse este mediante un estudo de simulación, onde se consideraron distintos modelos con distribucións condicionais normal, Bernoulli, Poisson e gamma. O parámetro de concentración seleccionouse mediante a nova proposta da regra refinada, e comparouse o seu funcionamento cos estimadores resultantes ao seleccionar este parámetro directamente coa regra piloto ou co criterio da validación cruzada. Cos tres métodos obtívose un bo comportamento asintótico, destacando o funcionamento da regra refinada para escoller o parámetro de suavizado, que dá lugar, en xeral, a un menor valor da aproximación Monte Carlo do Erro Cadrático Medio Integrado (ECMI).

Para finalizar, ilustrouse a metodoloxía proposta con catro conxuntos de datos reais (ver Apéndice A para unha breve descrición dos mesmos). En primeiro lugar, explorouse a relación entre o ángulo da man e o *H-reflex index* en participantes dun experimento psicolóxico, asumindo que a distribución condicional é normal. En segundo lugar, estudouse a relación entre a dirección de escape de pulgas de praia ao ser soltadas na area e o tipo de especie ao que pertencen, realizando así unha regresión loxística non paramétrica con explicativa circular. Para ilustrar a metodoloxía no caso no que a distribución condicional é unha Poisson, analizouse o número de *spikes* neuronais en células de macacos e a súa relación co ángulo do estímulo visual que desencadena os *spikes*. Por último, explorouse a relación entre a concentración de partículas pm10 en Pontevedra, España, e a dirección do vento, asumindo unha distribución condicional gamma dada a asimetría condicional que presenta a concentración de pm10. Para considerar o efecto adicional da velocidade do vento, estendeuse a metodoloxía proposta a contextos máis xerais que permiten a entrada no modelo de máis variables explicativas, e de distinta natureza, como modelos aditivos ou modelos parcialmente lineais. Este estudo está recollido en Alonso-Pena et al. (2022b).

Capítulo 4: Estimación conxunta para modelos de regresión con sobre/infra-dispersión. A relación entre o número de *spikes* neuronais en macacos e a dirección de un estímulo visual estudouse no Capítulo 3, como ilustración do modelado do parámetro de localización (condicionalmente á covariable) cando a distribución condicional é unha Poisson. Non obstante, aínda que no exemplo de datos dos *spikes* a variable resposta é unha variable de conteo, unha análise exploratoria mostrou que a distribución condicional está lonxe de ser unha Poisson, dado que nesta distribución a media e a varianza coinciden. Aínda máis sorprendente resulta a fundamentada sospeita de que, neste conxunto de datos, a dispersión (conciente entre a varianza e a media) varía co valor da covariable circular. Ademais, na rama de estudo da neuroloxía, o correcto modelado da dispersión resulta fundamental para poder entender as respostas neuronais, suxeitas aos estímulos visuais (Lombardo et al., 2018).

Motivado por este exemplo real, de gran interese no campo neurolóxico, propúxose no Capítulo 4 unha metodoloxía para estimar, non parametricamente e de xeito conxunto, tanto a media como a dispersión en modelos con infra-dispersión e/ou sobre-dispersión. Para lograr este obxectivo, empregouse a metodoloxía da log-verosimilitude local, estudada no Capítulo 3, pero considerando como distribución condicional a familia exponencial dobre (Efron, 1986). Propúxose un algoritmo de estimación para modelar, conxuntamente, transformacións dos parámetros de localización e de dispersión en modelos con distribución condicional pertencente á familia exponencial dobre.

O funcionamento deste procedemento de estimación analizouse mediante un estudo de simulación, no que se consideraron modelos con sobre-dispersión, infra-dispersión ou con ambas, onde a función de dispersión cambia segundo o valor da covariable. Obtívose un comportamento satisfactorio dos estimadores, con aproximacións Monte Carlo do ECMI tendendo a cero a medida que aumenta o tamaño mostral. Concluiuse tamén, como era esperado, que a estimación da dispersión obtén maiores erros que a estimación da función de localización, dado que é un problema moito máis complexo, pero aínda así o estimador da dispersión permite recuperar a forma da función verdadeira. Ademais, propúxose unha representación gráfica que permite visualizar, ao mesmo tempo, a variabilidade dos datos e as estimacións das funcións media e dispersión.

Por último, aplicouse a métodoloxía proposta ao conxunto de datos dos macacos, comparando os resultados obtidos co método de Stevenson (2016). Concluíuse que, neste conxunto de datos, existe unha pronunciada sobre-dispersión con respecto ao modelo Poisson, pero que esta dispersión cambia considerablemente co valor do ángulo de estímulo visual, facendo que esta sobre-dispersión sexa menor cando o estímulo visual é horizontal. Os contidos deste capítulo pódense atopar en Alonso-Pena et al. (2022a).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Capítulo 5: Regresión multimodal circular. Este capítulo céntrase nas modas locais como característica condicional que é obxectivo do proceso de estimación, dentro dos tres escenarios de regresión circular: circular-lineal, lineal-circular e circular-circular. Ao contrario que nos dous capítulos anteriores, non se especifica a distribución condicional para poder estimar

as modas locais, senón que estas se obteñen directamente dos datos.

No caso circular-lineal, onde se ten unha variable explicativa circular Θ e unha resposta escalar Y, a multifunción de regresión defínese como

$$M(\theta) = \left\{ y \in \mathbb{R} : \frac{\partial}{\partial y} f(y|\theta) = 0, \frac{\partial^2}{\partial y^2} f(y|\theta) < 0 \right\},$$

onde $f(y|\theta)$ denota a densidade condicional. Nos outros dous escenarios, a multifunción de regresión defínese de xeito análogo como

$$M(\delta) = \left\{ \phi \in \mathbb{T} : \frac{\partial}{\partial \phi} f(\phi|\delta) = 0, \frac{\partial^2}{\partial \phi^2} f(\phi|\delta) < 0 \right\},$$

onde Δ denota unha variable explicativa xeral (escalar ou circular, tomando valor δ) e Φ é a variable resposta circular, , tomando valor ϕ. A estimación das multifuncións realizouse de xeito indirecto, onde primeiramente a densidade condicional é estimada cun método tipo núcleo e, a continuación, as modas locais da densidade condicional son obtidas mediante un método númerico. Este método numérico é, no escenario circular-lineal, unha adaptación da versión condicional do algoritmo *mean shift*, onde os pesos dados polas funcións tipo núcleo foron debidamente modificados para ter en conta o carácter circular da variable explicativa. Nos dous escenarios restantes, empregouse unha versión condicional do *mean shift* circular, que se pode interpretar como un tipo de algoritmo de ascenso de gradiente na circunferencia.

Analizáronse as propiedades dos estimadores nos tres escenarios, obtendo as ordes de converxencia destes. Concluiuse que, baixo certas hipóteses, estas ordes coinciden coas ordes de converxencia da primeira derivada parcial do estimador da densidade conxunta. Ademais, levouse a cabo un estudo de simulación para comprobar o funcionamento da metodoloxía proposta con tamaños mostrais finitos. Neste estudo, os parámetros de suavizado foron seleccionados mediante un novo criterio de validación cruzada modal, baseado no proposto por Zhou e Huang (2019). Este criterio non está baseado en resultados teóricos, polo que o seu comportamento foi comparado coa selección *óptima* dos parámetros, buscados de xeito computacional de forma que minimicen o erro global da estimación, construído a partir da distancia de Hausdorff. No estudo de simulación obtívose que, aínda que, como se esperaba, a validación cruzada modal presenta peores resultados que cando se utilizan os parámetros de referencia, a actuación dos estimadores cos dous métodos de selección dos parámetros de suavizado é bastante parecida, en especial cando o tamaño mostral é grande.

Para finalizar, aplicouse a metodoloxía da regresión multimodal circular a un conxunto de datos reais tomados dun estudo experimental sobre peixes cebra (ver Apéndice A). Neste exemplo, estudouse a dirección de escape dos peixes en función do ángulo de visión co que detectaban a presenza dun depredador robot. Os contidos deste capítulo están recollidos en Alonso-Pena e Crujeiras (2022).

Capítulo 6: Conclusións, discusión, e traballo futuro. Este capítulo contén as liñas de discusión finais e conclusións da tese, así como as distintas vías de investigación que abre esta disertación. En concreto, faise fincapé nos principais resultados da investigación e nas conexións entre os distintos capítulos da tese. Ademais, como futuras liñas de investigación plantéxanse, entre outras, a extensión da metodoloxía actual a datos con soporte na hiperesfera, o uso de parámetros de suavizado locais, contrastes de hipóteses no ámbito da log-verosimilitude local, a estimación mediante esta en situacións onde a variable resposta é circular e distintos avances no marco da regresión multimodal.

