



FACULTADE DE MATEMÁTICAS

Traballo Fin de Grao

# Especificación de Modelos en el Análisis de Supervivencia

Ángel Antonio Pérez Balo

2024-2025

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



GRAO DE MATEMÁTICAS

Traballo Fin de Grao

# Especificación de Modelos en el Análisis de Supervivencia

Ángel Antonio Pérez Balo

Junio, 2025

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



# Trabajo propuesto

|   |
|---|
| <b>Área de Coñecemento:</b> Estadística e Investigación Operativa   |
| <b>Título:</b> Especificación de Modelos en el Análisis de Supervivencia  |
| <b>Breve descripción do contido</b>   |
| <p>En el contexto del “Análisis de Supervivencia” se trata de revisar algunos modelos de especificación para la función de distribución y para la función de densidad de una variable poblacional, con datos de partida que pueden ser susceptibles de censura aleatoria por la derecha, muy típica en el “Análisis de Supervivencia”. Un posible guión es el siguiente:</p> <ol style="list-style-type: none"><li>1) Una breve revisión de los contrastes de especificación para datos completos relativos a la distribución y función de densidad.</li><li>2) Modelos paramétricos y no paramétricos en el “Análisis de Supervivencia”. El fenómeno de la censura. El estimador de Kaplan-Meier.</li><li>3) Contrastes de especificación para la distribución en base al estimador de Kaplan-Meier.</li><li>4) Contrastes de especificación para la densidad en base al estimador de Kaplan Meier.</li><li>5) Aplicaciones a datos reales</li></ol> |
| <b>Recomendacións</b>   |
|   |
| <b>Outras observacións</b>  |
|   |



# Índice

|   |           |
|---|-----------|
| <b>Resumen</b>  | <b>VI</b> |
| <b>Introducción</b>   | <b>IX</b> |
| <b>1. Contrastes de especificación con datos completos</b>  | <b>1</b>  |
| 1.1. Introducción . . . . .   | 1         |
| 1.2. Elementos de un contraste . . . . .  | 2         |
| 1.3. Contrastes basados en la función de distribución . . . . .                                   | 5         |
| 1.4. Contrastes basados en la función de densidad . . . . .                                       | 14        |
| 1.4.1. Estimación no paramétrica de la función de densidad . . . . .                              | 15        |
| 1.4.2. Contrastes . . . . .   | 18        |
| 1.4.3. Comparación con los contrastes basados en la función de distribución . . . . .             | 20        |
| <b>2. Introducción al Análisis de Supervivencia</b>   | <b>23</b> |
| 2.1. Introducción . . . . .   | 23        |
| 2.2. Funciones de interés . . . . .   | 23        |
| 2.3. Modelos paramétricos notables . . . . .  | 26        |
| 2.4. Censura . . . . .  | 29        |
| <b>3. Contrastes de especificación con datos censurados basados en la función de distribución</b> | <b>31</b> |
| 3.1. Introducción . . . . .   | 31        |

|   |           |
|---|-----------|
| 3.2. Estimación no paramétrica de la distribución bajo censura aleatoria por la derecha   | 32        |
| 3.3. Estimación paramétrica de la distribución bajo censura aleatoria por la derecha .  | 33        |
| 3.4. Contrastes basados en la función de distribución bajo censura aleatoria por la derecha . . . . .                           | 34        |
| <b>4. Contrastes de especificación con datos censurados basados en la función de densidad</b>                                   | <b>37</b> |
| 4.1. Introducción . . . . .   | 37        |
| 4.2. Estimación de la densidad bajo censura aleatoria por la derecha . . . . .  | 37        |
| 4.3. Contrastes basados en la función de densidad bajo censura aleatoria por la derecha   | 38        |
| 4.4. Estimación continua de la distribución . . . . .   | 40        |
| 4.5. Contrastes sobre la función de distribución usando una estimación continua bajo censura aleatoria por la derecha . . . . . | 42        |
| <b>5. Ilustración por simulación</b>  | <b>45</b> |
| <b>I. Notación y código</b>   | <b>51</b> |
| I.1. Notación . . . . .   | 51        |
| I.2. Código del ejemplo 1.19 . . . . .  | 51        |
| I.3. Código del ejemplo 1.28 . . . . .  | 58        |
| I.4. Código del ejemplo 5.1 . . . . .   | 62        |
| I.5. Código del ejemplo 5.2 . . . . .   | 73        |
| <b>Bibliografía</b>   | <b>79</b> |

## Resumen

Este trabajo hace un repaso de algunos contrastes de especificación basados en la función de distribución y en la función de densidad para muestras de datos completos y muestras de datos con censura aleatoria por la derecha. Se explica lo que es un contraste de especificación y cuáles son sus elementos. Se propone un estimador no paramétrico de la función de distribución para muestras de datos completos, se estudian sus propiedades y se plantean contrastes sobre la distribución utilizando distintas distancias. Se comenta la estimación no paramétrica de la función de densidad para muestras de datos completos y algunos contrastes sobre la misma. Se introduce el campo del Análisis de la Supervivencia, parándose en las funciones más importantes para estudiar tiempos de vida y en los modelos paramétricos más comunes; así como el fenómeno de la censura, haciendo hincapié en la censura aleatoria por la derecha. Se propone una estimación no paramétrica de la función de distribución para el caso de datos con censura aleatoria por la derecha y se comenta cómo se puede obtener una estimación paramétrica; además, se sugiere una adaptación de los contrastes vistos para datos completos. Se proporciona una estimación no paramétrica de la función de densidad para datos con censura aleatoria por la derecha y un contraste basado en ella. Se estudia una estimación no paramétrica y continua de la función de distribución para datos con censura aleatoria por la derecha y un contraste utilizándola.

## Abstract

This project goes over some specification contrasts based on the distribution function and the density function for both complete data samples and samples with random right censoring. Specification contrasts are explained, along with their elements. A nonparametric estimator for the distribution function is proposed in case of a complete sample, its properties are studied and contrasts are laid out using different distances. Nonparametric estimation of the density function for complete samples is discussed, as well as some contrasts for it. Survival analysis

is presented, focusing on some of the most important functions for studying lifetime data and frequently used parametric families; censoring, and especially random right censoring, are also introduced. A nonparametric estimator for the distribution function under random right censoring is proposed and parametric estimation is discussed; contrasts for complete samples are adapted to this scenario. A nonparametric estimator for the density function under random right censoring is given, together with a contrast based on it. A nonparametric continuous estimator for the distribution function is studied along with a contrast using it.

# Introducción

Para llevar a cabo un estudio estadístico de una población es muy conveniente conocer de qué distribución procede la variable aleatoria de interés y si esta distribución pertenece a una familia conocida, pues en ese caso resultará más fácil el estudio al disponer de más resultados. Existen distintas funciones que caracterizan la distribución de una variable aleatoria real unidimensional, como la función de distribución, la función de densidad, la función característica o la función cuantil. Para estudiar si una muestra procede de cierta distribución conocida, se puede comparar alguna de estas funciones asociada a la distribución elegida con la estimación de la misma obtenida a partir de la muestra. Este trabajo se centra en la metodología basada en la función de distribución y de densidad.

Además, pueden darse circunstancias en las que la muestra de la que se dispone no está formada únicamente por observaciones exactas de la variable de interés, pero aun así aportan información que puede ser útil. Por ejemplo, en ensayos clínicos en los que interesa medir el tiempo que sobrevive una persona con cierta enfermedad o el tiempo que tarda en curarse, puede haber pacientes que abandonen el ensayo antes de alcanzar dicho evento de interés por diversos motivos que no están directamente relacionados con la enfermedad estudiada, por lo que los investigadores sólo saben que el evento de interés habría ocurrido después del momento en que le perdieron la pista al paciente, sin poder precisar más. A este fenómeno se le llama censura y se habla de muestras de datos censurados cuando algunas observaciones no se corresponden con tiempos de vida sino con tiempos de censura. Este tipo de muestras requieren un tratamiento especial, pero también permiten hacer estimaciones de las funciones de interés que se pueden comparar con las de una distribución conocida. En oposición a las muestras de datos censurados, se hará referencia a las muestras en las que todas las observaciones proceden de la variable de interés como muestras de datos completos.



# Capítulo 1

## Contrastes de especificación con datos completos

### 1.1. Introducción

Los contrastes de hipótesis son técnicas de inferencia estadística que sirven para determinar la validez de cierta afirmación sobre la distribución de una población. A estas afirmaciones se les llama hipótesis. Para llevar a cabo un contraste se necesitan dos hipótesis:

- La hipótesis nula, denotada por  $H_0$ , es la afirmación que se supone cierta en un principio.
- La hipótesis alternativa,  $H_1$ , es la que se quiere contrastar contra la nula.

El objetivo de un contraste es decidir si hay evidencias estadísticas suficientes como para rechazar la hipótesis nula en base a una muestra, que en este trabajo se supone formada por observaciones independientes e idénticamente distribuidas (i.i.d. a partir de ahora).

Sea  $X$  una variable aleatoria real unidimensional y sea  $F$  su función de distribución. Dentro de los contrastes de hipótesis sobre la función de distribución, se distinguen los paramétricos y los no paramétricos:

- En los paramétricos se asume que  $F$  pertenece a una familia paramétrica de distribuciones,  $F \in \{F_\theta, \theta \in \Theta\}$ , donde  $\Theta \subset \mathbb{R}^n$  es el conjunto de valores posibles del parámetro, y se plantea un contraste sobre el parámetro:

$$\begin{cases} H_0 : \theta \in \Theta_0, \\ H_1 : \theta \notin \Theta_0, \end{cases}$$

donde  $\Theta_0 \subsetneq \Theta$  es un subconjunto de los posibles valores del parámetro para el que se quiere realizar el contraste.

- En los no paramétricos, también conocidos como contrastes de especificación, el objetivo es estudiar hipótesis estructurales sobre la distribución de  $X$ , sin asumir ninguna familia paramétrica. Tienen la siguiente forma:

$$\begin{cases} H_0 : F \in \mathcal{F}, \\ H_1 : F \notin \mathcal{F}, \end{cases}$$

donde  $\mathcal{F}$  es un conjunto de funciones de distribución. En lo que sigue se estudiarán contrastes de este tipo.

En ambos escenarios las hipótesis pueden ser simples o compuestas. En el contexto de los contrastes de especificación, una hipótesis nula simple proporciona una función concreta  $F = F_0$  y se dice que está completamente especificada (por ejemplo, normal estándar). Por otro lado, una hipótesis compuesta proporciona un conjunto de funciones posibles,  $F \in \mathcal{F}$  y se dice que aporta una especificación incompleta (por ejemplo, cuando afirma que  $F$  es normal sin decir la media ni la varianza).

Habitualmente, la hipótesis nula propone un modelo más sencillo que la alternativa y el contraste consiste en estudiar si hay pruebas significativas de que el modelo bajo la nula no es adecuado. Si no las hay, no es necesario recurrir a un modelo más complejo.

## 1.2. Elementos de un contraste

Se llama estadístico de contraste a un valor numérico que se calcula a partir de la muestra y que se utiliza para tomar una decisión sobre el contraste en base a un criterio. También depende de la información proporcionada por la hipótesis nula y su distribución bajo la suposición de que  $H_0$  es cierta puede ser conocida o estimable. Una vez decidido el estadístico, se clasifican sus posibles valores en dos regiones complementarias dependiendo de la magnitud de esta discrepancia:

- La región crítica, o región de rechazo, es el conjunto de los valores del estadístico de contraste que manifiestan gran discrepancia entre los datos y la hipótesis nula. Si el valor del estadístico evaluado sobre la muestra cae en la región crítica, se rechaza la hipótesis nula y se adopta la alternativa.
- La región de aceptación, o de no rechazo, es el conjunto de valores del estadístico de contraste para los que no hay mucha discrepancia entre la hipótesis nula y los datos. Si el

valor del estadístico evaluado sobre la muestra cae en la región de aceptación, no se rechaza la hipótesis nula.

La principal complejidad en el planteamiento de un contraste está precisamente en la determinación de estas regiones, lo que se hace tratando de controlar los errores a la hora de tomar la decisión de rechazar o no rechazar la hipótesis nula. Estos errores pueden producirse de dos maneras: rechazando  $H_0$  cuando es cierta (error de tipo I) o no rechazándola cuando es falsa (error de tipo II).

**Definición 1.1.** [17, Sección 3.3] Se llama nivel de significación del contraste a

$$\alpha = \mathbb{P}(\text{Error de tipo I}) = \mathbb{P}(\text{Rechazar } H_0 \mid H_0 \text{ es cierta})$$

**Definición 1.2.** [17, Sección 3.3] Se denota

$$\beta = \mathbb{P}(\text{Error de tipo II}) = \mathbb{P}(\text{No rechazar } H_0 \mid H_0 \text{ es falsa})$$

y se llama potencia del contraste a

$$1 - \beta = \mathbb{P}(\text{Rechazar } H_0 \mid H_0 \text{ es falsa}).$$

**Definición 1.3.** [17, Sección 3.3] Se dice que un contraste es consistente si

$$1 - \beta \xrightarrow{n \rightarrow \infty} 1.$$

Lo ideal sería encontrar un contraste que minimizase simultáneamente las probabilidades de cometer ambos errores. Sin embargo, esto no suele ser posible porque al reducir la probabilidad de cometer uno aumenta la del otro. Normalmente se prioriza tener controlado el nivel de significación, fijando una cota superior de antemano. Se suele usar 0.05 o 0.01, pero puede ser otra dependiendo del contexto al que se aplique. Una vez hecho esto, se escoge el contraste que tenga mayor potencia bajo la hipótesis alternativa. Si un contraste es más potente que el resto para cualquier hipótesis alternativa, se dice que es el uniformemente más potente, aunque no siempre existe tal contraste.

Una vez fijado el nivel de significación máximo tolerable de acuerdo con el investigador, el valor del estadístico evaluado en la muestra caerá en la región de aceptación con probabilidad de por lo menos  $1 - \alpha$  cuando  $H_0$  es cierta. En ese caso no se rechaza la hipótesis nula para un nivel de significación  $\alpha$  y se dice que el contraste no es significativo. Por otro lado, caerá en la región crítica con probabilidad de como mucho  $\alpha$  cuando  $H_0$  es cierta. Ahora sí que hay pruebas significativas a favor de rechazar la hipótesis nula, el contraste es significativo.

A menudo, la decisión de rechazar o no rechazar la hipótesis nula no aporta suficiente información y es preferible indicar para qué valores de  $\alpha$  el contraste rechaza  $H_0$ . Si el contraste

rechaza  $H_0$  para un nivel  $\alpha$ , también lo hará para todo nivel  $\alpha'$  tal que  $\alpha' > \alpha$ , luego existe cierto nivel que es el más pequeño para el cual el contraste rechaza  $H_0$ , al que se llama p-valor. Por tanto, la hipótesis nula se rechaza si, y sólo si, el p-valor es menor que el nivel del contraste.

**Definición 1.4.** El p-valor es el ínfimo de los niveles  $\alpha \in (0, 1)$  para los que el estadístico de contraste evaluado sobre la muestra cae en la región crítica.

También se puede interpretar como la probabilidad, si  $H_0$  es cierta, de que el valor del estadístico represente una mayor discrepancia con la hipótesis nula que el valor obtenido al evaluarlo sobre la muestra.

El hecho de que el p-valor sea grande no significa que haya pruebas a favor de  $H_0$ . Si la distribución del estadístico de contraste bajo  $H_0$  es continua, el p-valor sigue una distribución *Uniforme* $[0, 1]$  cuando la hipótesis nula es cierta y se acumula cerca del 0 cuando es falsa, haciéndolo de forma más notable cuanto más elevada sea la potencia. De este modo, un p-valor elevado puede darse porque  $H_0$  sea cierta o porque el contraste tenga poca potencia. En definitiva, el p-valor no es la probabilidad de que  $H_0$  sea cierta.

Para poder calcular el p-valor de un contraste será necesario trabajar con la distribución del estadístico bajo la hipótesis nula, que no siempre es conocida. Pueden darse varias situaciones: Si es una distribución conocida, el p-valor se obtiene inmediatamente. En otro caso, es posible que el p-valor se pueda aproximar por simulación (Montecarlo) o que la distribución dependa de parámetros desconocidos, por lo que habría que estimar esos parámetros o utilizar técnicas de remuestreo como un bootstrap para aproximarla. Estos métodos se comentarán en la siguiente sección. Otro escenario posible es que la distribución sea desconocida pero se sepa la distribución asintótica (o que esta dependa de parámetros que se pueden estimar) y se pueda emplear para calcular los p-valores.