Apéndice A: Conxuntos de datos reais. Neste apéndice, descríbense os distintos conxuntos de datos reais empregados ao longo da tese como ilustración dos distintos métodos propostos. O primeiro deles é relativo ao ámbito da mecánica, e contén os ángulos de desequilibrio de volantes de inercia (pezas mecánicas utilizadas nos sistemas de transmisión dos vehículos), ademais da magnitude da medida correctora necesaria para equilibralos. O segundo conxunto de datos contén observacións sobre o ángulo de fuga dun grupo de pulgas de praia, que son soltadas na area, ademais doutros datos como a especie dos animais ou certas características ambientais. En terceiro lugar, preséntase un conxunto de datos tamén relacionado co escape animal, desta vez en grilos, onde ademais da dirección de escape tense a dirección de estímulo (dado por unha corrente de aire) e o tipo de resposta ao escapar (correndo ou saltando). O cuarto conxunto de datos provén dun estudo psicolóxico onde se estuda o ángulo de posición da man dos participantes e unha medida de resposta obtida mediante a técnica *H-reflex*. En quinto lugar, descríbese un conxunto de datos pertencente ao campo da neuroloxía, acerca de un estudo experimental en macacos, onde os macacos reciben un estímulo visual. O conxunto de datos contén o ángulo do estímulo visual e o número de *spikes* neuronais nunha determinada célula na rexión MT/V5. O seguinte conxunto de datos contén a concentración de partículas pm10 en Pontevedra, España, recollida durante 2019, ademais da dirección e da velocidade do vento. Por último, preséntanse datos relativos a un experimento conductal en larvas de peixe cebra, que foron perseguidos por un robot imitando un depredador. Neste caso, téñense datos do ángulo de escape das larvas no instante no que detectan o ataque, ademais do ángulo no que perciben a presenza do robot.

Apéndice B: Resultados de simulación adicionais. Neste apéndice, preséntanse resultados de simulación adicionais aos expostos na parte principal da memoria. En primeiro lugar, dánse resultados que complementan os obtidos no Capítulo 2 da memoria, completando así o estudo dos contrastes de non efecto, igualdade e paralelismo das curvas de regresión. En segundo lugar, proporciónanse resultados de simulación relativos á selección dos parámetros de suavizado no contexto da regresión multimodal circular, complementando así o Capítulo 5 da memoria.

Apéndice C: Resultados teóricos. Este apéndice contén as demostracións técnicas correspondentes aos capítulos da memoria, en concreto os resultados derivados dos Capítulos 2, 3 e 5. Ademais, expóñense as suposicións necesarias para obter os resultados relativos ao Capítulo 5 da memoria.

Apéndice D: Software: actualización do paquete `NPCirc`. Co obxectivo de facer as propostas metodolóxicas desa tese doutoral accesibles á comunidade científica e permitir o seu uso en contextos aplicados, programáronse funcións de `R` (R Core Team, 2021) que permiten poñer estas ferramentas en práctica. Estas funcións foron incluídas no paquete `NPCirc` (Oliveira et al., 2014b). A versión `3.0.1` do paquete foi publicada en Xullo de 2021, coas funcións correspondentes ao Capítulo 2 da memoria. A nova versión `3.1.1` contén tamén as ferramentas relativas aos Capítulos 3, 4 e 5, así como os conxuntos de datos descritos no Apéndice A. Así, o Apéndice D da memoria contén unha breve introdución ao paquete e ás súas actualizacións, así como a documentación das novas funcións.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Resumen en castellano

Una de las tareas más importantes en el campo de la estadística es el estudio de la relación de dependencia entre dos o más variables aleatorias. Esta dependencia se puede modelar de formas muy diversas, siendo uno de los enfoques más empleados el uso de modelos de regresión donde, en el marco clásico, se suele tratar de explicar la esperanza condicional de una variable respuesta Y mediante una función de la variable explicativa X. En general, se asume que la función de regresión tiene una forma explícita, que depende de unos parámetros que es necesario estimar y sobre los que es posible realizar inferencia. La forma paramétrica más simple es una linea recta, lo que da lugar al modelo de regresión lineal.

Sin embargo, en la práctica, asumir que la relación entre dos variables es lineal puede resultar muy restrictivo y no reflejar fielmente la relación existente entre las variables. Aunque es posible expresar la media condicional de la variable respuesta como una función más compleja dependiente de un vector de parámetros, a menudo el uso de modelos demasiado complicados puede resultar en problemas de mala especificación o de sobreajuse. Es en este contexto donde entran en juego los modelos de regresión no paramétricos (Wand y Jones, 1995; Fan y Gijbels, 1996), en los que no se asume ninguna forma específica para la media condicional, sino que ésta se estima directamente a partir de los propios datos. En la mayor parte de la literatura relativa a la regresión no paramétrica, se entiende la función de regresión como la media o esperanza condicional, pero existen otro tipo de técnicas donde la función de interés no es necesariamente la media, sino otra característica condicional de la variable respuesta, dado el valor de la variable explicativa.

Más allá del interés por la generación de conocimiento en este ámbito y la comprensión matemática de los procesos de relación que involucran variables de naturaleza circular, en muchos campos aplicados como la biología, la psicología o la meteorología, existen medidas de interés que son direcciones o ángulos, por lo que su soporte no es el espacio euclídeo. Este tipo de datos deben ser tratados con técnicas estadísticas que tengan en cuenta la naturaleza periódica de las variables, para lo que es necesario definir una dirección inicial y un sentido de rotación. Algunas referencias básicas en el contexto de la estadística circular (o su extensión a otras variedades direccionales) son Mardia (1972), Fisher (1993), Mardia y Jupp (2000), Jammalamadaka y SenGupta (2001) o Ley y Verdebout (2017). En estos volúmenes se recogen herramientas inferenciales, mayoritariamente paramétricas y centradas en el estudio de distribuciones de probabilidad.

Las observaciones circulares pueden venir acompañadas de otras mediciones, que pueden

ser tanto datos escalares como circulares, y en muchos casos resulta necesario estudiar la relación de dependencia entre estas variables. Aunque existen algunos modelos de regresión paramétricos que contemplan la existencia de variables circulares, la literatura de modelos de regresión circulares no paramétricos es bastante escasa, en especial la relacionada con la estimación de otras características condicionales distintas de la media. Por lo tanto, el objetivo principal de esta tesis es la propuesta y estudio de modelos de regresión no paramétricos con variables circulares, considerando distintas características condicionales como funciones de interés.

En lo que sigue, se expone un resumen de los capítulos que conforman esta tesis doctoral, destacando las principales aportaciones y avances recogidos en cada uno de ellos, así como un breve resumen de los apéndices de la memoria.

Capítulo 1: Introducción. En este capítulo se presentan distintos conceptos que serán abordados a lo largo de la memoria y se exponen, de forma detallada, los principales objetivos de la tesis.

En primer lugar, se proporciona una introducción a las técnicas tipo núcleo para la estimación de la función de regresión en el contexto escalar, considerando la estimación de diferentes características condicionales. La función de regresión clásica se define en un punto dado x_0 como

$$m(x_0) = \mathbb{E}(Y|X = x_0),$$

es decir, la esperanza de la variable respuesta condicionada al valor de la variable explicativa. La función $m(\cdot)$ se puede estimar de forma no paramétrica con estimadores tipo núcleo, como pueden ser el estimador de Nadaraya-Watson (Nadaraya, 1964; Watson, 1964) o el estimador polinómico local (Wand y Jones, 1995; Fan y Gijbels, 1996) donde, para llevar a cabo el proceso de estimación, es necesario seleccionar un parámetro ventana, que determina la suavidad de las curvas de regresión estimadas. Una cuestión importante en la estimación de la función de regresión es la determinación de la significación del efecto de la variable explicativa en el valor esperado de la respuesta. Con la finalidad de resolver esta problemática, Bowman y Azzalini (1997) propusieron un contraste de hipótesis basado en la utilización de estimadores tipo núcleo que permite evaluar la significación de la covariable. El modelo de regresión basado en la modelización de la media se puede extender a contextos más complejos, donde la existencia de una variable aleatoria categórica determina la división de la muestra en distintos grupos de observaciones. En estos casos, se pueden emplear los llamados modelos de análisis de la covarianza no paramétricos (Young y Bowman, 1995). Estos modelos permiten, además, llevar a cabo contrastes de hipótesis para determinar la igualdad o paralelismo de las curvas de regresión entre los distintos grupos de observaciones.

Aunque los métodos de estimación tipo núcleo proporcionan una forma flexible para modelar la función de regresión sin especificar una forma concreta para ésta, este enfoque resulta de utilidad cuando la esperanza del error asociado al modelo de regresión es cero, suposición que no siempre se cumple en la práctica. Una clase más amplia de estimadores tipo

núcleo, adecuados para variables respuesta de tipo genérico, fue introducida por Staniswalis (1989) y Fan et al. (1995), que extendieron la idea de la verosimilitud local propuesta por Tibshirani y Hastie (1987). En este contexto, la distribución condicional de las variables se asume conocida, y se considera como función de regresión una característica condicional de ésta. El proceso de estimación consiste en la maximización de la log-verosimilitud local, en la que las contribuciones individuales de cada observación se ponderan empleando una función tipo núcleo, dependiente de un parámetro ventana. Este método permite la estimación de características condicionales en contextos donde la variable respuesta es tanto continua como discreta. Además, Fan et al. (1998) propusieron una metodología para aproximar el sesgo y la varianza de los estimadores, así como un método general para seleccionar el parámetro de suavizado.