Para estudiar empíricamente el nivel y la potencia de un contraste se pueden llevar a cabo simulaciones. Este tipo de estudios consisten en generar  $M$  muestras aleatorias simples (esto es, muestras de observaciones i.i.d.) de una distribución dada y evaluar el estadístico de contraste en cada una para calcular la proporción de estos que caen en la región crítica. De esta forma se puede comprobar si un contraste es consistente viendo si al elegir distribuciones bajo la hipótesis nula la proporción de rechazos está cerca del nivel de significación o se va acercando según aumenta el tamaño muestral y si al elegir distribuciones bajo la hipótesis alternativa la proporción de rechazos se aproxima a 1 a medida que se incrementa el tamaño muestral. Este proceso se debe repetir para distintos niveles de significación y distribuciones tanto bajo la hipótesis nula como bajo la alternativa.

También es posible comparar el comportamiento de dos o más contrastes a través de una simulación. En ese caso, una vez comprobado el nivel de significación, se compara la potencia de

los contrastes bajo distintos modelos en la hipótesis alternativa. Es habitual que distintas formas de construir contrastes supongan ventajas frente a cierto tipo de alternativas y desventajas ante otros. Por tanto, no suele haber un contraste que sea uniformemente más potente que los demás bajo cualquier alternativa.

### 1.3. Contrastes basados en la función de distribución

En esta sección se presentan tres tipos de contrastes basados en la función de distribución. Se introducen en primer lugar para el caso de hipótesis nula simple y, a continuación, se explica cómo se pueden adaptar estos contrastes para una hipótesis nula compuesta.

Dada una muestra  $X_1, X_2, \dots, X_n$  de observaciones i.i.d. de una variable aleatoria real unidimensional cuya función de distribución  $F$  es continua y desconocida, se quiere contrastar:

$$\begin{cases} H_0 : F = F_0, \\ H_1 : F \neq F_0. \end{cases}$$

**Definición 1.5.** La función de distribución empírica se define como:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x) = \begin{cases} 0, & \text{si } x < X_{(1)}, \\ \frac{k}{n}, & \text{si } X_{(k)} \leq x < X_{(k+1)}, \\ 1, & \text{si } x \geq X_{(n)}, \end{cases}$$

donde  $I(\cdot)$  es la función indicadora, que vale 1 cuando la condición que aparece dentro es cierta y 0 en otro caso; y  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  es la muestra ordenada.

Los contrastes basados en la función de distribución generalmente se centran en medir la distancia entre un estimador de la función de distribución, que en este caso será la distribución empírica,  $F_n$ , y la función de distribución teórica dada en la hipótesis nula,  $F_0$ . Esta distancia servirá de estadístico de contraste, por lo que se rechazará la hipótesis nula para valores grandes de este. Establecer el umbral entre las regiones de rechazo y no rechazo es la dificultad técnica en la construcción de estos tests. Dependiendo de la distancia empleada, se obtendrán distintos estadísticos: los más conocidos son los de tipo Kolmogorov-Smirnov, Cramér-von Mises y Anderson-Darling. Los tres se basan en las propiedades de la distribución empírica como estimador.

**Definición 1.6.** [29, Sección 5.2] Se define el error cuadrático medio del estimador  $\hat{\theta}$  del parámetro  $\theta$  como

$$ECM(\hat{\theta}) = \mathbb{E} \left[ \left( \mathbb{E}[\hat{\theta}] - \theta \right)^2 \right].$$

**Proposición 1.7.** [29, Sección 5.3] El error cuadrático medio también se puede expresar como

$$ECM(\hat{\theta}) = \text{Var}(\hat{\theta}) + \left(\mathbb{E}[\hat{\theta}] - \theta\right)^2.$$

**Definición 1.8.** [29, Sección 5.3] Se dice que un estimador es consistente si

$$\hat{\theta} \xrightarrow{\mathbb{P}} \theta, \text{ esto es, } \mathbb{P}\left(|\hat{\theta} - \theta| > \epsilon\right) \rightarrow 0 \text{ para todo } \epsilon > 0.$$

*Observación 1.9.* [29, Sección 5.3] En particular, si  $ECM(\hat{\theta}) \rightarrow 0$  el estimador es consistente.

**Proposición 1.10.** [29, Sección 2.3] Fijado  $x \in \mathbb{R}$ , se tiene que:

- La distribución empírica es una variable aleatoria, pues depende de la muestra.
- $nF_n(x) \sim \text{Binomial}(n, F(x))$ , luego para  $k = 0, 1, 2, \dots, n$ ,

$$\mathbb{P}\left(F_n(x) = \frac{k}{n}\right) = \mathbb{P}(nF_n(x) = k) = \binom{n}{k} F(x)^k (1 - F(x))^{n-k}.$$

- $F_n(x)$  es un estimador insesgado de  $F(x)$ :

$$\mathbb{E}[F_n(x)] = \frac{1}{n} \mathbb{E}[nF_n(x)] = \frac{nF(x)}{n} = F(x)$$

- $F_n(x)$  es un estimador consistente:

$$\begin{aligned} ECM(F_n(x)) &= (\mathbb{E}[F_n(x)] - F(x))^2 + \text{Var}(F_n(x)) \\ &= \frac{1}{n^2} \text{Var}(nF_n(x)) = \frac{nF(x)(1 - F(x))}{n^2} \rightarrow 0. \end{aligned}$$

- Por el Teorema Central del Límite, asintóticamente, la distribución empírica sigue una distribución normal:

$$\sqrt{n}(F_n(x) - F(x)) \xrightarrow{d} N(0, \sqrt{F(x)(1 - F(x))}).$$

**Teorema 1.11.** (de Glivenko-Cantelli sobre la consistencia uniforme de  $F_n$ ). [31, Tema 7] Si  $F_n(x)$  es la distribución empírica calculada a partir de una muestra  $X_1, X_2, \dots, X_n$  de observaciones i.i.d. provenientes de una población con función de distribución  $F(x)$ , entonces

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{c.s.} 0.$$

El test de Kolmogorov-Smirnov propone utilizar la distancia del supremo para medir la discrepancia entre la función de distribución dada en la hipótesis nula,  $F_0$ , y la función de distribución empírica,  $F_n$ . De este modo, se obtiene el siguiente estadístico de contraste:

$$D_{KS} = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|.$$

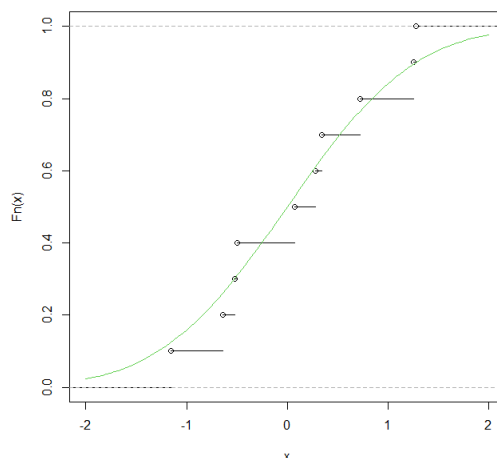


Figura 1.1: Función de distribución empírica

**Lema 1.12.** [29, Sección 10.5] Si  $F = F_0$  y  $F_0$  es continua, entonces la distribución de  $D_{KS}$  no depende de  $F_0$ .

Se suele hacer referencia a esta propiedad diciendo que el estadístico es de distribución libre.

En la práctica, el cálculo de  $D_{KS}$  a partir de la muestra se reduce al siguiente problema de optimización discreto [9, Sección 4.2.3]:

$$\begin{aligned} D_{KS} &= \max \left\{ \max_{1 \leq i \leq n} |F_n(X_i) - F_0(X_i)|, \max_{1 \leq i \leq n} |F_n(X_i^-) - F_0(X_i)| \right\} \\ &= \max \left\{ \max_{1 \leq i \leq n} \left| \frac{i}{n} - F_0(X_i) \right|, \max_{1 \leq i \leq n} \left| F_0(X_i) - \frac{i-1}{n} \right| \right\}, \end{aligned}$$

ya que  $F_n$  es una función escalonada creciente y  $F_0$  es no decreciente, por lo que la máxima diferencia entre  $F_n$  y  $F_0$  se produce en una observación o en el límite por la izquierda de una. Esto se puede apreciar en la Figura 1.1, en la que se representan la función de distribución empírica calculada a partir de una muestra de 10 observaciones i.i.d. de una población normal estándar (en negro) y la función de distribución de la normal estándar (en verde).

Los valores críticos del contraste se obtienen de la distribución del estadístico  $D_{KS}$  bajo la hipótesis nula. Para valores pequeños de  $n$ , la distribución de  $D_{KS}$  no tiene una expresión explícita, pero está tabulada ([20]). A su vez, para valores grandes de  $n$  ( $n \geq 35$ ) se puede utilizar la distribución asintótica obtenida por Kolmogorov ([16]) y tabulada por Smirnov ([27]):

**Teorema 1.13.** Si  $F = F_0$  es continua, se tiene que cuando  $n \rightarrow \infty$ ,

$$\mathbb{P}(\sqrt{n}D_{KS} \leq z) \longrightarrow 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 z^2}$$

uniformemente para  $z > 0$ . (Esto es, dado  $\epsilon > 0$ , existe  $N \in \mathbb{N}$  tal que para todo  $n \geq N$  y para todo  $z > 0$ ,

$$\left| \mathbb{P}(\sqrt{n}D_{KS} \leq z) - \left( 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 z^2} \right) \right| \leq \epsilon.$$

Además, los valores críticos también se pueden aproximar por simulación empleando el método de Montecarlo, que se explica a continuación. Como  $D_{KS}$  tiene distribución libre bajo  $H_0$ , los cálculos pueden llevarse a cabo para una distribución sencilla como la *Uniforme*[0, 1]. La aproximación consiste en simular  $M$  muestras de tamaño  $n$  de una *Uniforme*[0, 1] y calcular el estadístico de Kolmogorov-Smirnov para contrastar la hipótesis nula simple  $H_0 : X \sim \text{Uniforme}[0, 1]$  en cada muestra. Los  $M$  valores obtenidos de esta forma pueden usarse para aproximar la distribución de  $D_{KS}$  bajo la hipótesis nula. En particular, se utilizan los cuantiles muestrales para aproximar los valores críticos.

Hasta ahora se ha visto cómo contrastar una hipótesis nula simple, por lo que cabe preguntarse si el test de Kolmogorov-Smirnov es capaz de contrastar una hipótesis nula compuesta

$$\begin{cases} H_0 : F \in \{F_\theta, \theta \in \Theta\}, \\ H_1 : F \notin \{F_\theta, \theta \in \Theta\}. \end{cases}$$

Para ello se necesita un estimador consistente del parámetro,  $\hat{\theta}$ . Normalmente se usa el de máxima verosimilitud u otros obtenidos a partir de este, aunque hay excepciones dependiendo de la familia paramétrica especificada en la hipótesis nula ([9, Secciones 4.8 a 4.15]).

**Definición 1.14.** [29, Sección 7.3] Dada una muestra  $X_1, \dots, X_n$  de observaciones i.i.d. de una variable aleatoria continua cuya función de distribución pertenece a  $\{F_\theta : \theta \in \Theta\}$ , se denomina función de verosimilitud a

$$L(\theta) = f_\theta(X_1, \dots, X_n) = \prod_{i=1}^n f_\theta(X_i)$$

y estimador de máxima verosimilitud a

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta),$$

siendo  $f_\theta$  la función de densidad asociada a la distribución  $F_\theta$ .

Una vez elegido el estimador, se trabaja con el estadístico

$$\hat{D}_{KS} = \sup_{x \in \mathbb{R}} |F_n(x) - F_{\hat{\theta}}(x)|.$$

En este caso, como ambas distribuciones se estiman a partir de la misma muestra, tienden a ser más parecidas y el valor del estadístico tiende a ser más bajo. Así, el uso de las regiones

críticas obtenidas en el caso de la hipótesis nula simple llevaría a un contraste conservador, que favorecería excesivamente a la hipótesis nula. Por tanto, hay que estudiar la distribución del nuevo estadístico. En general, la distribución de  $\hat{D}_{KS}$  tiene que aproximarse por simulación (se puede recurrir al método de Montecarlo de nuevo). Algunos casos importantes, como las familias normal y exponencial, han sido estudiados detalladamente y existen tablas con los p-valores (obtenidas por Lilliefors en los años 60 usando el método de Montecarlo [18], [19]).

Una forma alternativa de aproximar los valores críticos del test de Kolmogorov-Smirnov para una hipótesis nula compuesta es a través de un bootstrap paramétrico ([11, Sección 21.5]). Para este método es necesario utilizar un estimador consistente  $\hat{\theta}$  del parámetro. Lo que se hace es simular B muestras provenientes de la distribución  $F_{\hat{\theta}}$  del mismo tamaño que la original y en cada una de ellas evaluar el estimador de máxima verosimilitud y calcular el estadístico. Como las muestras se simularon bajo  $H_0$ , estos valores permiten aproximar la distribución del estadístico de contraste bajo la hipótesis nula. De nuevo, los cuantiles muestrales sirven como aproximación de los p-valores.

Al estar basado en la distancia del supremo, el valor del estadístico de Kolmogorov-Smirnov calculado a partir de una muestra puede depender de un único punto, lo que puede llegar a ser problemático. Para evitar esto se puede utilizar la distancia cuadrática, la cual da lugar a la familia de estadísticos de Cramér-von Mises, que tienen la forma:

$$D = \int_{-\infty}^{+\infty} (F_n(x) - F_0(x))^2 \psi(F_0(x)) dF_0(x)$$

donde la función  $\psi(F_0(x))$  asigna pesos al término  $(F_n(x) - F_0(x))^2$ , por lo que el contraste será más sensible allí donde la ponderación sea mayor.

Cuando  $\psi(F_0(x)) = 1$ , se tiene el estadístico de Cramér-von Mises:

$$D_{CvM} = \int_{-\infty}^{+\infty} (F_n(x) - F_0(x))^2 dF_0(x).$$

Por otro lado, si se desea obtener un contraste que sea más potente bajo alternativas en las que la verdadera función de distribución  $F$  difiere de  $F_0$  cerca de las colas de esta (es decir, en los valores de  $x$  para los cuales  $F_0(x)$  esté cerca de 0 o de 1) y para ello se está dispuesto a perder potencia bajo alternativas que difieren de  $F_0$  cerca de la mediana, habría que elegir una función  $\psi$  que fuese grande cerca de 0 y de 1 y pequeña cerca de 1/2. Una manera de construir esta función de pesos es basándose en la siguiente propiedad de la función de distribución empírica.

**Proposición 1.15.** [2, Sección 3] Si  $F = F_0$ , para cada  $x \in \mathbb{R}$  se tiene que

$$\text{Var}(F_n(x) - F_0(x)) = \frac{F_0(x)(1 - F_0(x))}{n}.$$

Para compensar esta dispersión dependiente de  $x$  se pondera la discrepancia por  $\frac{1}{F_0(x)(1-F_0(x))}$ , lo que se traduce en usar como función de pesos  $\psi(u) = \frac{1}{u(1-u)}$ . El estadístico al que se llega es el de Anderson-Darling:

$$D_{AD} = \int_{-\infty}^{+\infty} \frac{(F_n(x) - F_0(x))^2}{F_0(x)(1-F_0(x))} dF_0(x).$$

Como esta elección de la función de  $\psi$  asigna mucho peso a los valores cercanos a 0 y a 1, el estadístico de Anderson-Darling será más sensible a las discrepancias en las colas de las distribuciones, mientras que el de Cramér-von Mises será preferible cuando difieran en el centro.

**Lema 1.16.** [6, Sección 2] Si  $F = F_0$  y  $F_0$  es continua, los estadísticos de la familia Cramér-von Mises son de distribución libre.

Aunque inicialmente pueda parecer que el cálculo de estos estadísticos es complicado, se pueden dar fórmulas explícitas para ambos. En la práctica, se pueden evaluar sobre la muestra ordenada  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  de la siguiente forma [9, Sección 4.2.3]:

- Estadístico de Cramér-von Mises

$$D_{CvM} = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^n \left( F_0(X_{(i)}) - \frac{2i-1}{2n} \right)^2$$

- Estadístico de Anderson-Darling

$$D_{AD} = -1 - \frac{1}{n^2} \sum_{i=1}^n \{ (2i-1) \log(F_0(X_{(i)})) + (2n+1-2i) \log(1-F_0(X_{(i)})) \}$$

Además, las distribuciones asintóticas fueron estudiadas por Anderson y Darling [1]:

**Teorema 1.17.** [1] Dado  $z > 0$  se tiene que, cuando  $n \rightarrow \infty$ ,

$$\mathbb{P}(nD_{CvM} \leq z) \rightarrow \frac{1}{\pi\sqrt{z}} \sum_{j=0}^{\infty} (-1)^j \binom{-1/2}{j} (4j+1)^{1/2} e^{-(4j+1)^2/(16z)} K_{1/4} \left( \frac{(4j+1)^2}{16z} \right),$$

donde  $K_{1/4}(x)$  es la función de Bessel ([32, Sección 3.7]):

$$K_{\nu}(x) = \frac{\pi}{2} \frac{I_{-\nu}(x) - I_{\nu}(x)}{\sin(\nu\pi)}, \text{ siendo } I_{\nu}(x) = \sum_{m=0}^{\infty} \frac{(x/2)^{\nu+2m}}{m!\Gamma(\nu+m+1)} \text{ y } \Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

**Teorema 1.18.** [1] Dado  $z > 0$  se tiene que, cuando  $n \rightarrow \infty$ ,

$$\mathbb{P}(nD_{AD} \leq z) \rightarrow \frac{\sqrt{2\pi}}{z} \sum_{j=0}^{\infty} \binom{-1/2}{j} (4j+1) e^{-((4j+1)^2\pi^2)/(8z)} \int_0^{\infty} e^{z/(8(w^2+1)) - ((4j+1)^2\pi^2 w^2)/(8z)} dw.$$

Si la hipótesis nula fuese compuesta, los estadísticos serían:

$$\hat{D}_{CvM} = \int_{-\infty}^{+\infty} (F_n(x) - F_{\hat{\theta}}(x))^2 dF_{\hat{\theta}}(x),$$

$$\hat{D}_{AD} = \int_{-\infty}^{+\infty} \frac{(F_n(x) - F_{\hat{\theta}}(x))^2}{F_{\hat{\theta}}(x)(1 - F_{\hat{\theta}}(x))} dF_{\hat{\theta}}(x).$$

Tanto para el contraste con hipótesis nula simple como compuesta es posible aproximar los valores críticos y p-valores computacionalmente, utilizando el método de Montecarlo y el bootstrap paramétrico, respectivamente.

**Ejemplo 1.19.** (Comparación de la potencia de los estadísticos de contraste propuestos para el caso de una hipótesis nula simple). Se plantea el siguiente contraste de especificación:

$$\begin{cases} H_0 : F = I_x(0.2, 0.8), \\ H_1 : F \neq I_x(0.2, 0.8), \end{cases}$$

donde  $I_x(\alpha, \beta)$  es la función de distribución de una  $Beta(\alpha, \beta)$ , en este caso  $Beta(0.2, 0.8)$ .

En este ejemplo se estudian tres alternativas diferentes, con el objetivo de plantear un escenario favorable para cada uno de los tests presentados en este capítulo. En el escenario A los datos se generaron de una distribución  $Beta(0.2, 0.85) + 0.01$ , en el escenario B de una  $Beta(0.5, 1)$  y en el escenario C de una  $Beta(0.1, 0.4)$ . En la Figura 1.4 aparecen dos gráficas para cada escenario: a la izquierda se representan las funciones de distribución bajo la hipótesis alternativa y bajo la hipótesis nula, mientras que a la derecha se comparan las respectivas funciones de densidad, lo que permite apreciar de qué manera se diferencian las hipótesis nula y alternativa en cada escenario. El código de R utilizado está en el Anexo I.2.

Los resultados se pueden ver en las Tablas 1.2 y 1.3. En la primera se recogen las proporciones de rechazo de cada test bajo la hipótesis nula para varios tamaños muestrales. En la segunda, las proporciones de rechazo de cada test bajo la hipótesis alternativa para los distintos escenarios y varios tamaños muestrales.

En concreto, la Tabla 1.2 confirma que los contrastes están bien calibrados pues las proporciones de rechazo bajo la hipótesis nula están cerca del nivel de significación deseado de 0.05. Que no sean exactamente iguales a 0.05 se debe a que las distribuciones de los estadísticos son desconocidas y se está haciendo una aproximación.

Lo que se observa en la Tabla 1.3 es lo siguiente: En el escenario A, el test de Kolmogorov-Smirnov es más potente que los otros (salvo para muestras de 10 observaciones, pero ninguno de los tests es muy potente en ese caso). Esto se debe a que existe una gran distancia vertical

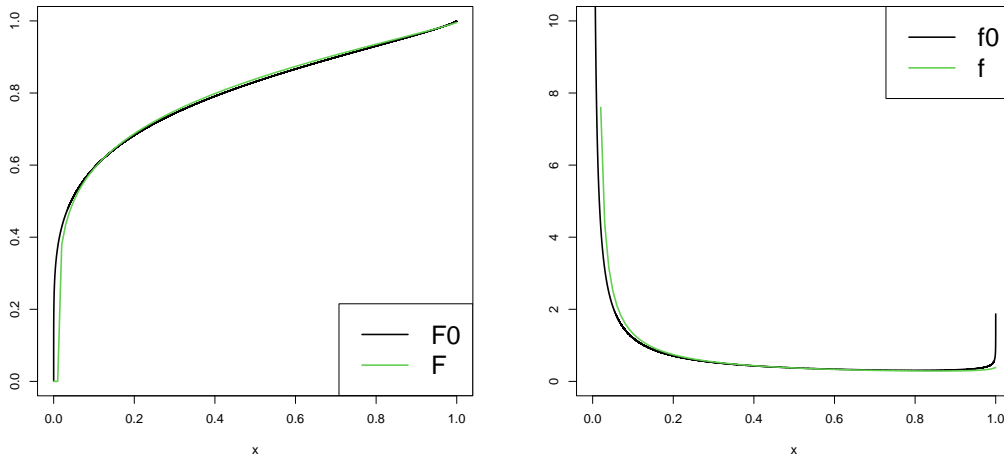
entre las gráficas de las funciones de distribución sin que el área que queda entre ellas sea muy grande. En el escenario B es más potente el test de Cramér-von Mises porque sí que hay un área notable entre las gráficas de las distribuciones. Por último, el test de Anderson-Darling es más potente en el escenario C ya que el área entre las gráficas cerca de la cola superior (cuando  $F_0(x)$  se acerca a 1) es considerable.

| Tamaño muestral $n$ | Prop. rechazos bajo $H_0$ |        |        |
|---------------------|---------------------------|--------|--------|
|                     | KS                        | CvM    | AD     |
| 10                  | 0.0495                    | 0.0473 | 0.0450 |
| 15                  | 0.0471                    | 0.0445 | 0.0457 |
| 20                  | 0.0491                    | 0.0496 | 0.0471 |
| 25                  | 0.0503                    | 0.0479 | 0.0470 |
| 50                  | 0.0531                    | 0.0475 | 0.0478 |
| 75                  | 0.0472                    | 0.0510 | 0.0487 |
| 100                 | 0.0476                    | 0.0434 | 0.0438 |
| 125                 | 0.0553                    | 0.0553 | 0.0540 |

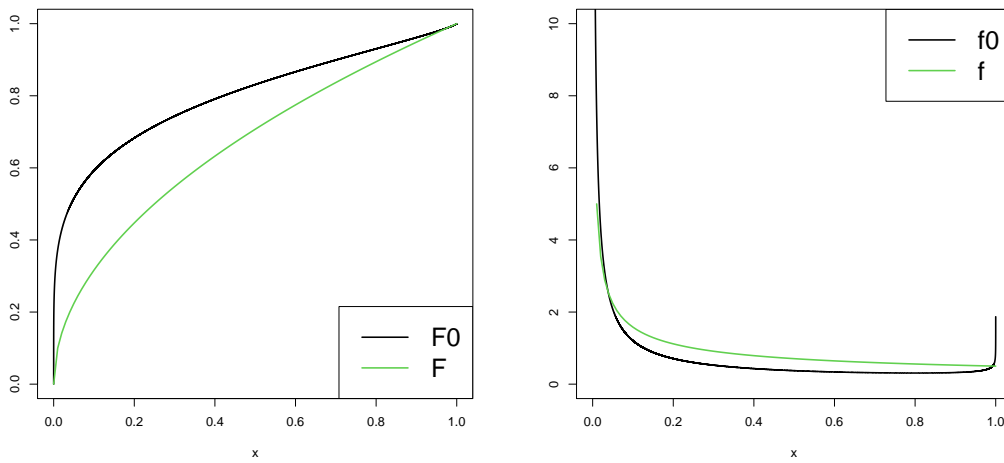
Tabla 1.2: Resultados obtenidos para un nivel de significación de 0.05 y  $M=10000$  muestras.

| Tamaño muestral $n$ | Proporción de rechazos bajo la hipótesis alternativa |        |        |             |        |        |             |        |        |
|---------------------|--|--------|--------|-------------|--------|--------|-------------|--------|--------|
|                     | Escenario A  |        |        | Escenario B |        |        | Escenario C |        |        |
|                     | KS   | CvM    | AD     | KS          | CvM    | AD     | KS          | CvM    | AD     |
| 10                  | 0.0438   | 0.0749 | 0.0904 | 0.4913      | 0.5400 | 0.4696 | 0.2336      | 0.2446 | 0.4474 |
| 15                  | 1  | 0.1455 | 0.1789 | 0.6869      | 0.7492 | 0.7039 | 0.3086      | 0.3193 | 0.5604 |
| 20                  | 1  | 0.3743 | 0.3907 | 0.8367      | 0.8877 | 0.8636 | 0.3944      | 0.4112 | 0.6596 |
| 25                  | 1  | 0.9209 | 0.9433 | 0.9261      | 0.9540 | 0.9478 | 0.4702      | 0.4797 | 0.7449 |
| 50                  | 1  | 1      | 1      | 0.9994      | 0.9998 | 0.9997 | 0.7851      | 0.8068 | 0.9581 |
| 75                  | 1  | 1      | 1      | 1           | 1      | 1      | 0.9179      | 0.9413 | 0.9928 |
| 100                 | 1  | 1      | 1      | 1           | 1      | 1      | 0.9731      | 0.9814 | 0.9993 |
| 125                 | 1  | 1      | 1      | 1           | 1      | 1      | 0.9936      | 0.9965 | 0.9999 |

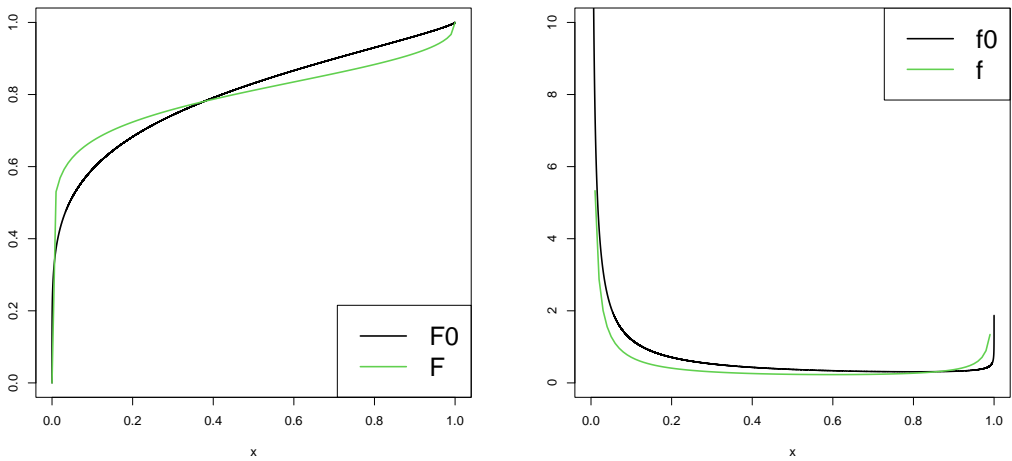
Tabla 1.3: Resultados obtenidos para un nivel de significación de 0.05 y  $M=10000$  muestras.



(a) Escenario A



(b) Escenario B



(c) Escenario C

Figura 1.4: Comparación de la hipótesis nula y la alternativa en cada escenario del Ejemplo 1.19.

### 1.4. Contrastes basados en la función de densidad

La función de densidad de probabilidad describe cómo se distribuyen las probabilidades a lo largo de los valores de una variable aleatoria continua. Se define como sigue:

**Definición 1.20.** [31, Sección 2.2] Se dice que una variable aleatoria real unidimensional  $X$  es continua si existe una función  $f$  tal que

- $f(x) \geq 0$  para todo  $x \in \mathbb{R}$ ,
- $\int_{-\infty}^{\infty} f(x) dx = 1$  y
- para cualquier par de puntos  $a \leq b$ ,  $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx$ .

A  $f$  se le llama función de densidad. Algunas propiedades relevantes de la función de densidad son que  $F(x) = \int_{-\infty}^x f(t) dt$  y  $f(x) = F'(x)$  allí donde  $F$  es derivable.

*Observación 1.21.* A partir de ahora, cuando una integral sea sobre toda la recta real se omitirán los índices de integración.

En particular,  $f$  está definida en casi todo punto de  $\mathbb{R}$  si  $F$  es absolutamente continua. En este caso, también se puede llevar a cabo el contraste de especificación sobre la función de densidad.

Dada una muestra  $X_1, X_2, \dots, X_n$  de observaciones i.i.d. de  $X$ , se puede considerar un contraste con una hipótesis nula simple:

$$\begin{cases} H_0 : f = f_0, \\ H_1 : f \neq f_0, \end{cases}$$

donde  $f_0$  es una función de densidad conocida, o con una hipótesis nula compuesta:

$$\begin{cases} H_0 : f \in \{f_\theta, \theta \in \Theta\}, \\ H_1 : f \notin \{f_\theta, \theta \in \Theta\}, \end{cases}$$

donde  $\{f_\theta, \theta \in \Theta\}$  es una familia paramétrica de densidades.

La idea de los contrastes basados en la función de densidad es medir la discrepancia entre una densidad hipotética y un estimador no paramétrico de la densidad de  $X$  calculado a partir de una muestra.

### 1.4.1. Estimación no paramétrica de la función de densidad

En los contrastes discutidos en esta sección se va a utilizar el estimador kernel de la densidad, un estimador no paramétrico que se construye a partir de una muestra de la siguiente forma ([30, Sección 2.2]):

$$f_{n,h}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

donde  $K$  es la función kernel, que cumple  $K(x) \geq 0$  para todo  $x \in \mathbb{R}$  e  $\int K(t) dt = 1$ ; y  $h > 0$  es la ventana o parámetro de suavizado. A veces se escribe usando el kernel reescalado  $K_h = h^{-1}K(u/h)$  y el estimador queda

$$f_{n,h}(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i).$$

La forma del estimador hace que herede las propiedades relativas a la continuidad y la diferenciabilidad de la función kernel  $K$ , por lo que hay que elegir esta adecuadamente ([26, Sección 3.1]). Normalmente se escoge una función de densidad unimodal y simétrica respecto al cero. En la Figura 1.5 se pueden ver las estimaciones kernel de la densidad obtenidas a partir de una muestra usando distintas funciones de densidad como kernels y un histograma de esa muestra. Se observa que la estimación que usa la densidad de una uniforme no es diferenciable, mientras que el resto son suaves y bastante parecidas entre ellas. En general, la elección de  $K$  (siempre que sea continua, unimodal y simétrica) no tiene mucho impacto sobre la estimación. Los kernels más utilizados son el gaussiano (la función de densidad de una normal estándar):

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

y el Epanechnikov:

$$K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1).$$

Como se verá a continuación, el parámetro ventana  $h$  tiene efecto sobre el sesgo y la varianza. Así, valores muy grandes de  $h$  resultan en un estimador demasiado suavizado que tiene poca varianza y mucho sesgo, mientras que valores pequeños de  $h$  dan lugar a un estimador con poco sesgo y mucha varianza (las relaciones se explicitarán en lo que sigue). Si bien existen situaciones en las que se puede elegir un  $h$  óptimo según cierto criterio (como pueden ser validación cruzada o plug-in), en general no está claro cuál es ese valor óptimo de  $h$  a la hora de usar esta estimación en un contraste de especificación (pues dependería de la hipótesis alternativa y puede no ser posible encontrarlo). En la Figura 1.6 se representan las estimaciones kernel de la densidad para una muestra usando el kernel gaussiano y varias ventanas distintas. Se aprecia como al elegir una

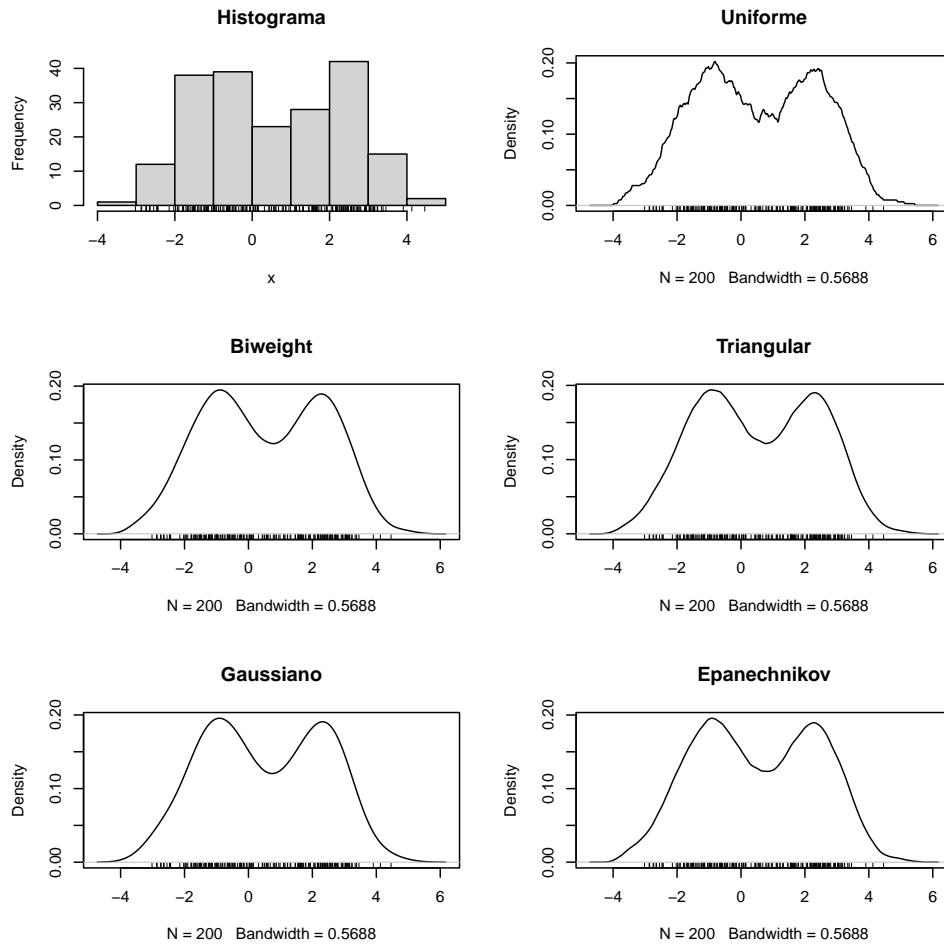


Figura 1.5: Forma del estimador kernel de la densidad dependiendo de la función kernel  $K$  escogida.

ventana grande se pierde la bimodalidad, mientras que al escoger una pequeña aparecen picos espúreos.