No obstante, si la distribución condicional presenta múltiples modas, la regresión multimodal se erige como una alternativa adecuada. La idea de la estimación de las modas locales condicionales fue brevemente abordada por Scott (1992), y estudiada en más profundidad por Einbeck y Tutz (2006) y Chen et al. (2016a). En este contexto, la función de interés no es realmente una función, sino una multifunción o función multievaluada. El proceso de estimación es indirecto, obteniendo en primer lugar un estimador tipo núcleo de la densidad condicional, para después calcular las modas locales de dicho estimador mediante una versión condicional del algoritmo *mean shift*. Sendas revisiones de este tema pueden consultarse en Chen (2018) y Alonso-Pena (2020).

La estimación tipo núcleo de ciertas características condicionales también se puede llevar a cabo sin especificar la distribución condicional. Por ejemplo, la estimación de la mediana o de los cuantiles condicionales fue abordada por Chaudhuri (1991a) y Chaudhuri (1991b). Menos conocida es la estimación de la moda condicional (Yao et al., 2012), que da lugar a una forma de estimación más robusta que las anteriormente mencionadas. Tanto la regresión en mediana como la regresión modal son de utilidad cuando la distribución condicional, aunque desconocida, es unimodal. Sin embargo, si la distribución condicional exhibe múltiples modas, la regresión multimodal se presenta como una alternativa adecuada. La idea de la estimación de las modas locales condicionales fue brevemente abordada por Scott (1992), y estudiada en más profundidad por Einbeck y Tutz (2006) y Chen et al. (2016a). En este contexto, la función de interés no es realmente una función, sino una multifunción o función multievaluada. El proceso de estimación es indirecto, obteniendo en primer lugar un estimador tipo núcleo de la densidad condicional, para después calcular las modas locales de dicho estimador mediante una versión condicional del algoritmo *mean shift*. Sendas revisiones de este tema pueden consultarse en Chen (2018) y Alonso-Pena (2020).

La segunda parte del Capítulo 1 introduce las herramientas clásicas en el contexto de la estadística circular, en la que se parte de observaciones con soporte la circunferencia unidad $\mathbb{T} = [0, 2\pi)$. Se introduce el concepto de densidad de una variable aleatoria circular, que es análogo al de una función de densidad usual con la condición extra de la periodicidad de la función. Esta frecuencia con período 2π hace que las medidas usuales de localización y

dispersión de una variable aleatoria circular tengan que ser redefinidas, así como sus análogos muestrales. Se introducen también las principales funciones de densidad circulares, entre las que destaca la densidad de von Mises, dada por

$$f(\theta;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa\cos(\theta-\mu)], \quad \theta \in \mathbb{T},$$

donde μ representa el parámetro de localización, κ es el parámetro de concentración y $I_0(\cdot)$ es la función de Bessel de clase uno y orden cero, que actúa como constante normalizadora. Aunque hay funciones de densidad específicamente enfocadas al modelado de variables circulares, existen métodos para construir densidades circulares a partir de funciones de densidad para variables aleatorias escalares como, por ejemplo, las densidades enrolladas o proyectadas. No obstante, al igual que sucede en el contexto de la regresión, estas formas paramétricas suelen ser demasiado restrictivas en la práctica, por lo que, como herramienta exploratoria e inferencial, se puede recurrir al estimador tipo núcleo de la densidad circular, dado por

$$\hat{f}(\theta_0) = \frac{1}{n}\sum_{i=1}^{n} K_\kappa(\Theta_i - \theta_0),$$

donde $K_\kappa(\cdot)$ es una función tipo núcleo circular con parámetro de concentración κ, que actúa como parámetro de suavizado.

Por último, se presentan los escenarios de regresión con variables circulares, contexto donde se enmarca esta tesis doctoral, y los principales modelos paramétricos. Se pueden distinguir tres escenarios diferentes en el marco de la regresión circular: el caso con una variable explicativa circular Θ y una respuesta escalar Y, que se conoce como regresión circular-lineal; la situación contraria, en la que la variable explicativa X es escalar y la respuesta, Φ, tiene naturaleza circular, conocida como regresión lineal-circular; y el escenario donde ambas variables, Θ y Φ, son circulares, *i.e.*, regresión circular-circular. En el contexto de la regresión circular-lineal, se suele modelar la esperanza de Y condicionada al valor de X mediante una combinación lineal del seno y coseno de X, aunque modelos más flexibles, con más términos trigonométricos, son plausibles. En lo referente a la regresión lineal-circular, los modelos más utilizados son la propuesta de Fisher y Lee (1992) y el modelo proyectado de Presnell et al. (1998). La literatura correspondiente a la regresión circular-circular es más escasa, destacando el modelo presentado por Jammalamadaka y SenGupta (2001), donde el tratamiento de la variable circular se realiza considerando marginalmente sus componentes seno y coseno.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Capítulo 2: Inferencia no paramétrica para regresión circular. En este capítulo se presentan, en primer lugar, los estimadores tipo núcleo de la regresión en media para variables circulares. En el contexto de la regresión con explicativa circular y respuesta escalar, el mo-

delo de regresión en media viene dado por

$$Y = m(\Theta) + \varepsilon,$$

donde ε es un error aleatorio de media cero. Di Marzio et al. (2009) propusieron un estimador tipo núcleo de $m(\cdot)$ basado en una modificación de la expansión de Taylor, en la que la función identidad es aproximada por la función seno para valores cercanos al cero. Por otra parte, cuando la variable respuesta es circular, se tiene el siguiente modelo de regresión:

$$\Phi = [m(\Delta) + \varepsilon](\text{mod } 2\pi),$$

donde Δ es una variable aleatoria que puede ser tanto circular como escalar, y ε es un error aleatorio circular con dirección media cero. En este escenario, Di Marzio et al. (2013) propusieron un estimador de $m(\cdot)$ mediante la aplicación del operador atan2$(\cdot, \cdot)$ a los estimadores tipo núcleo de la media condicional del seno y coseno de la respuesta.

Con el objetivo de comprobar el efecto de las variables explicativas en los modelos anteriores, se propusieron en esta tesis contrastes del efecto de la covariable en los tres escenarios de regresión. En el contexto circular-lineal, el estadístico de contraste se construye midiendo la distancia cuadrática entre los estimadores del modelo bajo la hipótesis nula y la alternativa, mientras que en los escenarios lineal-circular y circular-circular, la distancia cuadrática se sustituye por la distancia en coseno. Se plantearon también modelos de análisis de la covarianza (ANCOVA) no paramétricos, en los que se considera una variable explicativa categórica adicional, determinando I grupos de observaciones. En cada uno de los tres contextos de regresión circular, se propusieron contrastes de hipótesis para determinar tanto la igualdad como el paralelismo de las curvas de regresión entre los I grupos, siguiendo las ideas de Young y Bowman (1995).

Para poder analizar el funcionamiento de las distintas propuestas de contraste, se realizó un estudio de simulación en el que se consideraron diferentes modelos, tamaños muestrales y niveles de significación. En los test de no-efecto, se exploró la influencia del parámetro de suavizado, determinando que un parámetro adecuado para la estimación da lugar a un incorrecto calibrado. Se determinó que este efecto viene dado por el sesgo del estimador no paramétrico bajo la hipótesis alternativa, y que la infrasuavización del estimador de la curva de regresión conlleva un correcto calibrado bajo la hipótesis nula y una potencia adecuada. Por otra parte, el problema del sesgo desaparece en los contrastes relativos a los modelos ANCOVA cuando el mismo parámetro de suavizado es utilizado bajo cada una de las hipótesis, cancelando así el sesgo de los estimadores. Se concluyó que cuando el mismo parámetro de suavizado es utilizado, los test de igualdad y paralelismo presentan un correcto calibrado bajo la hipótesis nula y una potencia también adecuada.

Las propuestas de este capítulo (test de no-efecto y contrastes de igualdad y paralelismo) fueron ilustradas con datos reales (ver Apéndice A) en los tres contextos de regresión circular. En el escenario circular-lineal, se estudió el efecto del ángulo de desequilibrio en la magnitud

correctora en volantes de inercia en los sistemas de transmisión de vehículos, donde el tipo de metal empleado en la construcción de los mismos determina cuatro grupos de observaciones. En el caso lineal-circular, se exploró el efecto de la temperatura en la dirección de escape de pulgas de playa, donde el tipo de experimento actúa como variable explicativa categórica. Por último, en el contexto circular-circular, se analizó la dirección de escape de grillos en función del ángulo de estímulo, consistente en una corriente de aire, donde el tipo de respuesta de escape (corriendo o saltando) conforma dos grupos de observaciones. Los contenidos de este capítulo están publicados en Alonso-Pena et al. (2021).