A continuación se enuncian algunas propiedades de este estimador y su comportamiento asintótico.

**Proposición 1.22.** [30, Sección 2.3] *Se tiene que:*

- $\mathbb{E}[f_{n,h}(x)] = \mathbb{E}[K_h(x - X)] = \int K_h(x - y)f(y) dy = (K_h * f)(x),$
- $Var(f_{n,h}(x)) = n^{-1} [(K_h^2 * f)(x) - (K_h * f)^2(x)].$

Como consecuencia de lo anterior,

$$Sesgo(f_{n,h}(x)) = \mathbb{E}[f_{n,h}(x)] - f(x) = (K_h * f)(x) - f(x). \quad (1.1)$$

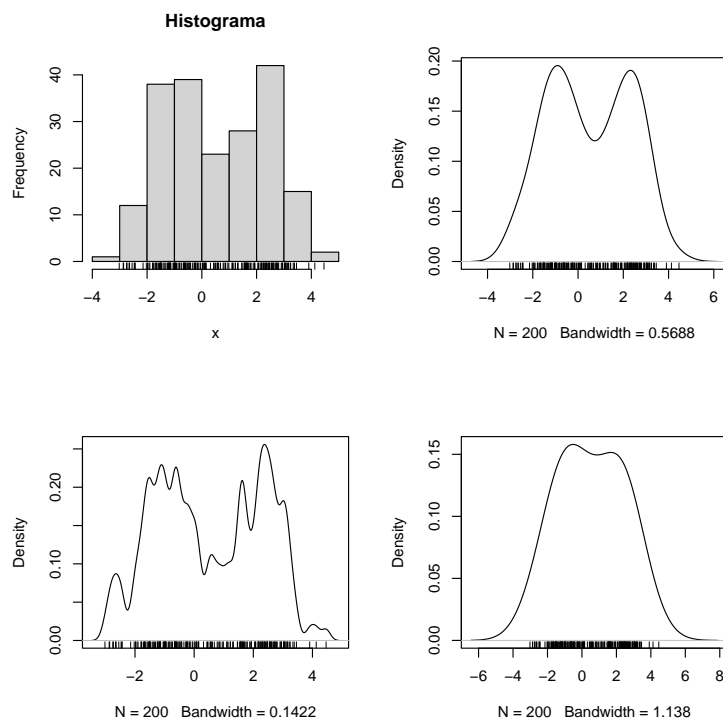


Figura 1.6: Forma del estimador kernel de la densidad dependiendo de la ventana  $h$  escogida.

La convolución por el kernel representa una suavización, por lo que el sesgo del estimador es la diferencia entre la verdadera función de densidad y una suavización de la misma.

Además,

$$ECM[f_{n,h}(x)] = n^{-1} [(K_h^2 * f)(x) - (K_h * f)^2(x)] + [(K_h * f)(x) - f(x)]^2$$

y, para obtener una medida global del error, se puede integrar en su dominio de definición, quedando

$$\begin{aligned} \int ECM[f_{n,h}(x)] dx &= (nh)^{-1} \int K^2(x) dx + (1 - n^{-1}) \int (K_h * f)^2(x) dx \\ &\quad - 2 \int (K_h * f)(x) f(x) dx + \int f(x)^2 dx \end{aligned}$$

**Proposición 1.23.** [30, Sección 2.5] (Comportamiento asintótico del sesgo y la varianza). *Asumiendo que:*

- la segunda derivada de la densidad,  $f''$ , es continua, su cuadrado es integrable y existe  $M > 0$  tal que es monótona en  $(-\infty, M)$  y  $(M, \infty)$ ,
- la ventana  $h = h_n$  es una sucesión que cumple  $\lim_{n \rightarrow \infty} h = 0$  y  $\lim_{n \rightarrow \infty} nh = \infty$  (es decir, se acerca a cero más despacio que  $n^{-1}$ ),

- el kernel  $K$  es una densidad acotada y simétrica respecto a cero con momento de orden 4 finito,

y utilizando expansiones en series de Taylor, se llega a:

- $\text{Sesgo}(f_{n,h}(x)) = \frac{1}{2}h^2 f''(x) \int t^2 K(t) dt + o(h^2),$
- $\text{Var}(f_{n,h}(x)) = (nh)^{-1} f(x) \int K^2(t) dt + o\{(nh)^{-1}\}.$

*Observación 1.24.* El sesgo depende de la curvatura de la verdadera densidad,  $f''$ , esto se debe a que el estimador tiende a suavizar los picos y los valles de  $f$ . Asintóticamente, tanto el sesgo como la varianza tienden a cero. Se puede escribir

$$ECM[f_{n,h}(x)] = (nh)^{-1} f(x) \int K^2(t) dt + \frac{1}{4}h^4 f''(x)^2 \left( \int t^2 K(t) dt \right)^2 + o\{(nh)^{-1} + h^4\}$$

y entonces

$$\int ECM[f_{n,h}(x)] dx = (nh)^{-1} \int K^2(t) dt + \frac{1}{4}h^4 \int f''(x)^2 dx \left( \int t^2 K(t) dt \right)^2 + o\{(nh)^{-1} + h^4\},$$

donde la notación  $a_n = o(b_n)$  (cuando  $n \rightarrow \infty$ ) significa que  $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$ .

### 1.4.2. Contrastes

Para el caso de un contraste con hipótesis nula simple, Bickel y Rosenblatt ([5, Sección 4]) propusieron el siguiente estadístico, que mide la discrepancia cuadrática entre  $f_0(x)$  y  $f_{n,h}(x)$ :

$$D_{BR} = \int (f_{n,h}(x) - (K_h * f_0)(x))^2 dx. \quad (1.2)$$

Ya se vio en (1.1) que el estimador kernel de la densidad no es insesgado, pues  $\mathbb{E}[f_{n,h}(x)] = (K_h * f)(x)$ . Entonces la convolución que aparece en el estadístico es la esperanza del estimador si la hipótesis nula es cierta, y se plantea así para evitar que el sesgo del estimador pueda distorsionar el valor del estadístico.

**Teorema 1.25.** [5, Sección 4] (Sobre la distribución asintótica del estadístico de Bickel y Rosenblatt) Si  $H_0$  es cierta y se cumplen:

- $\int K(t) dt = 1$  y el kernel  $K$  verifica que: o bien tiene soporte compacto en el que es absolutamente continuo con derivada  $K'$  o bien es absolutamente continuo en  $\mathbb{R}$  y su derivada  $K'$  cumple  $\int |K'(t)|^k dt < \infty$ , para  $k = 1, 2$ .

- La densidad  $f$  es continua y acotada.
- La función  $f^{1/2}$  es absolutamente continua y su derivada,  $\frac{1}{2}f'f^{-1/2}$ , está acotada en valor absoluto. Además,

$$\int_{|z| \geq 3} |z|^{3/2} (\log \log z)^{1/2} (|K'(z)| + |K(z)|) dz < \infty.$$

- La ventana  $h = h_n$  cumple  $\lim_{n \rightarrow \infty} h = 0$  y  $\lim_{n \rightarrow \infty} nh = \infty$ .

Entonces se tiene que

$$n\sqrt{h} \frac{D_{BR} - \mu(K)}{\sigma(K, f_0)} \xrightarrow{d} N(0, 1),$$

$$\text{donde } \mu(K) = (nh)^{-1} \int K^2(u) du \text{ y } \sigma^2(K, f_0) = 2 \left( \int \left[ \int K(u+v)K(v) dv \right]^2 du \right) \left( \int f_0^2(x) dx \right).$$

*Observación 1.26.* En las condiciones del teorema anterior, se puede obtener una aproximación asintótica de los valores críticos del contraste. En efecto, fijado un nivel de significación  $\alpha$ , la hipótesis nula se rechaza si

$$D_{BR} > \mu(K) + z_{1-\alpha} \frac{\sigma(K, f_0)}{n\sqrt{h}},$$

donde  $z_p$  denota el cuantil  $p$  de la distribución normal estándar.

Por otro lado, para un contraste con hipótesis nula compuesta, de nuevo se necesita un estimador del parámetro  $\theta$  indexando la familia de densidades bajo la nula. Dada una muestra  $X_1, X_2, \dots, X_n$  de observaciones i.i.d. de la variable aleatoria  $X$  y tomando un estimador adecuado de  $\theta$  (cumpliendo las condiciones requeridas en el Teorema 1.27, como por ejemplo podría ser el estimador de máxima verosimilitud),  $f_{\hat{\theta}}(x)$  será el estimador de máxima verosimilitud de  $f(x)$  bajo la hipótesis nula.

Fan ([12, Sección 4]) propuso el siguiente estadístico de contraste:

$$D_F = \int (f_{n,h}(x) - (K_h * f_{\hat{\theta}})(x))^2 dx,$$

Del mismo modo que en el contraste de Bickel y Rosenblatt, la convolución se introduce para evitar que la discrepancia observada entre  $f_{\hat{\theta}}(x)$  y  $f_{n,h}(x)$  provenga del sesgo del estimador kernel  $f_{n,h}$  pues, si la hipótesis nula es cierta,  $\mathbb{E}[f_{n,h}(x)] = (K_h * f_{\hat{\theta}})(x)$ .

**Teorema 1.27.** [12, Sección 4] (Sobre la distribución asintótica del estadístico de Fan) Si  $H_0$  es cierta y se cumplen las siguientes condiciones:

- El kernel  $K$  está acotado, es simétrico, integra 1 en  $\mathbb{R}$  y verifica  $|u|K(u) \rightarrow 0$  cuando  $|u| \rightarrow \infty$  y  $\int |u^m|K(u) du < \infty$  para algún  $m$  entero positivo.

- La ventana  $h = h_n$  cumple  $\lim_{n \rightarrow \infty} h = 0$  y  $\lim_{n \rightarrow \infty} nh = \infty$ .
- $f$  y su  $m$ -ésima derivada están acotadas y son uniformemente continuas en  $\mathbb{R}$ .
- Existe  $\theta_* \in \Theta$  tal que  $\hat{\theta} \xrightarrow{c.s.} \theta_*$  y

$$\hat{\theta} - \theta_* = (nA(\theta_*))^{-1} \sum_{i=1}^n D_\theta \log f_{\theta_*}(X_i) + o(n^{-1/2}),$$

donde  $D_\theta \log f_{\theta_*}(X_i)$  es el vector de derivadas parciales de primer orden de  $\log f_\theta(X_i)$  con respecto a  $\theta$  evaluadas en  $\theta = \theta_*$  y  $A(\theta_*) = \mathbb{E} \left[ \frac{\partial^2 \log f_\theta(X_i)}{\partial \theta^2} \Big|_{\theta=\theta_*} \right]$ .

- La función de densidad paramétrica  $f_\theta$  y sus derivadas parciales de segundo orden respecto a  $\theta$  están uniformemente acotadas y son uniformemente continuas en  $\mathbb{R} \times \Theta$ .  $f_{\theta_*}$  y su  $m$ -ésima derivada son acotadas y uniformemente continuas en  $\mathbb{R}$ . Además,  $\int |D_\theta f_{\theta_*}(x)| dx < \infty$ , donde  $D_\theta f_{\theta_*}(x)$  es el vector de derivadas parciales de primer orden de  $f_\theta(x)$  con respecto a  $\theta$  evaluadas en  $\theta = \theta_*$ .

Entonces se tiene que

$$n\sqrt{h} \frac{D_F - \mu(K)}{\sigma(K, f_{n,h})} \xrightarrow{d} N(0, 1),$$

donde

$$\mu(K) = \frac{1}{nh} \int K^2(u) du$$

y

$$\sigma^2(K, f_{n,h}) = 2 \left( \int \left[ \int K(u+v)K(v) dv \right]^2 du \right) \left( \int f_{n,h}^2(x) dx \right).$$

De nuevo, este resultado proporciona una aproximación asintótica de los valores críticos: fijado un nivel de significación  $\alpha$ , el contraste rechaza la hipótesis nula cuando

$$D_F > \mu(K) + z_{1-\alpha} \frac{\sigma(K, f_{n,h})}{n\sqrt{h}}.$$

Dado un contraste, distintos valores del parámetro ventana  $h$  darán lugar a distintos valores del correspondiente estadístico, lo que podría llevar a distintas conclusiones sobre la validez de la hipótesis nula  $H_0$  para una misma muestra ([12, Sección 5]).

### 1.4.3. Comparación con los contrastes basados en la función de distribución

Algunas ventajas de los contrastes basados en la función de densidad con respecto a los de la función de distribución son:

- Los estadísticos siguen asintóticamente una distribución normal, cuyas propiedades han sido ampliamente estudiadas. En el caso de los contrastes basados en la función de distribución, aunque se conocen las distribuciones asintóticas de los estadísticos, no pertenecen a familias conocidas de distribuciones y no se sabe tanto sobre sus propiedades.
- Son más potentes frente a alternativas que difieren localmente de la hipótesis nula ([24, Sección 3]).

**Ejemplo 1.28.** (Potencia del estadístico de Bickel y Rosenblatt para el caso de una hipótesis nula simple). Se plantea el siguiente contraste de especificación:

$$\begin{cases} H_0 : f = \phi, \\ H_1 : f \neq \phi, \end{cases}$$

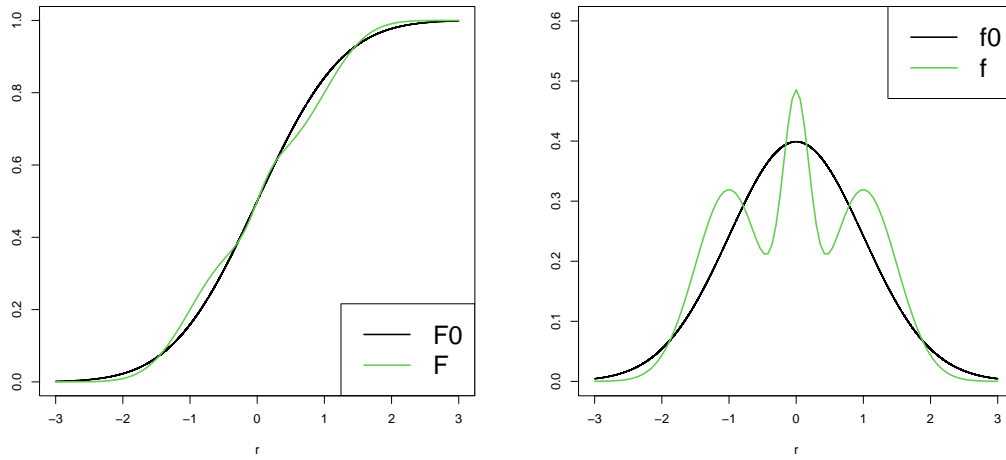
donde  $\phi$  es la función de densidad de una  $N(0, 1)$ .

En este ejemplo se estudian tres alternativas diferentes para ilustrar como se comporta el test. En el escenario A los datos se generaron de una mixtura de distribuciones normales  $N(-1,0.5)$ ,  $N(0,0.2)$  y  $N(1,0.5)$  con probabilidades 0.4,0.2 y 0.4, respectivamente; en el escenario B de una  $N(0,1.2)$  y en el escenario C de una  $N(0.2,1)$ . En la Figura 1.8 se representan las funciones de distribución y de densidad de la hipótesis nula junto a las de las hipótesis alternativas. Para la estimación de la densidad se ha usado un kernel gaussiano y la ventana seleccionada por la regla del dedo [25, Subsección 3.4.2]. El código de R utilizado está en el Anexo I.3.

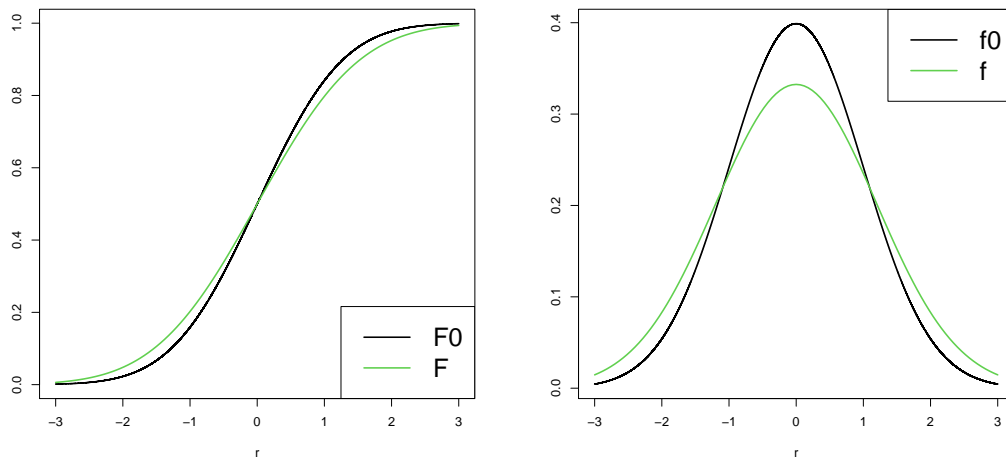
En la Tabla 1.7 se recogen las proporciones de rechazo bajo la hipótesis nula (que están cerca del valor deseado de 0.05, por lo que el contraste está bien calibrado) y bajo las distintas alternativas (que se van acercando a 1 según aumenta el tamaño de las muestras, por lo que el contraste es consistente bajo todas las alternativas) para varios tamaños muestrales.

| Tamaño muestral $n$ | Prop. rechazos bajo $H_0$ | Proporción de rechazos bajo la alternativa |             |             |
|---------------------|---------------------------|--|-------------|-------------|
|                     |                           | Escenario A                                | Escenario B | Escenario C |
| 100                 | 0.038                     | 0.459                                      | 0.450       | 0.268       |
| 200                 | 0.047                     | 0.711                                      | 0.674       | 0.479       |
| 300                 | 0.047                     | 0.892                                      | 0.886       | 0.626       |
| 400                 | 0.057                     | 0.960                                      | 0.957       | 0.801       |
| 500                 | 0.048                     | 0.983                                      | 0.977       | 0.863       |
| 600                 | 0.058                     | 0.994                                      | 0.995       | 0.925       |
| 700                 | 0.056                     | 0.999                                      | 0.997       | 0.956       |
| 800                 | 0.044                     | 1  | 1           | 0.968       |

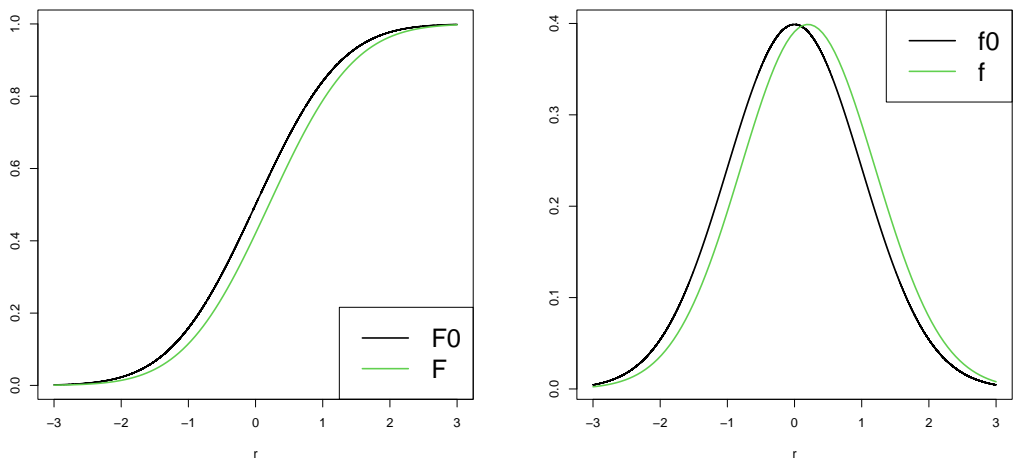
Tabla 1.7: Resultados obtenidos para un nivel de significación de 0.05 y  $M=1000$  muestras usando el método de Montecarlo.



(a) Escenario A



(b) Escenario B



(c) Escenario C

Figura 1.8: Comparación de la hipótesis nula y la alternativa en cada escenario.

## Capítulo 2

# Introducción al Análisis de Supervivencia

### 2.1. Introducción

El término análisis de supervivencia se refiere al estudio de datos que miden el tiempo que pasa hasta que ocurre un evento. Este evento puede ser la muerte, la aparición de un tumor, el desarrollo de una enfermedad, el fallo de una máquina. . . O puede ser un evento positivo, como la remisión de una enfermedad tras un tratamiento, el abandono del tabaco. . . A la variable de interés se le suele llamar tiempo de vida, tiempo de supervivencia o tiempo de fallo dependiendo del contexto, y al evento de interés, muerte o fallo.

El tiempo de vida será una variable aleatoria real unidimensional no negativa  $X$ , que en este trabajo se considerará continua. Para estudiar esta variable se podría usar la función de densidad  $f(x)$ , que se interpreta como la probabilidad relativa de que el evento de interés ocurra en el instante  $x > 0$ , o la de distribución  $F(x)$ , que representa la probabilidad de que el evento de interés suceda en el instante  $x$  o antes. Sin embargo, en el análisis de supervivencia resulta interesante trabajar también con otras funciones relacionadas.

### 2.2. Funciones de interés

Algunas de las funciones más importantes que caracterizan la distribución de  $X$  son la función de supervivencia, que expresa la probabilidad de que un individuo sobreviva más allá del tiempo  $x$ ; la función de riesgo, que indica la probabilidad de que un individuo que ha sobrevivido hasta el tiempo  $x$  experimente el evento de interés en el instante siguiente; y la función de riesgo

acumulado, que representa la cantidad total de riesgo acumulado hasta el tiempo  $x$  (es como sumar todos los riesgos instantáneos desde el inicio hasta ese momento).

**Definición 2.1.** [15, Sección 2.2] La función de supervivencia se define como

$$S(x) := \mathbb{P}(X > x).$$

$S(x)$  es el complemento de la función de distribución,  $S(x) = 1 - F(x)$ , por lo que es una función continua, monótona decreciente y se tiene que  $S(0) = 1$  y  $\lim_{x \rightarrow \infty} S(x) = 0$ . También se puede obtener integrando la función de densidad:

$$S(x) = \mathbb{P}(X > x) = \int_x^{\infty} f(s) ds,$$

y entonces

$$f(x) = -S'(x). \quad (2.1)$$

**Definición 2.2.** [15, Sección 2.3] La función de riesgo, o razón de fallo, es

$$h(x) := \lim_{\Delta x \rightarrow 0} \frac{\mathbb{P}(x \leq X < x + \Delta x \mid X \geq x)}{\Delta x}.$$

Se tiene que

$$h(x) = \frac{f(x)}{S(x)} = -\frac{d}{dx} [\log S(x)], \quad (2.2)$$

por tanto,

$$S(x) = \exp \left\{ - \int_0^x h(s) ds \right\}.$$

**Definición 2.3.** [15, Sección 2.3] Se llama función de riesgo acumulado, o razón de fallo acumulada, a

$$H(x) := \int_0^x h(s) ds.$$

De esta manera,  $H(x) = -\log S(x)$  y  $S(x) = e^{-H(x)}$ .

**Ejemplo 2.4.** En la Figura 2.1 se representan las funciones comentadas para el caso en el que  $X$  sigue una distribución exponencial de parámetro 2, esto es,  $F(x) = 1 - e^{-2x}$ . Se tiene que  $S(x) = 1 - F(x) = e^{-2x}$ ,  $f(x) = F'(x) = 2e^{-2x}$ ,  $h(x) = f(x)/S(x) = 2$  y  $H(x) = -\log S(x) = 2x$ . Se puede interpretar del siguiente modo: como el riesgo instantáneo  $h(x)$  es constante, el riesgo acumulado  $H(x)$  crece linealmente y por tanto la supervivencia  $S(x)$  tiene un decaimiento exponencial.

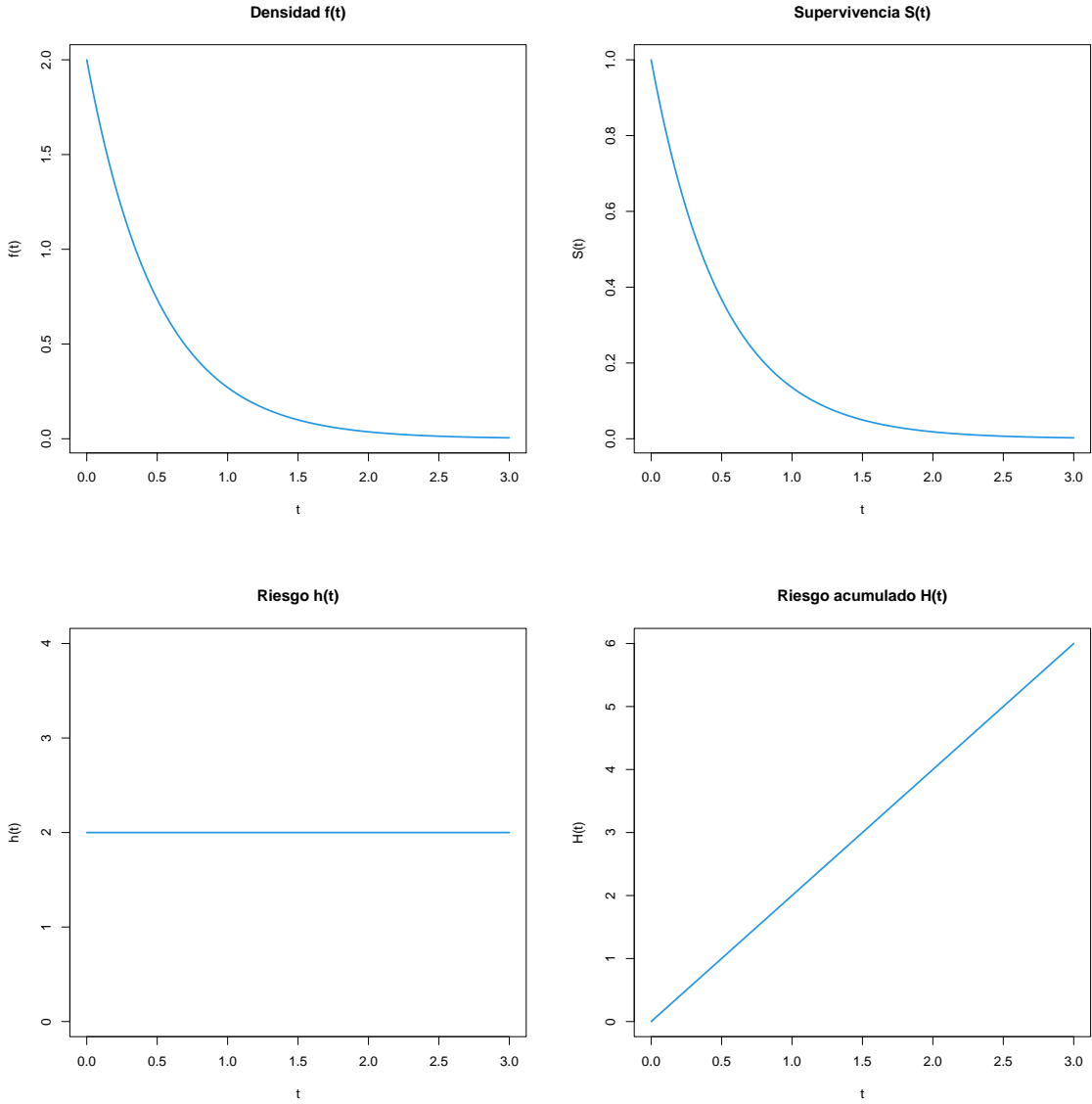


Figura 2.1: Funciones de interés de la  $Exp(2)$ .

### 2.3. Modelos paramétricos notables

Aunque a menudo se usarán modelos no paramétricos o semiparamétricos, es apropiado comentar también los modelos paramétricos, que son más comunes. Estos modelos son muy populares en el análisis de supervivencia y ayudan a entender la naturaleza de los parámetros y funciones utilizadas, en especial la función de riesgo. Algunos modelos paramétricos importantes son las distribuciones exponencial, Weibull, gamma, gamma generalizada, normal logarítmica y logística logarítmica. A continuación se expondrán algunas propiedades de estas distribuciones, empezando por la más utilizada históricamente:

- Exponencial. Su función de supervivencia es

$$S(x) = e^{-\lambda x}, \quad (2.3)$$

donde  $\lambda > 0$ ,  $x > 0$  y su función de densidad es  $f(x) = \lambda e^{-\lambda x}$ . Se caracteriza por una función de riesgo constante

$$h(x) = \lambda, \quad (2.4)$$

debida a que la distribución exponencial "no tiene memoria", esto es,

$$\mathbb{P}(X \geq x + z | X \geq x) = \mathbb{P}(X \geq x),$$

lo que facilita los cálculos a la vez que puede ser poco realista en algunos contextos. Además,  $H(x) = \lambda x$  y tanto la media como la desviación típica de la distribución exponencial valen  $1/\lambda$ . La interpretación de este modelo se vio en el ejemplo 2.4.

A continuación se comentan otras dos distribuciones, la Weibull y la gamma, que generalizan la exponencial. Otras propiedades de la exponencial vendrán dadas implícitamente de las de estas distribuciones.

- Weibull. Su función de supervivencia es  $S(x) = e^{-\lambda x^\alpha}$ , para  $x > 0$ .  $\lambda > 0$  es un parámetro de escala, y  $\alpha > 0$  es un parámetro de forma que cuando vale 1 da lugar a la distribución exponencial.

La distribución Weibull generaliza la exponencial añadiendo un segundo parámetro,  $\alpha$ , que permite modelar tasas de riesgo crecientes ( $\alpha > 1$ ) o decrecientes ( $\alpha < 1$ ), lo que la hace notablemente más flexible y permite su aplicación a contextos en los que el riesgo cambia con el tiempo (fallos mecánicos, evolución de enfermedades...). Este hecho, junto con la relativa sencillez de sus funciones de supervivencia, densidad  $f(x) = \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha}$  y riesgo  $h(x) = \lambda \alpha x^{\alpha-1}$  han hecho que sea un modelo paramétrico muy popular.

El momento de orden  $r$  de la distribución Weibull es  $[\Gamma(1 + \frac{r}{\alpha})] \lambda^{-r/\alpha}$ , por lo que la media y la varianza son  $[\Gamma(1 + \frac{1}{\alpha})] \lambda^{-1/\alpha}$  y  $\{\Gamma(1 + \frac{2}{\alpha}) - [\Gamma(1 + \frac{1}{\alpha})]^2\} \lambda^{-2/\alpha}$ , respectivamente, donde  $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$  es la función gamma.