Capítulo 3: Un marco general para regresión circular. La estimación de características condicionales en contextos donde la variable explicativa Θ es circular y la variable respuesta Y es general, con distribución tanto continua como discreta, fue abordada en este capítulo. Una vez especificada la distribución condicional de $Y|\Theta$, dependiente de una característica condicional de interés $g(\theta)$, se construye la función de log-verosimilitud local, dada por

$$\mathcal{L}_p(\boldsymbol{\beta}; \kappa, \theta_0) = \sum_{i=1}^{n} \ell(\boldsymbol{\Theta}_i^T \boldsymbol{\beta}, Y_i) K_\kappa(\Theta_i - \theta_0),$$

donde $K_\kappa(\cdot)$ es una función núcleo circular con parámetro de suavizado κ, $\boldsymbol{\beta}$ es un vector de parámetros locales, $\boldsymbol{\Theta_i} = [1, \sin(\Theta_i - \theta_0), \ldots, \sin^p(\Theta_i - \theta_0)]^T$ con p dado, y $\ell(\cdot, \cdot)$ es la contribución de (Θ_i, Y_i) a la log-verosimilitud, con la función de interés, $g(\Theta_i)$, approximada mediante el modelo local $\boldsymbol{\Theta}_i^T \boldsymbol{\beta}$. La estimación de la función $g(\cdot)$ y de sus derivadas se realiza mediante la maximización de la log-verosimilitud local con respecto al vector de parámetros locales.

Con el objetivo de realizar inferencia sobre la función de interés, se desarrollaron métodos generales para la estimación del sesgo y de la varianza de los estimadores, basados en la selección de un parámetro de suavizado (concentración) piloto, y un ajuste local previo de grado superior. Estas estimaciones del sesgo y de la varianza permitieron la realización de tareas inferenciales como es la construcción de bandas de confinanza puntuales. Además, estas estimaciones dan lugar a un método natural para la selección del parámetro de concentración en la práctica, llamado regla refinada, ya que depende de la selección previa de un parámetro de suavizado piloto. Se propuso también una regla para seleccionar el parámetro piloto, mediante la construcción y estudio teórico de la función CRSC, diseñada para hacer un balance entre el sesgo y la varianza de los estimadores.

Para examinar el comportamiento de las propuestas en muestras finitas, se exploró éste mediante un estudio de simulación, donde se consideraron distintos modelos con distribuciones condicionales normal, Bernoulli, Poisson y gamma. El parámetro de concentración se seleccionó mediante la nueva propuesta de la regla refinada, y se comparó su funcionamiento con los estimadores resultantes al seleccionar este parámetro directamente con la regla piloto o con el criterio de la validación cruzada. Con los tres métodos se obtuvo un buen comportamiento asintótico, destacando el funcionamiento de la regla refinada para escoger

el parámetro de suavizado, que da lugar, en general, a un menor valor de la aproximación Monte Carlo del Error Cuadrático Medio Integrado (ECMI).

Para finalizar, se ilustró la metodología propuesta con cuatro conjuntos de datos reales (ver el Apéndice A para una breve descripción de los mismos). En primer lugar, se exploró la relación entre el ángulo de la mano y el *H-reflex index* en participantes de un experimento psicológico, asumiendo que la distribución condicional es normal. En segundo lugar, se estudió la relación entre la dirección de escape de pulgas de playa al ser soltadas en la arena y el tipo de especie al que pertenecen, realizando así una regresión logística no paramétrica con explicativa circular. Para ilustrar la metodología en el caso en el que la distribución condicional es una Poisson, se analizó el número de *spikes* neuronales en células de macacos y su relación con el ángulo del estímulo visual que desencadena los *spikes*. Por último, se exploró la relación entre la concentración de partículas pm10 en Pontevedra, España, y la dirección del viento, asumiendo una distribución condicional gamma dada la asimetría condicional que presenta la concentración de pm10. Para considerar el efecto adicional de la velocidad del viento, se extendió la metodología propuesta a contextos más generales que permiten la entrada en el modelo de más variables explicativas, y de distinta naturaleza, como modelos aditivos o modelos parcialmente lineales. Este estudio está recogido en Alonso-Pena et al. (2022b).

Capítulo 4: Estimación conjunta para modelos de regresión con sobre/infra-dispersión. La relación entre el número de *spikes* neuronales en macacos y la dirección de un estímulo visual se estudió en el Capítulo 3, como ilustración del modelado del parámetro de localización (condicionalmente a la covariable) cuando la distribución condicional es una Poisson. Sin embargo, aunque en el ejemplo de datos de los *spikes* la variable respuesta es una variable de conteo, un análisis exploratorio mostró que la distribución condicional está lejos de ser una Poisson, dado que en esta distribución la media y la varianza coinciden. Aún más sorprendente resulta la fundamentada sospecha de que, en este conjunto de datos, la dispersión (cociente entre la varianza y la media) varía con el valor de la covariable circular. Además, en la rama de estudio de la neurología, el correcto modelado de la dispersión resulta fundamental para poder entender las respuestas neuronales, sugetas a los estímulos visuales (Lombardo et al., 2018).

Motivado por este ejemplo real, de gran interés en el campo neurológico, se propuso, en el Capítulo 4, una metodología para estimar, no paramétricamente y de forma conjunta, tanto la media como la dispersión en modelos con infra-dispersión y/o sobre-dispersión. Para lograr este objetivo, se empleó la metodología de la log-verosimilitud local, estudiada en el Capítulo 3, pero considerando como distribución condicional a la familia exponencial doble (Efron, 1986). Se propuso un algoritmo de estimación para modelar, conjuntamente, transformaciones de los parámetros de localización y de dispersión en modelos con distribución condicional perteneciente a la familia exponencial doble.

El funcionamiento de este procedimiento de estimación se analizó mediante un estudió de

simulación, en el que se consideraron modelos con sobre-dispersión, infra-dispersión o con ambas, donde la función de dispersión cambia según el valor de la covariable. Se obtuvo un comportamiento satisfactorio de los estimadores, con aproximaciones Monte Carlo del ECMI tendiendo a cero a medida que aumenta el tamaño muestral. Se concluyó también, como era esperado, que la estimación de la dispersión da lugar a mayores errores que la estimación de la función de localización, dado que es un problema mucho más complejo, pero aún así el estimador de la dispersión permite recuperar la forma de la función verdadera. Además, se propuso una representación gráfica que permite visualizar, al mismo tiempo, la variabilidad de los datos y las estimaciones de las funciones media y dispersión.

Por último, se aplicó la metodología propuesta al conjunto de datos de los macacos, comparando los resultados obtenidos con el método de Stevenson (2016). Se concluyó que, en este conjunto de datos, existe una pronunciada sobre-dispersión con respecto al modelo Poisson, pero que esta dispersión cambia considerablemente con el valor del ángulo de estímulo visual, haciendo que esta sobre-dispersión sea menor cuando el estímulo visual es horizontal. Los contenidos de este capítulo se pueden encontrar en Alonso-Pena et al. (2022a).

Capítulo 5: Regresión multimodal circular. Este capítulo se centra en las modas locales como característica condicional que es objetivo del proceso de estimación, dentro de los tres escenarios de regresión circular: circular-lineal, lineal-circular y circular-circular. Al contrario que en los dos capítulos anteriores, no se especifica la distribución condicional para poder estimar las modas locales, sino que éstas se obtienen directamente de los datos.

En el caso circular-lineal, donde se tiene una variable explicativa circular Θ y una respuesta escalar Y, la multifunción de regresión se define como

$$M(\theta) = \left\{ y \in \mathbb{R} : \frac{\partial}{\partial y} f(y|\theta) = 0, \frac{\partial^2}{\partial y^2} f(y|\theta) < 0 \right\},$$

donde $f(y|\theta)$ denota la densidad condicional. En los otros dos escenarios, la multifunción de regresión se define de forma análoga como

$$M(\delta) = \left\{ \phi \in \mathbb{T} : \frac{\partial}{\partial \phi} f(\phi|\delta) = 0, \frac{\partial^2}{\partial \phi^2} f(\phi|\delta) < 0 \right\},$$

donde Δ denota una variable explicativa general (escalar o circular, tomando valor δ) y Φ es la variable respuesta circular (tomando valor ϕ). La estimación de las multifunciones se realizó de forma indirecta, donde, en primer lugar, la densidad condicional es estimada con un método tipo núcleo y, a continuación, las modas locales de la densidad condicional son obtenidas mediante un método númerico. Este método numérico es, en el escenario circular-lineal, una adaptación de la versión condicional del algoritmo *mean shift*, donde los pesos dados por las funciones tipo núcleo fueron debidamente modificados para tener en cuenta el carácter circular de la variable explicativa. En los dos escenarios restantes, se empleó una versión condicional del *mean shift* circular, que se puede interpretar como un tipo de

algoritmo de ascenso de gradiente en la circunferencia.

Se analizaron las propiedades de los estimadores en los tres escenarios, obteniendo los órdenes de convergencia de éstos. Se concluyó que, bajo ciertas hipótesis, estos órdenes coinciden con los órdenes de convergencia de la primera derivada parcial del estimador de la densidad conjunta. Además, se llevó a cabo un estudio de simulación para comprobar el funcionamiento de la metodología propursta con tamaños muestrales finitos. En este estudio, los parámetros de suavizado fueron seleccionados mediante un nuevo criterio de validación cruzada modal, basado en el propuesto por Zhou y Huang (2019). Este criterio no está basado en resultados teóricos, por lo que su comportamiento fue comparado con la selección *óptima* de los parámetros, buscados de forma computacional de modo que minimicen el error global de la estimación, construido con la distancia de Hausdorff. En el estudio de simulación se obtuvo que, aunque como se esperaba, la validación cruzada modal presenta peores resultados que cuando se utilizan los parámetros de referencia, la actuación de los estimadores con los dos métodos de selección de los parámetros de suavizado es bastante parecida, en especial cuando el tamaño muestral es grande.

Para finalizar, se aplicó la metodología de la regresión multimodal circular a un conjunto de datos reales tomados de un estudio experimental sobre peces cebra (ver Apéndice A). En este ejemplo, se estudió la dirección de escape de los peces en función del ángulo de visión con el que detectaban la presencia de un depredador robot. Los contenidos de este capítulo están recogidos en Alonso-Pena y Crujeiras (2022).

Capítulo 6: Conclusiones, discusión, y trabajo futuro. Este capítulo contiene las lineas de discusión finales y conclusiones de la tesis, así como las distintas vías de investigación que abre esta disertación. En concreto, se hace hincapié en los principales resultados de la investigación y en las conexiones entre los distintos capítulos de la tesis. Además, como futuras lineas de investigación se plantean, entre otras, la extensión de la metodología actual a datos con soporte en la hiperesfera, el uso de parámetros de suavizado locales, contrastes de hipótesis en el ámbito de la log-verosimilitud local, la estimación mediante ésta en situaciones donde la variable respuesta es circular y distintos avances en el marco de la regresión multimodal.

Apéndice A: Conjuntos de datos reales. En este apéndice, se describen los distintos conjuntos de datos reales empleados a lo largo de la tesis como ilustración de los distintos métodos propuestos. El primero de ellos es relativo al ámbito de la mecánica, y contiene los ángulos de desequilibrio de volantes de inercia (piezas mecánicas utilizadas en los sistemas de transmisión de los vehículos), además de la magnitud de la medida correctora necesaria para equilibrarlos. El segundo conjunto de datos contiene observaciones sobre el ángulo de fuga de un grupo de pulgas de playa, que son soltadas en la arena, además de otros datos como la especie de los animales o ciertas características ambientales. En tercer lugar, se presenta un conjunto de datos también relacionado con el escape animal, esta vez en grillos,

donde además de la dirección de escape se tiene la dirección de estímulo (dado por una corriente de aire) y el tipo de respuesta al escapar (corriendo o saltando). El cuarto conjunto de datos proviene de un estudio psicológico donde se estudia el ángulo de posición de la mano de los participantes y una medida de respuesta obtenida mediante la técnica *H-reflex*. En quinto lugar, se describe un conjunto de datos pertenciente al campo de la neurología, acerca de un estudio experimental en macacos, donde los macacos reciben un estímulo visual. El conjunto de datos contiene el ángulo del estímulo visual y el número de *spikes* neuronales en una determinada célula en la región MT/V5. El siguiente conjunto de datos contiene la concentración de partículas pm10 en Pontevedra, España, recogida durante 2019, además de la dirección y de la velocidad del viento. Por último, se presentan datos relativos a un experimento conductual en larvas de pez cebra, que fueron perseguidos por un robot imitando un depredador. En este caso, se tienen datos del ángulo de escape de las larvas en el instante en el que detectan el ataque, además del ángulo en el que perciben la presencia del robot.

Apéndice B: Resultados de simulación adicionales. En este apéndice, se presentan resultados de simulación adicionales a los expuestos en la parte principal de la memoria. En primer lugar, se dan resultados que complementan a los obtenidos en el Capítulo 2 de la memoria, completando así el estudio de los contrastes de no efecto, igualdad y paralelismo de las curvas de regresión. En segundo lugar, se proporcionan resultados de simulación relativos a la selección de los parámetros de suavizado en el contexto de la regresión multimodal circular, complementando así el Capítulo 5 de la memoria.

Apéndice C: Resultados teóricos. Este apéndice contiene las demostraciones técnicas correspondientes a los capítulos de la memoria, en concreto los resultados derivados de los Capítulos 2, 3 y 5. Además, se exponen las suposiciones necesarias para obtener los resultados relativos al Capítulo 5 de la memoria.

Apéndice D: Software: actualización del paquete `NPCirc`. Con el objetivo de hacer las propuestas metodológicas de esta tesis doctoral accesibles a la comunidad científica y permitir su uso en contextos aplicados, se programaron funciones de `R` (R Core Team, 2021) que permiten poner estas herramientas en práctica. Estas funciones fueron incluidas en el paquete `NPCirc` (Oliveira et al., 2014b). La versión `3.0.1` del paquete fue publicada en julio de 2021, con las funciones correspondientes al Capítulo 2 de la memoria. La nueva versión `3.1.1` contiene también las herramientas relativas a los Capítulos 3, 4 y 5, así como los conjuntos de datos descritos en el Apéndice A. Así, el Apéndice D de la memoria contiene una breve introducción al paquete y a sus actualizaciones, así como la documentación de las nuevas funciones.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Further information

In compliance with the rules of doctoral studies at Universidade de Santiago de Compostela collected in *Regulamento dos estudos de doutoramento na USC, DOG de 16 de setembro de 2020*, some information on the articles that support the contributions of this thesis is provided, as well as information about the journals in which such works are published. In particular, the authors' names, journals' names, publishing years, ISSN, publishers, DOI-type links, impact factors, JCR and Scopus quartiles and CiteScore rating are provided, in addition to the specific contribution of the PhD candidate. Furthermore, some relevant information regarding copyright permissions for the published articles is given.

Articles and journals

Information about the works in which this dissertation is based is given in this section. First, details about the journals in which some of the works are published are provideded. A list of unpublished works is given afterwards.

Published and accepted works

The following works, corresponding to parts of Chapters 1, 2 and 5, have been either published or accepted. Information about the articles, journals and copyright permissions is provided hereafter.

Chapter 1: Alonso-Pena (2020)

TITLE: An introduction to nonparametric multimodal regression.
AUTHOR: M. Alonso-Pena[1].
AFFILIATIONS:
[1] Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela.
JOURNAL: Boletín de Estadística e Investigación Operativa.
YEAR: 2020.
STATUS: Published.
ISSN: 2387-1725.
PUBLISHER: Sociedad de Estadística e Investigación Operativa (SEIO).

LINK: www.seio.es/BBEIO/BEIOVol36Num1/index.html.
SJR IMPACT FACTOR: 0.127 [Q4 in Statistics, Probability and Uncertainty (147/153)].
CITESCORE: 0.2.
INFORMATION REGARDING COPYRIGHT AND USE: The following statement can be found in the copyright transfer agreement (see the Permissions section):
The author retains the right to use the substance of the above work in future works, including lectures, press releases and reviews, provided that you acknowledge its prior publication in the journal.

Chapter 2: Alonso-Pena et al. (2021)

TITLE: Nonparametric tests for circular regression.
AUTHORS: M. Alonso-Pena[1], J. Ameijeiras-Alonso[2] and R.M. Crujeiras[1].
AFFILIATIONS:
[1] Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela.
[2] Department of Mathematics, KU Leuven.
JOURNAL: Journal of Statistical Computation and Simulation.
YEAR: 2021.
STATUS: Published.
ISSN: 0094-9655.
PUBLISHER: Taylor & Francis.
LINK: https://doi.org/10.1080/00949655.2020.1818243.
JCR IMPACT FACTOR: 1.424 [Q3 in Statistics and Probability (67/125)].
CITESCORE: 1.9.
INFORMATION REGARDING COPYRIGHT AND USE: Information about copyright and reuses can be found at https://www.tandfonline.com/doi/abs/10.1080/00949655.2020.1818243?tab=permissions&scroll=top. For a thesis/dissertation reuse request, it holds:
Taylor & Francis is pleased to offer reuses of its content for a thesis or dissertation free of charge contingent on resubmission of permission request if work is published.
CONTRIBUTIONS OF THE PHD CANDIDATE: The candidate contributed to the conceptualization of the works, methodology, proofs of the theoretical results, software implementation, formal analysis of the results and to the manuscripts preparation.

Chapter 5: Alonso-Pena and Crujeiras (2022)

TITLE: Analyzing animal escape data with circular nonparametric multimodal regression.
AUTHORS: M. Alonso-Pena[1] and R.M. Crujeiras[1].
AFFILIATIONS:
[1] Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela.

JOURNAL: Annals of Applied Statistics.
YEAR: To appear.
STATUS: Accepted.
ISSN: 1932-6157.
PUBLISHER: Mattson Publishing Services/Institute of Mathematical Statistics.
JCR IMPACT FACTOR: 2.083 [Q2 in Statistics and Probability (43/125)].
CITESCORE: 3.9.
INFORMATION REGARDING COPYRIGHT AND USE: The following statement can be found in the copyright transfer agreement (see the Permissions section):
The author(s) reserve the following:

1. *All proprietary rights other than copyright, such as patent rights.*
2. *The rights to use all or part of this work (with appropriate reference to the journal and article) in future works of their own, such as lectures, press releases, reviews, textbooks, or reprint books.*
3. *The right to place the final version of this article (exactly as published in the journal) on their own homepage or in a public digital repository, provided there is a link to the official journal site.*
4. *The right to provide a copy of the final peer-reviewed manuscript to the NIH upon acceptance for publication, for public archiving in PubMed Central as soon as possible but no later than 12 months after publication.*

CONTRIBUTIONS OF THE PHD CANDIDATE: The candidate contributed to the conceptualization of the works, methodology, proofs of the theoretical results, software implementation, formal analysis of the results and to the manuscripts preparation.