- Gamma. Su función de densidad viene dada por

$$f(x) = \frac{\lambda^\beta x^{\beta-1} e^{-\lambda x}}{\Gamma(\beta)},$$

donde  $\lambda > 0$ ,  $\beta > 0$ ,  $x > 0$  y  $\Gamma(\beta)$  es la función gamma. Por razones similares a las expuestas en el caso de la distribución Weibull,  $\lambda$  es un parámetro de escala y  $\beta$ , de forma.

La distribución gamma generaliza la exponencial añadiendo un segundo parámetro,  $\beta$ . Se puede entender la gamma como una suma de  $\beta$  exponenciales independientes con el mismo parámetro  $\lambda$ . Trivialmente, cuando  $\beta = 1$  se reduce a un modelo exponencial. También permite modelar variables en las que el riesgo varía con el tiempo (la función de riesgo es incluso más flexible que para la Weibull), pero con la desventaja de que no tiene una expresión cerrada. Este tipo de modelos es adecuado para situaciones en las que el tiempo hasta un evento se puede modelar como la suma de otros tiempos positivos (por ejemplo, el tiempo hasta la falla de un sistema compuesto por varias partes).

La función de supervivencia y la función de riesgo se pueden obtener a partir de la densidad usando las relaciones expuestas en las ecuaciones (2.1) y (2.2). La función de riesgo es monótona creciente si  $\beta > 1$ , con  $h(0) = 0$ ; y monótona decreciente si  $\beta < 1$ , con  $\lim_{x \rightarrow 0} h(x) = \infty$ . En cualquier caso, se acerca a  $\lambda$  a medida que va creciendo  $t$ . Cuando  $\beta = 1$  se obtiene la distribución exponencial, por lo que la función de riesgo será constante. Todo esto se puede ver en la Figura 2.2.

La media y la varianza de la distribución gamma son  $\beta/\lambda$  y  $\beta/\lambda^2$ , respectivamente.

Tanto la Weibull como la gamma son generalizaciones de la exponencial. Si se introducen ambas generalizaciones simultáneamente se obtiene el modelo siguiente.

- Gamma generalizada. La distribución gamma generalizada se obtiene introduciendo un parámetro adicional  $\alpha$  que permite mayor flexibilidad a la hora de escoger una función de riesgo. Este modelo tiene función de densidad

$$f(x) = \frac{\alpha \lambda^\beta x^{\alpha\beta-1} e^{-\lambda x^\alpha}}{\Gamma(\beta)}.$$

Esta distribución se reduce a la exponencial cuando  $\alpha = \beta = 1$ , a la Weibull cuando  $\beta = 1$ , a la gamma cuando  $\alpha = 1$  y se acerca a la normal logarítmica cuando  $\beta \rightarrow \infty$ . Es una distribución muy útil para la especificación de modelos.

- Normal logarítmica. Se dice que una variable aleatoria  $X$  sigue la distribución normal logarítmica o log-normal si su logaritmo  $Y = \log X$  sigue una distribución normal. Del mismo modo que la distribución normal, la distribución normal logarítmica queda completamente especificada dados dos parámetros  $\mu$  y  $\sigma$ , la media y la varianza de  $Y$ . Su función de

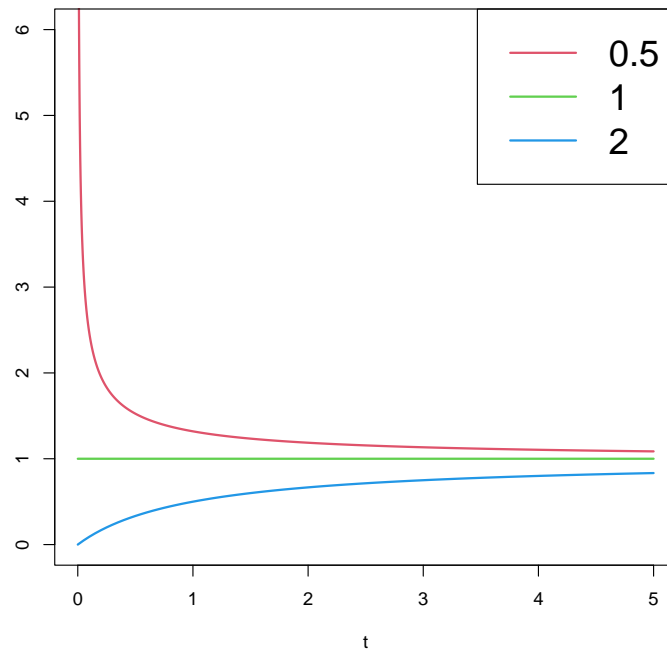


Figura 2.2: Función de riesgo de la distribución  $Gamma(1, \beta)$  dependiendo del valor de  $\beta$ , cuando  $\beta = 1$  se corresponde con una  $Exp(1)$ .

supervivencia viene dada por

$$S(x) = 1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right),$$

donde  $\Phi$  es la función de distribución de una variable normal estándar.

La media viene dada por  $e^{\mu+\sigma^2/2}$  y la varianza por  $(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$ .

Este modelo es adecuado cuando los datos presentan colas largas.

- Logística logarítmica. Se dice que una variable  $X$  sigue una distribución log-logística o logística logarítmica si su logaritmo  $Y = \log X$  sigue una distribución logística. La función de supervivencia de la distribución log-logística se puede escribir de forma relativamente sencilla:

$$S(x) = \frac{1}{1 + \lambda x^\alpha},$$

donde  $\alpha = \frac{1}{\sigma} > 0$  y  $\lambda = e^{-\mu/\sigma}$ .

La media y la varianza de  $X$  vienen dadas por

$$\mathbb{E}[X] = \frac{\pi}{\alpha \lambda^{1/\alpha} \text{sen}(\pi/\alpha)}, \text{ si } \alpha > 1$$

y

$$\text{Var}(X) = \frac{2\pi}{\alpha\lambda^{2/\alpha} \text{sen}(2\pi/\alpha)} - \mathbb{E}[Y]^2, \text{ si } \alpha > 2.$$

Esta distribución es útil en modelos con riesgo no monótono.

## 2.4. Censura

A menudo, los tiempos de vida no son observables en todos los individuos de una muestra: puede suceder que el seguimiento se detenga antes de que se produzca el evento de interés para un sujeto o que este experimente otro tipo de evento que compite con el de interés e impide su observación. Por ejemplo, en el seguimiento de tiempos de supervivencia para una determinada enfermedad, es posible que al término del estudio algunos de los pacientes seguidos hayan sobrevivido, o puede ser que algunos hayan muerto por otras causas o incluso que hayan dejado de acudir a las citas. Las observaciones correspondientes a estos pacientes serían entonces observaciones censuradas. Este tipo de dificultades en la observación complican el análisis y requieren de un tratamiento estadístico específico. Para ello se introduce la variable aleatoria tiempo de censura,  $C$ , que también será real unidimensional, no negativa y continua.

Hay varios tipos de censura: por la derecha, por la izquierda o en un intervalo. Para lidiar adecuadamente con la censura en el análisis, se debe tener en cuenta la forma en la que se obtuvieron los datos. La censura por la derecha puede ser de los siguientes tipos:

- Tipo I. El evento sólo se observa si sucede antes de un tiempo preestablecido, que puede variar dependiendo del individuo.
- Tipo II. El estudio continúa hasta que han alcanzado el evento de interés los primeros  $r$  individuos, donde  $r < n$  es algún entero fijado con antelación.
- Aleatoria. Este tipo de censura puede producirse cuando se estudia un evento de interés y algunos individuos experimentan otro evento que los lleva a ser eliminados del estudio. En estos casos, el evento de interés no es observable para los individuos que experimentan el evento competidor, y sus tiempos de vida son censurados aleatoriamente por la derecha en ese instante.

Este trabajo se centra en la censura aleatoria por la derecha. Además, se asume que la censura es no informativa, esto es, el conocimiento del tiempo de censura para un individuo no aporta información sobre su supervivencia si hubiese continuado en el estudio: el tiempo del evento de interés y el tiempo de censura son independientes. Algunos ejemplos donde estos tiempos pueden ser independientes son muertes accidentales, emigración, etc.

En la censura aleatoria por la derecha, es conveniente usar la siguiente notación: para cada individuo  $i$  estudiado, se asume que hay un tiempo de vida  $X_i$  y un tiempo de censura  $C_i$ . El tiempo de vida  $X_i$  del individuo se conocerá si, y sólo si,  $X_i$  es menor o igual a su tiempo de censura  $C_i$ . En otro caso, el tiempo de vida del individuo queda censurado en  $C_i$ . Los datos de este experimento se pueden representar por pares de variables aleatorias  $(T_i, \delta_i)$ , donde  $\delta_i = I(X_i \leq C_i)$  y  $T_i = \min(X_i, C_i)$ .

En la Figura 2.3 aparece una representación típica de una muestra con censura aleatoria por la derecha, en la que los círculos blancos representan tiempos de censura y las cruces, tiempos de vida.

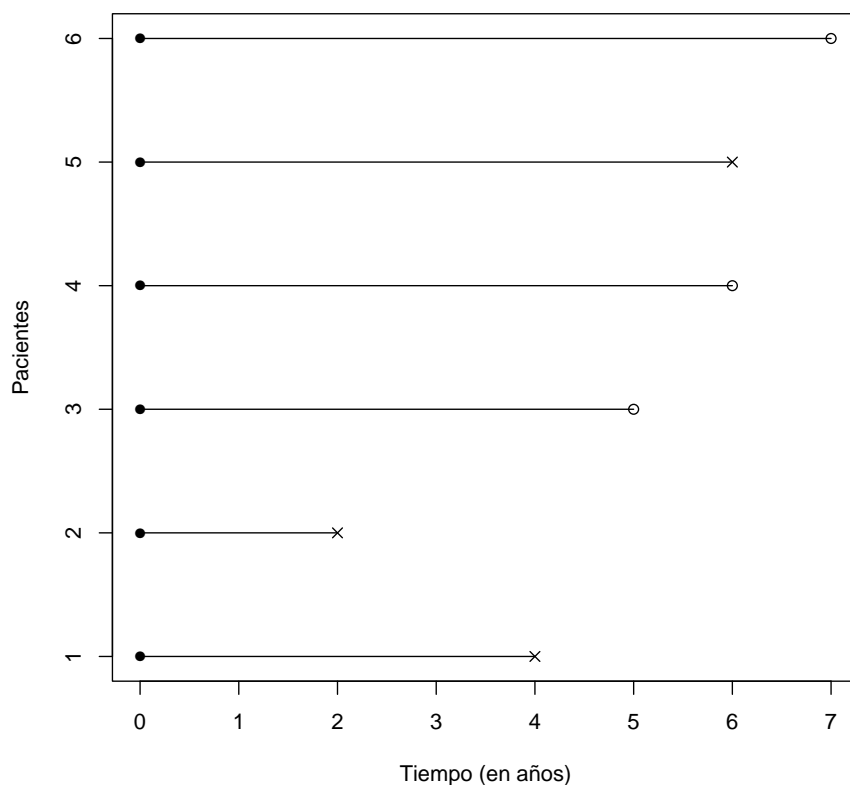


Figura 2.3: Ejemplo de representación gráfica de una muestra censurada aleatoriamente por la derecha. Para los pacientes 1, 2 y 5 se representa el tiempo hasta el evento de interés (observaciones completas), mientras que para los pacientes 3, 4 y 6 se representa el tiempo hasta la censura (observaciones censuradas).

## Capítulo 3

# Contrastes de especificación con datos censurados basados en la función de distribución

### 3.1. Introducción

En este capítulo y en el siguiente se adaptarán las ideas presentadas para datos completos al caso de censura aleatoria por la derecha. Esto requiere modificar los estimadores de las funciones de distribución y densidad, lo que afecta a la construcción y a las propiedades de los estadísticos.

En las siguientes dos secciones se verá cómo se puede estimar la función de distribución en el contexto de censura aleatoria por la derecha. Se recuerda que en este contexto existe una variable aleatoria de interés  $X$  y otra variable aleatoria  $C$  que puede censurar las observaciones de la primera. Ambas son variables aleatorias reales unidimensionales, continuas y no negativas. Se denotarán por  $F$  y  $G$  a las funciones de distribución y por  $f$  y  $g$  a las funciones de densidad de  $X$  y  $C$ , respectivamente. Se trabajará a partir de una muestra aleatoria simple obtenida bajo censura aleatoria por la derecha  $\{(T_i, \delta_i)\}_{i=1}^n$  de los pares  $(T, \delta)$ , donde

$$T = \min\{X, C\} \text{ y } \delta = I(X \leq C).$$

Además,  $X$  y  $C$  serán independientes.

### 3.2. Estimación no paramétrica de la distribución bajo censura aleatoria por la derecha

El estimador más común de la función de supervivencia de la variable de interés  $X$  para datos con censura aleatoria por la derecha es el conocido como estimador límite-producto y fue propuesto por Kaplan y Meier en 1958 ([14]). Para poder definir este estimador es necesario ordenar los  $n$  tiempos observados, sin tener en cuenta si son tiempos de vida o de censura:  $0 < T_{(1)} < T_{(2)} < \dots < T_{(n)}$ . De este modo,

$$\hat{S}_n(x) = \begin{cases} 1 & \text{si } 0 \leq x < T_{(1)}, \\ \prod_{T_{(i)} \leq x} \left( \frac{n-i}{n-i+1} \right)^{\delta_i} & \text{si } T_{(1)} \leq x \leq T_{(n)}. \end{cases}$$

Si el mayor tiempo observado se corresponde a un tiempo de vida, la curva de supervivencia estimada vale cero a partir de este punto. Por otro lado, si  $T_{(n)}$  es un tiempo de censura, el valor de  $\hat{S}_n(x)$  a partir de este punto no está determinado porque no se sabe cuánto habría sobrevivido el último individuo si no hubiese sido censurado.

Se han propuesto diferentes alternativas no paramétricas para lidiar con esta ambigüedad. Por un lado, Efron ([10]) propuso tomar  $\hat{S}_n(x) = 0$  para  $x > T_{(n)}$ , lo que se corresponde con asumir o forzar que la última observación  $T_{(n)}$  es un tiempo de fallo. Por otro lado, Gill ([13, Sección 3.2]) propuso tomar  $\hat{S}_n(x) = \hat{S}_n(T_{(n)})$  para  $x > T_{(n)}$ , lo que se corresponde con asumir que el individuo cuyo tiempo de vida queda censurado en  $T_{(n)}$  no muere nunca, y lleva a un estimador con sesgo. Estas dos técnicas se corresponden con las dos situaciones más extremas que se pueden dar, también existen otras opciones intermedias que utilizan modelos paramétricos para completar las colas. En lo que sigue se usará la versión de Efron.

El estimador de la función de distribución que se obtiene usando este método es

$$\hat{F}_n(x) = 1 - \hat{S}_n(x), \tag{3.1}$$

que coincide con la función de distribución empírica  $F_n$  en caso de que no haya censura. También se puede usar el estimador de Kaplan-Meier para estimar la función de riesgo acumulada  $H(x) = -\ln[S(x)]$ , siendo el estimador  $\hat{H}_n(x) = -\ln[\hat{S}_n(x)]$ .

*Observación 3.1.* Para diferenciar los estimadores de la función de distribución para datos completos y para datos censurados, en este trabajo se denota por  $F_n$  al primero y por  $\hat{F}_n$  al segundo.

La varianza de este estimador se puede estimar con la fórmula de Greenwood ([14]):

$$\widehat{Var}[\hat{S}_n(x)] = \hat{S}_n(x)^2 \sum_{Z_{(i)} \leq x} \frac{1}{(n-i)(n-i+1)}.$$

### 3.3. Estimación paramétrica de la distribución bajo censura aleatoria por la derecha

En caso de que no haya observaciones censuradas, queda una expresión análoga a la de la varianza de la función de distribución empírica:

$$\widehat{Var}[\hat{S}_n(x)] = \frac{\hat{S}_n(x)(1 - \hat{S}_n(x))}{n}.$$

**Teorema 3.2.** ([28], Sobre la consistencia uniforme del estimador de Kaplan-Meier). Se tiene que

$$\sup_{0 \leq x \leq \tau_H} |\hat{F}_n(x) - F(x)| \xrightarrow{c.s.} 0$$

si  $F(\tau_H) = 1$  o  $G(\tau_H) < 1$ , donde  $H$  es la función de distribución de  $T = \min\{X, C\}$  y  $\tau_H = \inf\{t : H(t) = 1\}$ .

### 3.3. Estimación paramétrica de la distribución bajo censura aleatoria por la derecha

Se empezará por obtener la función de verosimilitud de la variable de interés,  $X$ , para una muestra con censura aleatoria por la derecha,  $\{(T_i, \delta_i)\}_{i=1}^n$ .