Unpublished works

The following works, corresponding to Chapters 3 and 4 of the manuscript, are not published but are being considered for publication by JCR journals.

Chapter 3: Alonso-Pena et al. (2022b)

TITLE: A general framework for circular local likelihood regression.
AUTHORS: M. Alonso-Pena[1], I. Gijbels[2] and R.M. Crujeiras[1].
AFFILIATIONS:

[1] Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela.
[2] Department of Mathematics, KU Leuven.

CONTRIBUTIONS OF THE PHD CANDIDATE: The candidate contributed to the conceptualization of the works, methodology, proofs of the theoretical results, software implementation, formal analysis of the results and to the manuscripts preparation.

Chapter 4: Alonso-Pena et al. (2022a)

TITLE: Flexible joint modeling of mean and dispersion for the directional tuning of neuronal spike counts.
AUTHORS: M. Alonso-Pena[1], I. Gijbels[2] and R.M. Crujeiras[1].
AFFILIATIONS:

[1] Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela.

[2] Department of Mathematics and Leuven Statistics Research Center (LStat), KU Leuven.

CONTRIBUTIONS OF THE PHD CANDIDATE: The candidate contributed to the conceptualization of the works, methodology, proofs of the theoretical results, software implementation, formal analysis of the results and to the manuscripts preparation.

Permissions

In this section, the copyright transfer agreements for the works Alonso-Pena (2020) and Alonso-Pena and Crujeiras (2022) are respectively given.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Copyright Transfer Statement BEIO

Title of article:

An introduction to nonparametric multimodal regression

Author(s):

María Alonso-Pena

Submitting a manuscript to BEIO means that the author transfers the copyright for this contribution to Sociedad de Estadística e Investigación Operativa (SEIO). Therefore, the author agrees to the following:

- ✓ The author warrants the originality and authorship of the work.
- ✓ The copyright of this article is transferred to SEIO. The author signs for and accepts responsability for releasing this material on behalf of any and all co-authors. The copyright transfer covers the exclusive right to reproduce and distribute the article, including reprints, translations, photographic reproductions, microform, electronic form (offline, online) or any other reproductions of similar nature.
- ✓ The author retains the right to use the substance of the above work in future works, including lectures, press releases and reviews, provided that you acknowledge its prior publication in the journal.
- ✓ The author authorises to SEIO to include his/her article in catalogues and data bases belonging to institutions and organisations which have permission to include material from the journal BEIO.
- ✓ BEIO shall prepare the article for publication. We reserve the right to make such editorial changes as may be necessary to make the article suitable for publication. In any case, we reserve the right not to proceed with publication for whatever reason.

Date: 29/03/2020

Author's signature:





COPYRIGHT TRANSFER AGREEMENT

INSTITUTE OF MATHEMATICAL STATISTICS
BRAZILIAN STATISTICAL ASSOCIATION

ASSOCIATION PUBLICATIONS DE L'INSTITUT HENRI POINCARÉ
BERNOULLI SOCIETY FOR MATHEMATICAL STATISTICS & PROBABILITY

This signed Agreement must be received before the manuscript can be scheduled for publication. Please fill in and sign as indicated. Fax, email or mail to: Mattson Publishing Services, 2002 Holly Neck Road, Baltimore, MD 21221 USA
Email: assistant@mattsonpublishing.com, Fax: 410-391-2542

Transmittal #: AOAS1619

Author(s): María Alonso-Pena and Rosa M. Crujeiras

Title: Analyzing animal escape data with circular nonparametric multimodal regression

Copyright of the above work is hereby transferred to (please check the box next to the journal in which your article will appear).

Institute of Mathematical Statistics
- ☐ Annals of Applied Probability
- ☐ Annals of Probability
- ☑ Annals of Applied Statistics
- ☐ Annals of Statistics
- ☐ Statistical Science

Association Publications de l'Institute Henri Poincaré
- ☐ Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques

Bernoulli Society for Mathematical Statistics & Probability
- ☐ Bernoulli

Brazilian Statistical Association
- ☐ Brazilian Journal of Probability and Statistics

However, the author(s) reserve the following:

1. All proprietary rights other than copyright, such as patent rights.
2. The rights to use all or part of this work (with appropriate reference to the journal and article) in future works of their own, such as lectures, press releases, reviews, textbooks, or reprint books.
3. The right to place the final version of this article (exactly as published in the journal) on their own homepage or in a public digital repository, provided there is a link to the official journal site.
4. The right to provide a copy of the final peer-reviewed manuscript to the NIH upon acceptance for publication, for public archiving in PubMed Central as soon as possible but no later than 12 months after publication.

All authors must sign this form. This form may be duplicated for signing by co-authors.
In the case of a "work made for hire," a representative of the employer may send an equivalent form.

ALONSO PENA MARIA - 77420769D

Signature

María Alonso-Pena

Name

19/04/2022

Date

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

References

Aliyari Ghassabeh, Y. (2013). A sufficient condition for the convergence of the mean shift algorithm with Gaussian kernel. *Pattern Recognition Letters*, 34:1423–1427.

Aliyari Ghassabeh, Y. (2015). On the convergence of the mean shift algorithm in the one-dimensional space. *Journal of Multivariate Analysis*, 135:1–10.

Aliyari Ghassabeh, Y., Linder, T., and Takahara, G. (2013). On some convergence properties of the subspace constrained mean shift. *Pattern Recognition*, 46:3140–3147.

Alonso-Pena, M. (2020). An introduction to nonparametric multimodal regression. *Boletín de Estadística e Investigación Operativa (BEIO)*, 36:5–23.

Alonso-Pena, M., Ameijeiras-Alonso, J., and Crujeiras, R. M. (2021). Nonparametric tests for circular regression. *Journal of Statistical Computation and Simulation*, 19:477–500.

Alonso-Pena, M. and Crujeiras, R. M. (2022). Analizing animal escape data with circular nonparametric multimodal regression. *Annals of Applied Statistics (To appear)*.

Alonso-Pena, M., Gijbels, I., and Crujeiras, R. M. (2022a). Flexible joint modeling of mean and dispersion for the directional tuning of neuronal spike counts. *Submitted*.

Alonso-Pena, M., Gijbels, I., and Crujeiras, R. M. (2022b). A general framework for circular local likelihood regression. *Submitted*.

Ameijeiras-Alonso, J., Lagona, F., Ranalli, M., and Crujeiras, R. M. (2019). A circular non-homogeneous hidden Markov field for the spatial segmentation of wild fire occurrences. *Environmetrics*, 30:e2501.

Anderson-Cook, C. M. (1999). A tutorial on one-way analysis of circular-linear data. *Journal of Quality Technology*, 31:109–119.

Antoniadis, A., , Gijbels, I., Lambert-Lacroix, S., and Poggi, J.-M. (2016). Joint estimation and variable selection for mean and dispersion in proper dispersion models. *Electronic Journal of Statistics*, 10:1630–1676.

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

Bartels, R. (1984). Estimation in a bidirectional mixture of von Mises distributions. *Biometrics*, 40:777–784.

Batschelet, E. (1981). *Circular Statistics in Biology.* Academic Press.

Born, R. and Bradley, D. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*, 28:157–189.

Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations*. Oxford University Press.

Butts, D. A. and Goldman, M. S. (2006). Tuning curves, neuronal variability, and sensory coding. *PLOS Biology*, 4:e92.

Carandini, M. and Ferster, D. (2000). Membrane potential and firing rate in cat primary visual cortex. *Journal of Neuroscience*, 20:470–484.

Casa, A., Chacón, J. E., and Menardi, G. (2020). Modal clustering asymptotics with applications to bandwidth selection. *Electronic Journal of Statistics*, 14:835–856.

Cattaneo, M. D., Chandak, R., Jansson, M., and Ma, X. (2022). Boundary adaptive local polynomial conditional density estimators. *arXiv e-prints*, page arXiv:2204.10359.

Chacón, J. E. and Duong, T. (2013). Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electronic Journal of Statistics*, 7:499–532.

Chacón, J. E., Duong, T., and Wand, M. (2011). Asymptotics for general multivariate kernel density derivative estimators. *Statistica Sinica*, 21:807–840.

Chatla, S. B. and Shmueli, G. (2018). Efficient estimation of COM–Poisson regression and a generalized additive model. *Computational Statistics & Data Analysis*, 121:71–88.

Chaudhuri, P. (1991a). Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis*, 39:246–269.

Chaudhuri, P. (1991b). Nonparametric estimates of regression quantiles and their local bahadur representation. *The Annals of Statistics*, 19:760–777.