Se recuerda que, en caso de que no haya censura, se dispone de una muestra de  $n$  observaciones i. i. d. de los tiempo de supervivencia,  $\{X_i\}_{i=1}^n$ , y la función de verosimilitud tiene la forma

$$L_n = \prod_{i=1}^n f(X_i).$$

Para tener en cuenta la censura, que se produce cuando  $X_i$  es mayor que  $C_i$ , habría que sustituir los factores  $f(T_i)$  correspondientes a observaciones censuradas (que representan la probabilidad relativa de que la variable aleatoria  $X$  valga  $T_i$ ) por  $S(T_i)$  (que es la probabilidad de que la variable aleatoria  $X$  sea mayor que  $T_i$ ) ([21, Sección 2.6]). De este modo, quedaría

$$L_n = \prod_{i=1}^n f(T_i)^{\delta_i} S(T_i)^{1-\delta_i} = \prod_{i=1}^n \left( \frac{f(T_i)}{S(T_i)} \right)^{\delta_i} S(T_i) = \prod_{i=1}^n h(T_i)^{\delta_i} S(T_i),$$

donde la tercera igualdad se obtiene de la ecuación (2.2).

Si se intenta obtener una función de distribución que maximice la verosimilitud sin imponer ningún modelo paramétrico, se llega a la misma expresión que proporciona el estimador de Kaplan-Meier (pues tiene saltos en las observaciones de la variable de interés y no los tiene en los tiempos de censura). Si se impone un modelo paramétrico (indexado por un parámetro  $\theta$  perteneciente a un espacio paramétrico  $\Theta$ ), la expresión resultante es

$$L_n(\theta) = \prod_{i=1}^n f_\theta(T_i)^{\delta_i} S_\theta(T_i)^{1-\delta_i} = \prod_{i=1}^n h_\theta(T_i)^{\delta_i} S_\theta(T_i). \quad (3.2)$$

Se trata de encontrar el valor de  $\theta \in \Theta$  que maximiza esta función, que será el estimador de máxima verosimilitud del parámetro.

En el caso de la distribución exponencial, que es una de las más sencillas, sustituyendo en la expresión (3.2) las funciones vistas en las ecuaciones (2.3) y (2.4), queda

$$L_n(\lambda) = \prod_{i=1}^n \lambda^{\delta_i} e^{-\lambda T_i} = \lambda^{(\sum_{i=1}^n \delta_i)} e^{-\lambda(\sum_{i=1}^n T_i)} = \lambda^d e^{-\lambda V},$$

donde  $d = \sum_{i=1}^n \delta_i$  es el número de observaciones sin censura y  $V = \sum_{i=1}^n T_i$  es la suma de todos los tiempos observados. El estimador se obtiene maximizando la verosimilitud tal y como sucede en ausencia de censura

$$\hat{\lambda} = \arg \max_{\lambda > 0} L_n(\lambda)$$

Para ello es más fácil trabajar con el logaritmo de la verosimilitud, que al ser una función monótona creciente alcanzará el máximo en el mismo punto que la función de verosimilitud.

$$l(\lambda) = \log L_n(\lambda) = d \log(\lambda) - \lambda V$$

$$l'(\lambda) = \frac{d}{\lambda} - V$$

El máximo se alcanza en el punto en el que se anula la derivada,

$$\hat{\lambda} = \frac{d}{V},$$

y se comprueba que efectivamente se trata de un máximo analizando el valor de la segunda derivada en  $\lambda = \frac{d}{V}$

$$l''(\lambda) = -\frac{d}{\lambda^2} \implies l''\left(\frac{d}{V}\right) = -\frac{V^2}{d} < 0.$$

Se tiene que el estimador de máxima verosimilitud del parámetro de una distribución exponencial para una muestra con censura aleatoria por la derecha es

$$\hat{\lambda} = \frac{d}{V} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n T_i}.$$

Para otras distribuciones, las cuentas pueden no ser tan sencillas y la estimación del parámetro puede requerir recurrir a métodos numéricos.

### 3.4. Contrastes basados en la función de distribución bajo censura aleatoria por la derecha

En la Sección 1.3, se estudiaron varios tipos de estadísticos para contrastar hipótesis nulas simples y compuestas acerca de la función de distribución de una variable aleatoria de la que se

disponía de una muestra aleatoria simple de datos completos. Ahora se tratará de ver cómo se puede llevar a cabo ese mismo estudio a partir de una muestra de datos incompletos, con censura aleatoria por la derecha.

Empezando por la hipótesis nula simple

$$\begin{cases} H_0 : F = F_0, \\ H_1 : F \neq F_0, \end{cases}$$

se utilizará como criterio para decidir si se rechaza o no la hipótesis nula la discrepancia entre la función de distribución propuesta en la hipótesis nula,  $F_0$ , y la estimación no paramétrica de la distribución calculada sobre la muestra censurada usando el estimador límite-producto,  $\hat{F}_n$  (definido en la ecuación (3.1)). Igual que en el caso de datos completos, dependiendo de la distancia usada para medir la discrepancia se obtendrán distintos estadísticos de contraste ([23]):

- Estadístico de tipo Kolmogorov-Smirnov para la distancia del supremo

$$D_{KS}^C = \sup_{x>0} |\hat{F}_n(x) - F_0(x)|.$$

- Estadístico de tipo Cramér-von Mises para la distancia cuadrática

$$D_{CvM}^C = \int_0^\infty (\hat{F}_n(x) - F_0(x))^2 dF_0(x).$$

- Estadístico de tipo Anderson-Darling para la distancia cuadrática ponderada enfatizando las colas

$$D_{AD}^C = \int_0^\infty \frac{(\hat{F}_n(x) - F_0(x))^2}{F_0(x)(1 - F_0(x))} dF_0(x).$$

Estos estadísticos coinciden con los planteados para datos completos si la muestra no tiene observaciones censuradas, también son de distribución libre y se pueden calcular fácilmente a partir de una muestra con las siguientes fórmulas ([23],[9]):

$$D_{KS}^C = \max \left\{ \max_{i:\delta_i=1} |\hat{F}_n(T_{[i]}) - F_0(T_{[i]})|, \max_{i:\delta_i=1} |F_0(T_{[i]}) - \hat{F}_n(T_{[i-1]})| \right\},$$

$$D_{CvM}^C = \frac{d}{3} + d \sum_{i:\delta_i=1} \left\{ \hat{F}_n(T_{[i-1]}) [F_0(T_{[i]}) - F_0(T_{[i-1]})] \cdot \left[ \hat{F}_n(T_{[i-1]}) - F_0(T_{[i]}) + F_0(T_{[i-1]}) \right] \right\},$$

$$D_{AD}^C = -d + d \sum_{i:\delta_i=1} \left\{ \left[ \hat{F}_n(T_{[i-1]})^2 - \hat{F}_n(T_{[i]})^2 \right] \log F_0(T_{[i]}) - \left[ (1 - \hat{F}_n(T_{[i-1]}))^2 - (1 - \hat{F}_n(T_{[i]}))^2 \right] \log(1 - F_0(T_{[i]})) \right\},$$

donde  $d = \sum_{i=1}^n \delta_i$  es el número de observaciones no censuradas,  $T_{[1]}, \dots, T_{[d]}$  son las observaciones no censuradas ordenadas,  $T_{[0]} = 0$  y el estimador límite-producto  $\hat{F}_n$  se calcula utilizando la corrección de Efron del estimador de Kaplan-Meier introducida en la Sección 3.2. Su distribución bajo la hipótesis nula depende de la distribución de la variable aleatoria  $C$ , que es desconocida. Los valores críticos tendrán que obtenerse por simulación, por ejemplo usando el método de Montecarlo.

Si la hipótesis nula es compuesta,

$$\begin{cases} H_0 : F \in \{F_\theta, \theta \in \Theta\}, \\ H_1 : F \notin \{F_\theta, \theta \in \Theta\}, \end{cases}$$

se estudia la discrepancia entre  $\hat{F}_n$  y  $F_{\hat{\theta}}$ , siendo  $\hat{\theta}$  el estimador de máxima verosimilitud de  $\theta$ , cuya obtención para muestras con censura se vio en la Sección 3.3. De este modo, los distintos estadísticos tendrían la misma forma que los propuestos en la hipótesis nula simple sustituyendo  $F_0$  por  $F_{\hat{\theta}}$ :

$$\hat{D}_{KS}^C = \sup_{x>0} |\hat{F}_n(x) - F_{\hat{\theta}}(x)|$$

$$\hat{D}_{CvM}^C = \int_0^\infty (\hat{F}_n(x) - F_{\hat{\theta}}(x))^2 dF_{\hat{\theta}}(x)$$

$$\hat{D}_{AD}^C = \int_0^\infty \frac{(\hat{F}_n(x) - F_{\hat{\theta}}(x))^2}{F_{\hat{\theta}}(x)(1 - F_{\hat{\theta}}(x))} dF_{\hat{\theta}}(x)$$

Para aproximar los valores críticos y p-valores de estos contrastes también se puede usar un bootstrap con algunas modificaciones, pues habrá que estimar la distribución de los tiempos de censura,  $G$ , para poder generar nuevas muestras como la estudiada ([7]). Para ello se puede usar una estimación obtenida a partir de la de Kaplan-Meier, invirtiendo los papeles que juegan las variables aleatorias. Así, la modificación propuesta para estimar la función de distribución de la variable  $C$  es:

$$1 - \hat{G}_n(c) = \begin{cases} 1, & \text{si } 0 \leq c < T_{(1)}, \\ \prod_{T_{(i)} \leq c} \left( \frac{n-i+1}{n-i+2} \right)^{1-\delta_i}, & \text{si } T_{(1)} \leq c \leq T_{(n)}, \end{cases} \quad (3.3)$$

donde  $T_{(1)} < \dots < T_{(n)}$  son los tiempos observados ordenados.

## Capítulo 4

# Contrastes de especificación con datos censurados basados en la función de densidad

### 4.1. Introducción

A continuación se tratará la estimación no paramétrica de la función de densidad para datos con censura aleatoria por la derecha y se plantearán algunos contrastes de especificación basados en esa estimación. Además de adaptar los contrastes vistos en la Sección 1.4, se introducirá un nuevo tipo de contraste basado en la función de distribución que usa una estimación suavizada de la misma aprovechando el hecho de que existe función de densidad al ser la variable aleatoria de interés  $X$  continua.

### 4.2. Estimación de la densidad bajo censura aleatoria por la derecha

De nuevo, se trabaja a partir de una muestra censurada aleatoriamente por la derecha, como se explicó en la Sección 3.1. En la Sección 3.2 se introdujo el estimador de Kaplan-Meier de la función de supervivencia, a partir del cual se obtenía el estimador límite-producto de la función de distribución  $\hat{F}_n$  (3.1).

En el caso de datos completos, la función de densidad se estimaba utilizando la metodología kernel no paramétrica tal y como se introdujo en la Subsección 1.4.1. Para datos con censura

aleatoria por la derecha, se sustituirá la función de distribución empírica  $F_n$  por el estimador límite-producto de la distribución  $\hat{F}_n$  en la definición del estimador kernel de la densidad para datos completos, obteniendo el siguiente estimador no paramétrico de la densidad bajo censura aleatoria por la derecha [8, Sección 1]:

$$\hat{f}_{n,h}(x) = \int_{-\infty}^{\infty} K_h(x-u) d\hat{F}_n(u),$$

que coincidiría con  $f_{n,h}(x)$  si no hubiese censura. Del mismo modo que en la estimación para datos completos,  $h > 0$  es el parámetro ventana y el  $K$  es la función kernel.

*Observación 4.1.* Para diferenciar los estimadores de la función de densidad para datos completos y para datos censurados, en este trabajo se denota por  $f_{n,h}$  al primero y por  $\hat{f}_{n,h}$  al segundo.

### 4.3. Contrastes basados en la función de densidad bajo censura aleatoria por la derecha

Los contrastes sobre la función de densidad tendrán que centrarse en medir la discrepancia entre cierta función de densidad bajo la hipótesis nula  $f_0$  y la estimación obtenida a partir de la muestra  $\hat{f}_{n,h}$ . En el caso de datos completos, el estadístico de Bickel y Rosenblatt (expresión (1.2)) utilizaba la distancia de  $L^2$  para medir la discrepancia entre  $f_0$  y  $f_n$ , por tanto resulta natural tratar de plantear un estadístico análogo para datos con censura aleatoria por la derecha que mida una discrepancia cuadrática entre  $f_0$  y  $\hat{f}_{n,h}$ . En el siguiente teorema se estudia la distribución asintótica de

$$I_n = \int_0^{\tau_{F_0}} (\hat{f}_{n,h}(x) - f_0(x))^2 d\mu(x),$$

donde  $\mu(x)$  es una función de pesos que tendrá que cumplir ciertas condiciones explicadas en el enunciado y  $\tau_{F_0} = \inf\{x : F_0(x) = 1\}$  (para otras funciones también se denotará de esta forma). A continuación se propondrá un caso particular de  $\mu(x)$  cumpliendo las hipótesis del teorema y para la cual  $I_n$  puede ser evaluada sobre una muestra de datos censurados, proporcionando así un estadístico de contraste.

**Teorema 4.2.** [8, Sección 1, Teorema 3] Si  $f = f_0$ , se tiene que

$$(h\sigma^2(\tau_{F_0}))^{-1/2} (nhI_n - m_n(\tau_{F_0})) \xrightarrow{d} N(0, 1)$$

si se cumplen las siguientes hipótesis sobre la función kernel, las funciones de distribución y densidad de  $Y$  y  $C$  y el parámetro ventana:

- $K$  está acotada, tiene soporte compacto,  $\int K(t) dt = 1$ ,  $\int K^2(t) dt > 0$ ,  $\int |t| d|K(t)| < \infty$  y  $\int tK(t) dt = 0$ .

- $\tau_{F_0} \leq \tau_G$ ,

$$\limsup_{t \rightarrow \tau_{F_0}} \frac{(1 - F_0(t))^\beta}{1 - G(t)} < \infty \text{ para cierto } 0 \leq \beta < 1, \quad (4.1)$$

$f_0$  está acotada en  $[0, \tau_{F_0})$  y es monótona en un entorno de  $\tau_{F_0}$ ,  $\sup_{0 < c < \tau_{F_0}} g(c) < \infty$ ,

$$\sup_{0 < x < \tau} \left| \frac{f_0'(x)}{\sqrt{f_0(x)}} \right| < \infty \text{ para todo } \tau < \tau_{F_0}, \quad \sup_{0 < x < \tau_{F_0}} |f_0''(x)| < \infty \text{ y}$$

$$d\mu(x) = w(x) dx, \text{ donde } w(x) \geq 0 \text{ es continua en } [0, \tau_{F_0})$$

$$\text{y existe } \epsilon > 0 \text{ tal que } \limsup_{c \rightarrow \tau_F} \frac{c^{1+\epsilon} w(c)}{(1 - G(c))^4} < \infty. \quad (4.2)$$

- $h \rightarrow 0$ ,  $nh^4 \rightarrow \infty$ ,  $nh^5 \rightarrow 0$  y  $\frac{n^{\kappa-1/2}}{h} \rightarrow 0$  para algún  $\kappa > \frac{\beta}{\beta+1}$ .

Los coeficientes normalizantes vienen dados por:

$$\bar{f}_n(x) = \frac{1}{n} \int K\left(\frac{x-u}{h}\right) dF_0(u),$$

$$D = \left( \int K^2(t) dt \right)^{1/2},$$

$$l(t) = (1 - G(t))^{-1/2} (f_0(t))^{1/2},$$

$$m_n(\tau_{F_0}) = \int \int_0^{\tau_{F_0}} (l(t)Dx + (nh)^{1/2} [\bar{f}_n(t) - f_0(t)])^2 w(t)\phi(x) dt dx,$$

$$\eta(t) = \frac{\int K(u)K(t+u) du}{\int K^2(u) du},$$

$$\begin{aligned} \sigma_1^2 = (2\pi)^{-1} \int \left\{ \int \int (xy)^2 (1 - \eta^2(u))^{-1/2} \cdot \right. \\ \left. \exp \left\{ -\frac{1}{2(1 - \eta^2(u))} (x^2 - 2xy\eta(u) + y^2) \right\} dx dy - 1 \right\} du, \end{aligned}$$

$$\sigma^2(T) = \sigma_1^2 \int_0^T (1 - G(t))^{-2} f_0^2(t) w^2(t) dt \left( \int K^2(u) du \right)^2.$$

*Observación 4.3.* Las condiciones sobre la función kernel se cumplen para una función de densidad acotada con soporte compacto y media 0 (por ejemplo, el kernel de Epanechnikov).