Chaudhuri, P. and Marron, J. S. (1999). Sizer for exploration of structures in curves. *Journal of the American Statistical Association*, 94:807–823.

Chen, Y.-C. (2018). Modal regression using kernel density estimation: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10:e1431.

Chen, Y.-C., Genovese, C. R., , and Wasserman, L. (2014). Generalized mode and ridge estimation. *arXiv e-prints*, page arXiv:1406.1803.

Chen, Y.-C., Genovese, C. R., Tibshirani, R. J., and Wasserman, L. (2016a). Nonparametric modal regression. *The Annals of Statistics*, 44:489–514.

Chen, Y.-C., Genovese, C. R., and Wasserman, L. (2016b). A comprehensive approach to mode clustering. *Electronic Journal of Statistics*, 10:210–241.

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:790–799.

Colwill, R. M. and Creton, R. (2011). Imaging escape and avoidance behavior in zebrafish larvae. *Reviews in the neurosciences*, 22:63–73.

Comaniciu, D. and Meer, P. (2002). Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619.

Conway, R. W. and Maxwell, W. L. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, 12:132–136.

Cronin, B., Stevenson, I. H., Sur, M., and Kording, K. P. (2010). Hierarchical Bayesian modeling and Markov chain Monte Carlo sampling for tuning-curve analysis. *Journal of Neurophysiology*, 103:591–602.

Croux, C., Gijbels, I., and Prosdocimi, I. (2012). Robust estimation of mean and dispersion functions in extended generalized additive models. *Biometrics*, 68:31–44.

Dey, D. K., Gelfand, A. E., and Peng, F. (1997). Overdispersed generalized linear models. *Journal of Statistical Planning and Inference*, 64:93–107.

Di Marzio, M., Fensore, S., and Panzera, A. (2018). Nonparametric classification for circular data. In Ley, C. and Verdebout, T., editors, *Applied Directional Statistics: Modern Methods and Case Studies*. Chapman & Hall/CRC.

Di Marzio, M., Fensore, S., Panzera, A., and Taylor, C. C. (2016). A note on nonparametric estimation of circular conditional densities. *Journal of Statistical Computation and Simulation*, 86:2573–2582.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statistics & Probability Letters*, 798:2066–2075.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2011). Kernel density estimation on the torus. *Journal of Statistical Planning and Inference*, 141:2156–2173.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2013). Non-parametric regression for circular responses. *Scandinavian Journal of Statistics*, 40:238–255.

Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.

USC DE SANTIAGO DE COMPOSTELA

Edelsbrunner, H., Kirkpatrick, D. G., and Seidel, R. (1983). On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, 29:551–559.

Edelsbrunner, H. and Mucke, E. P. (1984). Three-dimensional alpha shapes. *ACM Transactions on Graphics*, 13:43–72.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81:709–721.

Efthimiou, C. and Frye, C. (2014). *Spherical Harmonics In p Dimensions.* World Scientific.

Einbeck, J. and Tutz, G. (2006). Modelling beyond regression functions: An application of multimodal regression to speedflow data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55:461–475.

Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, 33:1380–1403.

Ermentrout, G. B., Galán, R. F., and Urban, N. N. (2008). Reliability, synchrony and noise. *Trends in Neurosciences*, 31:428–434.

Fan, J., Farmen, M., and Gijbels, I. (1998). Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60:591–608.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 57:371–394.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman and Hall.

Fan, J., Heckman, N. E., and Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasilikelihood functions. *Journal of the American Statistical Association*, 90:141–150.

Fan, J., Yao, Q., and Howell, T. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, 83:189–206.

Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge University Press.

Fisher, N. I. and Lee, A. J. (1992). Regression models for angular responses. *Biometrics*, 48:645–677.

Fukunaga, K. and Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21:32–40.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

García-Portugués, E. (2014). *Nonparametric Inference with Directional and Linear Data*. PhD thesis, Universidade de Santiago de Compostela.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013). Kernel density estimation for directional-linear data. *Journal of Multivariate Analysis*, 121:152–275.

García-Portugués, E., Van Keilegom, I., Crujeiras, R. M., and González-Manteiga, W. (2016). Testing parametric models in linear-directional regression. *Scandinavian Journal of Statistics*, 43:1178–1191.

Gasser, T., Sroka, L., and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, 73:625–633.

Gatto, R. and Jammalamadaka, S. R. (2007). The generalized von Mises distribution. *Statistical Methodology*, 4:341–353.

Gaunt, R. E., Iyengar, S., Olde Daalhuis, A. B., and Simsek, B. (2019). An asymptotic expansion for the normalizing constant of the Conway–Maxwell–Poisson distribution. *Annals of the Institute of Statistical Mathematics*, 71:163–180.

Gelfand, A. E. and Dalal, S. R. (1990). A note on overdispersed exponential families. *Biometrika*, 77:55–64.

Georgopoulos, A. P., Kalaska, J., Caminiti, R., and Massey, J. T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience*, 2:1527–1537.

Gijbels, I. and Prosdocimi, I. (2011). Smooth estimation of mean and dispersion function in extended generalized additive models with application to italian induced abortion data. *Journal of Applied Statistics*, 38:2391–2411.

Gijbels, I. and Prosdocimi, I. (2012). Flexible mean and dispersion function estimation in extended generalized additive models. *Communications in Statistics - Theory and Methods*, 41:3259–3277.

Gijbels, I., Prosdocimi, I., and Claeskens, G. (2010). Nonparametric estimation of mean and dispersion functions in extended generalized linear models. *TEST*, 19:580–608.

Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincaré*, 38:907–921.

USC
DE SANTIAGO
DE COMPOSTELA

Gould, A. L. (1969). A regression technique for angular variates. *Biometrics*, 25:683–700.

Gumbel, E. J., Greenwood, J. A., and Durand, D. (1953). The circular normal distribution: theory and tables. *Journal of the American Statististical Association*, 48:131–152.

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

Huang, M., Li, R., and Wang, S. (2013). Nonparametric mixture of regression models. *Journal of the American Statistical Association*, 108:929–941.

Hurvich, C. M., Simonoff, J. S., and Tsai, C. (1988). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60:271–293.

Hyndman, R. J., Bashtannyk, D. M., and Grunwald, G. K. (1996). Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, 5:315–336.

Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics.* World Scientific.

Jander, R. (1957). Die optische richtungsorientierung der roten waldameise (formicaruea l.). *Zeitschrift für vergleichende Physiologie*, 40:162–238.

Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, 73:602–606.

Jona-Lasinio, G., Gelfand, A., and Jona-Lasinio, M. (2012). Spatial analysis of wave direction data using wrapped Gaussian processes. *Annals of Applied Statistics*, 6:1478–1498.

Jones, M. C. and Pewsey, A. (2005). A family of symmetric distributions on the circle. *Journal of the American Statistical Association*, 100:1422–1428.

Kato, S. and Jones, M. C. (2015). A tractable and interpretable four-parameter family of unimodal distributions on the circle. *Biometrika*, 102:181–190.

Kaufman, C. G., Ventura, V., and Kass, R. E. (2005). Spline-based non-parametric regression for periodic functions and its application to directional tuning of neurons. *Statistics in Medicine*, 24:2255–2265.

Kim, S. and SenGupta, A. (2017). Multivariate-multiple circular regression. *Journal of Statistical Computation and Simulation*, 87:1277–1291.

Kobayashi, T. and Otsu, N. (2010). Von Mises-Fisher mean shift for clustering on a hypersphere. In *Proceedings of the 20th international conference on pattern recognition*, pages 2130–2133. IEEE.

USC DE SANTIAGO DE COMPOSTELA

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46:33–50.

Kohn, A. and Movshon, J. A. (2003). Neuronal adaptation to visual motion in area MT of the macaque. *Neuron*, 39:681–691.

Lee, M. J. (1989). Mode regression. *Journal of Economics*, 42:337–349.

Lee, M. J. (1993). Quadratic mode regression. *Journal of Economics*, 57:1–19.

Lee, Y. and Nelder, J. A. (2000). The relationship between double-exponential families and extended quasilikelihood families, with application to modelling Geissler's human sex ratio data. *Journal of Applied Statistics*, 49:413–419.

Ley, C. and Verdebout, T. (2017). *Modern Directional Statistics*. CRC Press.

Li, X., Hu, Z., and Wu, F. (2007). A note on the convergence of the mean shift. *Pattern Recognition*, 40:1756–1762.

Lombardo, J. A., Macellaio, M. V., Liu, B., Palmer, S. E., and Osborne, L. C. (2018). State dependence of stimulus-induced variability tuning in macaque MT. *PLOS Computational Biology*, 14:e1006527.

Marchetti, G. and Scapini, F. (2003). Use of multiple regression models in the study of sandhopper orientation under natural conditions. *Estuarine, Coastal and Shelf Science*, 58:207–215.

Mardia, K. (1972). *Statistics of Directional Data.* Academic Press.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics.* Wiley.

Mardia, K. V. and Sutton, T. W. (1978). A model for cylindrical variables with applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 40:229–233.

Meilán-Vila, A., Francisco-Fernández, M., and Crujeiras, R. M. (2022). Goodness-of-fit tests for multiple regression with circular response. *Journal of Statistical Computation and Simulation*, 92:1941–1963.

Meilán-Vila, A., Francisco-Fernández, M., Crujeiras, R. M., and Panzera, A. (2021). Non-parametric multiple regression estimation for circular response. *TEST*, 30:650–672.

Montgomery-Smith, S. J. (1993). Comparison of sums of independent indentically distributed random vectors. *Probability and Mathematical Statistics*, 14:281–285.

Mooney, J. A., Helms, P. J., and Jollife, I. T. (2003). Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome. *Computational Statistics & Data Analysis*, 41:505–513.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, 9:141–142.

Nair, A., Changsing, K., Stewart, W. J., and McHenry, M. J. (2017a). Data from: Fish prey change strategy with the direction of a threat. https://doi.org/10.5061/dryad.47mq9.

Nair, A., Changsing, K., Stewart, W. J., and McHenry, M. J. (2017b). Fish prey change strategy with the direction of a threat. *Proceedings of the Royal Society B: Biological Sciences*, 284:20170393.

Oba, S., Kato, K., and Ishii, S. (2005). Multi-scale clustering for gene expression profiling data. In *5th IEEE Symposium on Bioinformatics and Bioengineering (BIBE'05)*, pages 210–217. IEEE.

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

Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2014a). CircSiZer: an exploratory tool for circular data. *Environmental and Ecological Statistics*, 21:143–159.

Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2014b). NPCirc : An R package for nonparametric circular methods. *Journal of Statistical Software*, 61:1–26.

Pardi, L. and Ercolini, A. (1986). Some angular-linear distributions and related regression models. *Bollettino di Zoologia Italiana*, 53:139–160.

Pateiro-Lopez, B. and Rodriguez-Casal, A. (2022). *alphahull: Generalization of the Convex Hull of a Sample of Points in the Plane*. R package version 2.4.

Patrangenaru, V. and Ellingson, L. (2016). *Nonparametric Statistics on Manifolds and their Applications to Object Data Analysis*. CRC Press.

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

Presnell, B., Morrison, S. P., and Little, R. C. (1998). Multivariate linear models for directional data. *Journal of the American Statistical Association*, 93:1068–1077.

Puglisi, G., Leonetti, A., Landau, A., Fornia, L., Cerri, G., and Borroni, P. (2017). The role of attention in human motor resonance. *PLoS One*, 12:e0177457.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ringach, D. L., Hawken, M. J., and Shapley, R. (2003). Dynamics of orientation tuning in macaque V1: the role of global and tuned suppression. *Journal of Neurophysiology*, 90:342–352.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90:1257–1270.

Saavedra-Nieves, P. and Crujeiras, R. M. (2021a). *HDiR: Directional Highest Density Regions*. R package version 1.1.1.

Saavedra-Nieves, P. and Crujeiras, R. M. (2021b). Nonparametric estimation of directional highest density regions. *Advances in Data Analysis and Classification (to appear)*.

Sager, T. W. and Thisted, R. A. (1982). Maximum likelihood estimation of isotonic modal regression. *The Annals of Statistics*, 10:690–707.

Sato, N., Shidara, H., and Ogawa, H. (2019). Trade-off between motor performance and behavioural flexibility in the action selection of cricket escape behaviour. *Scientific Reports*, 9:18112.

Scapini, F., Aloia, A., Bouslama, M., Chelazzi, L., Colombini, I., ElGtari, M., Fallaci, M., and Marchetti, G. M. (2002). Multiple regression analysis of the sources of variation in orientation of two sympatric sandhoppers, talitrus saltator and talorchestia brito, from an exposed mediterranean beach. *Behavioral Ecology*, 52:403–414.

Scott, D. W. (1992). *Multivariate Density Estimation*. Wiley.

Sellers, K. F. and Shmueli, G. (2010). A flexible regression model for count data. *Annals of Applied Statistics*, 4:943–961.

Sellers, K. F. and Shmueli, G. (2013). Data dispersion: now you see it... now you don't. *Communications in Statistics - Theory and Methods*, 42:3134–3147.

SenGupta, S. and Rao, J. S. (1966). Statistical analysis of cross-bedding azimuths from the kamthi formation around bheemaram, pranhita: Godavari valley. *Sankhya*, 28:165–174.

Shmueli, G., Minka, T., Kadane, J., Borle, S., and Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the Conway- Maxwell-Poisson distribution. *Journal of Applied Statistics*, 54:127–142.

Small, C. G. (1996). *The Statistical Theory of Shape*. Springer.

Spurr, B. D. (2015). On estimating the parameters in mixtures of circular normal distributions. *Mathematical Geology*, 13:163–173.

Staniswalis, J. G. (1989). The kernel estimate of a regression function in likelihood-based models. *Journal of the American Statistical Association*, 84:276–283.

Stein, R. B., Gossen, E. R., and Jones, K. E. (2005). Neuronal variability: noise or part of the signal? *Nature Reviews Neuroscience*, 6:389–397.

Stevenson, I. H. (2016). Flexible models for spike count data with both over- and under-dispersion. *Journal of Computational Neuroscience*, 41:29–43.

USC UNIVERSIDADE DE SANTIAGO

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, 82:559–567.

Von Mises, R. (1918). Uber die 'ganzzahligkeit' der atomgewichte und verwandte fragen. *Physikalische Zeitschrift*, 19:490–500.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman & Hall.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya, series A*, 26:359–372.

Wikimedia Commons (2014a). Larval zebrafish, from author: Csiro. https://commons.wikimedia.org/wiki/File:CSIRO_ScienceImage_7598_larval_zebra.jpg. [Online; accessed September 29, 2021].

Wikimedia Commons (2014b). Zebrafish, from author: Database center for life science (dbcls). https://commons.wikimedia.org/wiki/File:202101_Zebrafish.svg. [Online; accessed January 18, 2022].

Wikimedia Commons (2006). Different areas of cortex, from author: Juan. https://commons.wikimedia.org/wiki/File:Ence,_Pontevedra,_Galiza.jpg. [Online; accessed January 17, 2022].

Wikimedia Commons (2010). Gryllus bimaculatus. Female. Gryllidae, from author: gailhampshire. https://commons.wikimedia.org/wiki/File:Gryllus_bimaculatus._Female._Gryllidae_-_Flickr_-_gailhampshire.jpg. [Online; accessed January 18, 2022].

Wikimedia Commons (2013). Transmission, from author: Ford motor company. https://commons.wikimedia.org/wiki/File:Ford_model_t_1919_d029_transmission.png. [Online; accessed June 7, 2019].

Wikimedia Commons (2014). Ence, Pontevedra, Galiza, vista dende Lourido, San Salvador de Poio, Poio, provincia de Pontevedra, Galiza., from author: Drking1234. https://commons.wikimedia.org/wiki/File:Cortex_functional_areas.jpg. [Online; accessed January 17, 2022].

Wikimedia Commons (2017). Creepy crawlies beach closeups, from author: Magnus hagdorn. https://commons.wikimedia.org/wiki/File:Going_-_sandhopper_(Talitrus_saltator)_(8666909448).jpg. [Online; accessed June 7, 2019].

Xiang, S. Yao, W. (2018). Semiparametric mixtures of nonparametric regressions. *Annals of the Institute of Statistical Mathematics*, 70:131–154.

Yao, W. and Li, L. (2014). A new regression model: modal linear regression. *Scandinavian Journal of Statistics*, 41:656–671.

Yao, W., Lindsay, B. G., and Li, R. (2012). Local modal regression. *Journal of Nonparametric Statistics*, 24:647–663.

Young, S. and Bowman, A. W. (1995). Nonparametric analysis of covariance. *Biometrics*, 51:920–931.

Zhang, Y. and Chen, Y.-C. (2021a). Kernel smoothing, mean shift, and their learning theory with directional data. *Journal of Machine Learning Research (to appear)*.

Zhang, Y. and Chen, Y.-C. (2021b). Mode and ridge estimation in euclidean and directional product spaces: a mean shift approach. *arXiv e-prints*, page arXiv:2110.08505.

Zhang, Y. and Chen, Y.-C. (2021c). The EM perspective of directional mean shift algorithm. *arXiv e-prints*, page arXiv:2101.10058.

Zhou, H. and Huang, X. (2016). Nonparametric modal regression in the presence of measurement error. *Electronic Journal of Statistics*, 10:3579–3620.

Zhou, H. and Huang, X. (2019). Bandwidth selection for nonparametric modal regression. *Communications in Statistics - Simulation and Computation*, 48:968–984.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA





Nonparametric regression models are employed to examine the dependence between two or more random variables, without assuming a specific form for the regression function. However, complex data structures often arise in practice, leading to situations where the support of the variables is not Euclidean. This is the case of circular variables, defined on the unit circumference. Classical nonparametric regression methods do not take into account the periodicity of the data, and thus are not adequate for this kind of observations. This thesis provides new nonparametric regression models and inference tools to deal with circular variables. The performance of the proposed methodologies is analyzed and illustrated with real data applications.