*Observación 4.4.* [8, Sección 1, observación 1] Si  $\tau_{F_0} < \tau_G$ , la condición (4.1) se cumple para  $\beta = 0$  y se puede tomar  $w(c) = 1 - G(c)$  en la condición (4.2), quedando

$$\sigma^2(\tau_F) = \sigma_1^2 \int_0^{\tau_F} f_0^2(x) dx \left( \int K^2(u) du \right)^2,$$

que ya no depende de la función desconocida  $G(c)$ . Además, si se cumple que

$$\int_{-\infty}^{\tau_H} (1 - G(s))^{-1} dF_0(s) < \infty \quad \text{si } F_0(\tau_H) < 1$$

$$(1 - F_0(s))^\beta = O(1 - G(s)) \text{ cuando } s \rightarrow \tau_H \text{ para cierto } 0 \leq \beta < 1 \quad \text{si } F_0(\tau_H) = 1,$$

entonces se puede estimar  $G$  a partir de la muestra por su estimador límite-producto  $\hat{G}_n$  presentado en la expresión (3.3) y usar  $w(c) = 1 - \hat{G}_n(c)$  en vez de  $w(c) = 1 - G(c)$  en la definición de  $I_n$ , por lo que la expresión resultante se puede evaluar sobre una muestra censurada.

De este modo, podría usarse como estadístico de contraste

$$I_n = \int_0^{\tau_{F_0}} \left( \hat{f}_{n,h}(x) - f_0(x) \right)^2 \left( 1 - \hat{G}_n(x) \right) dx.$$

#### 4.4. Estimación continua de la distribución

Hasta ahora se han utilizado funciones escalonadas para estimar la distribución, sin tener en cuenta si la variable aleatoria estudiada era continua o discreta. En el caso de que sea continua, existe una función de densidad asociada  $f(x)$  tal que  $F(x) = \int_{-\infty}^x f(u) du$ . Esto puede aprovecharse para construir un nuevo estimador no paramétrico de la distribución que sea continuo y que se obtendrá integrando un estimador no paramétrico de la densidad como los planteados anteriormente.

Para datos completos, si se dispone de una muestra aleatoria simple  $X_1, \dots, X_n$ , quedaría

$$\tilde{F}_n(x) = \int_{-\infty}^x f_{n,h}(u) du = \int_{-\infty}^x \frac{1}{nh} \sum_{i=1}^n K\left(\frac{u - X_i}{h}\right) du = \frac{1}{n} \sum_{i=1}^n W\left(\frac{x - X_i}{h}\right), \quad (4.3)$$

donde  $W(s) = \int_{-\infty}^s K(u) du$ .

**Proposición 4.5.** [22, Sección 2] Si  $K(x)$  es una función de densidad acotada cumpliendo que  $\lim_{|x| \rightarrow \infty} |xK(x)| = 0$  y  $h \xrightarrow{n \rightarrow \infty} 0$ , entonces  $\tilde{F}_n(x)$  es un estimador consistente de  $F(x)$ , por ser asintóticamente insesgado y con

$$Var(\tilde{F}_n(x)) = \frac{F(x)(1 - F(x))}{n}.$$

Por otro lado, para datos con censura aleatoria por la derecha el estimador se calcula de la siguiente forma [3]:

$$\hat{\tilde{F}}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{1 - \hat{G}(T_i)} W\left(\frac{x - T_i}{h}\right).$$

En la práctica la función  $G$  es desconocida, por lo que no será posible evaluar esta expresión, pero se puede usar el estimador límite-producto  $\hat{G}_n$  dado en la expresión (3.3). De este modo, se llega a la siguiente expresión para un estimador continuo de la función de distribución, que sí se puede usar en la práctica:

$$\hat{\tilde{F}}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{1 - \hat{G}_n(T_i)} W\left(\frac{x - T_i}{h}\right).$$

*Observación 4.6.* Para distinguir el estimador que usa la verdadera función de distribución  $G$  del que usa su estimador límite-producto  $\hat{G}_n$ , en este trabajo se denota por  $\hat{F}$  al primero y por  $\hat{F}_n$  al segundo.

En lo que sigue se utiliza la siguiente notación:

$$\mu_i(K) = \int u^i K(u) du, \quad i = 0, 1, 2, \quad (4.4)$$

que es el momento de orden  $i$  del kernel  $K$  y

$$R(K) = \int K^2(u) du. \quad (4.5)$$

Se enumeran algunas hipótesis sobre  $F$ ,  $G$  y  $K$  [3].

- $F(x)$  es dos veces derivable y  $F''(x)$  está acotada y es uniformemente continua.
- Para  $l = 0, 1, 2$ , la  $l$ -ésima derivada de  $K$ ,  $K^{(l)}$ , está acotada y la integral de su valor absoluto es finita con momento de orden dos finito.
- $R(K)$  es finito y  $\mu_0(K) = 1$ ,  $\mu_1(K) = 0$ ,  $\mu_2(K) < \infty$ .
- Existe un  $h$  lo suficientemente pequeño como para que  $\frac{W(h^{-1}(y-x))}{(1-F(y))(1-G(y))}$  esté uniformemente acotado para  $|y-x| > M$ , para todo  $M > 0$ .

*Observación 4.7.* La segunda condición implica que  $W(s)$  está acotada y la tercera, en concreto la parte de  $\mu_0(K) = 1$ , implica que  $\lim_{s \rightarrow -\infty} W(s) = 0$  y  $\lim_{s \rightarrow +\infty} W(s) = 1$ . Casi todos los kernels utilizados frecuentemente satisfacen las tres primeras condiciones y todos los de soporte compacto verifican la cuarta.

A continuación, se presentan algunos resultados sobre el estimador  $\hat{F}_n$  para el caso en el que se verifican las hipótesis enunciadas.

**Teorema 4.8.** [3] (Sobre las propiedades asintóticas de  $\hat{F}(x)$ ). Si se cumplen las hipótesis anteriores y  $h \rightarrow 0$  cuando  $n \rightarrow \infty$  se tiene que:

$$\mathbb{E}[\hat{F}(x)] - F(x) = F''(x) \frac{h^2}{2} \mu_2(K) + o(h^2),$$

$$\text{Var}(\hat{F}(x)) = \frac{1}{n} \int_0^x \frac{f(u)}{1-G(u)} du - \frac{(1-F(x))^2}{n} - \frac{2h}{n} \frac{f(x)}{1-G(x)} \int uK(u)W(u) du + O(h^2 n^{-1}).$$

**Teorema 4.9.** [3] (Sobre la normalidad asintótica de  $\hat{F}(x)$ ). Bajo las hipótesis anteriores, si  $nh \rightarrow \infty$  y  $h \rightarrow 0$  cuando  $n \rightarrow \infty$ , entonces

$$\sqrt{\text{Var}(\hat{F}(x))^{-1}} (\hat{F}(x) - \mathbb{E}[\hat{F}(x)]) \xrightarrow{d} N(0, 1).$$

**Teorema 4.10.** [3] (Sobre la consistencia de  $\widehat{F}(x)$ ). Sea  $\tau = \sup\{x : F(x) < 1 - \epsilon\}$  para un  $\epsilon > 0$  pequeño. Bajo las hipótesis anteriores se cumple que:

$$\sup_{0 < x < \tau} |\widehat{F}(x) - F(x)| = O_p(h^2 + n^{-1/2}).$$

*Observación 4.11.* [3] Por la convergencia uniforme del estimador de Kaplan-Meier a la verdadera función de supervivencia, se tiene que

$$\widehat{F}_n(x) = \widehat{F}(x) + o_p(n^{-1/2}),$$

con lo que  $\widehat{F}(x)$  y  $\widehat{F}_n(x)$  son equivalentes asintóticamente.

La notación  $X_n = O_p(b_n)$ ,  $X_n = o_p(b_n)$  tiene un significado análogo al de  $a_n = O(b_n)$ ,  $a_n = o(b_n)$  pero se usa para sucesiones de variables aleatorias en vez de sucesiones de números reales y expresa convergencia en probabilidad. Se define rigurosamente en el apéndice I.1.

**Corolario 4.12.** [3] Por lo visto en la observación 4.11, y usando el teorema de Slutsky, se tiene que los teoremas enunciados para  $\widehat{F}(x)$  también se cumplen para  $\widehat{F}_n(x)$ .

## 4.5. Contrastes sobre la función de distribución usando una estimación continua bajo censura aleatoria por la derecha

En esta sección se supone que se siguen cumpliendo las hipótesis enunciadas en la sección anterior. Con el objetivo de obtener un contraste de especificación sobre la función de distribución usando el estimador continuo de la misma que se acaba de introducir, se estudiarán las propiedades de la distancia cuadrática entre  $F$  y  $\widehat{F}$ :

$$J_n = \int \left( \widehat{F}(x) - F(x) \right)^2 dx.$$

Sean  $k = \mu_2(K)/2$  (donde  $\mu_2(K)$  se definió en la ecuación (4.4)),  $B(u) = \int W(s)W(s+u) ds$  ( $W$  se definió en la ecuación (4.3)),

$$d(n) = \begin{cases} nh^{-3/2} & \text{si } nh^3 \rightarrow 0, \\ n^{1/2}h^{-3} & \text{si } nh^3 \rightarrow \infty, \\ n^{3/2} & \text{si } nh^3 \rightarrow \theta \neq 0, \end{cases}$$

y sea  $Z$  una distribución normal estándar.

**Teorema 4.13.** [4] (Sobre la distribución asintótica de  $J_n$ ) Bajo las hipótesis enunciadas en la sección anterior, si  $nh \rightarrow \infty$  y  $h \rightarrow 0$  cuando  $n \rightarrow \infty$ , entonces

$$d(n)(J_n - \mathbb{E}[J_n]) \xrightarrow{d} \begin{cases} 2^{1/2}\sigma_1 Z & \text{si } nh^3 \rightarrow 0, \\ 2k\sigma_4 Z & \text{si } nh^3 \rightarrow \infty, \\ (2^{1/2}\theta^{1/2}\sigma_1 + 2\theta k\sigma_4)Z & \text{si } nh^3 \rightarrow \theta, \end{cases}$$

donde  $0 < \theta < \infty$  y

$$\sigma_1^2 = R \left( \frac{f(s)}{1 - G(s)} \right) R(B(s)),$$

$$\sigma_4^2 = \left( \int F''(s)^2 \frac{f(s)}{1 - G(s)} ds \right) R(B(s)) - \int F''(x)^2 F(x) dx,$$

donde  $R$  se definió en la ecuación (4.5).

*Observación 4.14.* La distancia cuadrática  $J_n$  se puede descomponer de la siguiente forma:

$$\begin{aligned} J_n &= \int \left( \widehat{F}(x) - F(x) \right)^2 dx \\ &= \int \left( \widehat{F}(x) - \mathbb{E} \left[ \widehat{F}(x) \right] + \mathbb{E} \left[ \widehat{F}(x) \right] - F(x) \right)^2 dx \\ &= \int \left( \widehat{F}(x) - \mathbb{E} \left[ \widehat{F}(x) \right] \right)^2 dx + \int \left( \mathbb{E} \left[ \widehat{F}(x) \right] - F(x) \right)^2 dx \\ &\quad + \int 2 \left( \widehat{F}(x) - \mathbb{E} \left[ \widehat{F}(x) \right] \right) \left( \mathbb{E} \left[ \widehat{F}(x) \right] - F(x) \right) dx \end{aligned} \tag{4.6}$$

y se tiene que

$$J_n - \mathbb{E}[J_n] = 2^{1/2}n^{-1}h^{3/2}\sigma_1 Z + 2n^{-1/2}h^3 k\sigma_4 Z + O_p(n^{-3/2}h),$$

donde el primer sumando viene de  $\hat{I}_1$  y el segundo de  $\hat{I}_3$ , siendo estos el primer y el tercer sumando de la descomposición de  $J_n$  de la ecuación (4.6). Se aprecia claramente que la distribución asintótica de  $J_n$  depende del parámetro ventana  $h$ :

Si  $nh^3 \rightarrow 0$ , se está aplicando poco suavizado y  $\hat{I}_1$  domina a  $\hat{I}_3$ , por lo que el comportamiento asintótico de  $J_n$  viene de  $\hat{I}_1$ .

Si se suaviza demasiado, esto es,  $nh^3 \rightarrow \infty$ , entonces  $\hat{I}_1$  es asintóticamente despreciable en comparación con  $\hat{I}_3$ . Luego la distribución asintótica de  $J_n$  centrado y escalado se determina a partir del término  $\hat{I}_3$ .

Por último, para un suavizado óptimo, que se da cuando  $nh^3 \rightarrow \theta$ , ninguno de los términos es asintóticamente dominante pues ambos tienen la misma magnitud. En este caso, la distribución límite de  $J_n$  centrado y escalado será la suma de las distribuciones de los dos términos comentados.

Haciendo uso del teorema 4.13, se puede plantear un contraste de especificación para una hipótesis nula simple de la forma

$$\begin{cases} H_0 : F(x) = F_0(x), \\ H_1 : F(x) \neq F_0(x). \end{cases}$$

Siguiendo el corolario 4.12 y el teorema 4.13, bajo  $H_0$  se obtiene el estadístico de contraste  $T_n$  sustituyendo  $\widehat{F}(t)$  por  $\widehat{\widehat{F}}_n(t)$  y  $F(t)$  por  $F_0(t)$  en la expresión de  $J_n$ . Además, centrándolo y escalándolo se llega a

$$T_{n,*} = d(n) \left( \int \left( \widehat{\widehat{F}}_n(x) - F_0(x) \right)^2 dx - c(n) \right),$$

donde

$$c(n) = \int \left( \mathbb{E}[\widehat{\widehat{F}}_n(x)] - F_0(x) \right)^2 dx + \sigma_2^2$$

y

$$\sigma_2^2 = n^{-1} \left( hR(W) \int \frac{f(t)}{1-G(t)} dt - h^2 \left[ \int \int K(s)K(s+v) ds dv \right] \int F(t)F(t+hv) dt \right).$$

Así, se tiene que

$$T_{n,*} / \sqrt{\text{Var}(T_{n,*})} \rightarrow N(0, 1),$$

donde

$$\text{Var}(T_{n,*}) = \begin{cases} (2^{1/2}\sigma_1)^2 & \text{si } nh^3 \rightarrow 0, \\ (2k\sigma_4)^2 & \text{si } nh^3 \rightarrow \infty, \\ (2^{1/2}\theta^{1/2}\sigma_1 + 2\theta^{1/3}k\sigma_4)^2 & \text{si } nh^3 \rightarrow \theta. \end{cases}$$

De este modo, para un nivel de significación  $\alpha$ , el contraste rechaza la hipótesis nula cuando  $T_{n,*} / \sqrt{\text{Var}(T_{n,*})} > z_\alpha$ , donde  $z_\alpha$  es el cuantil  $\alpha$  de una normal estándar.

## Capítulo 5

# Ilustración por simulación

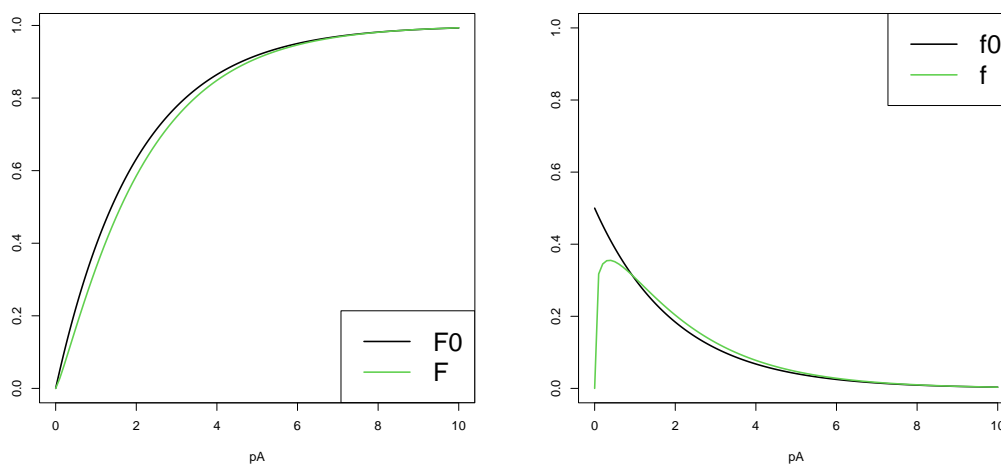
En este capítulo se incluye un estudio de simulación para ilustrar el comportamiento de algunas de las técnicas presentadas a lo largo del trabajo. En el Ejemplo 5.1 se ilustran los contrastes para datos completos e hipótesis nula simple introducidos en el Capítulo 1: los de Kolmogorov-Smirnov, Cramér-von Mises y Anderson-Darling basados en la función de distribución y el de Bickel y Rosenblatt basado en la función de densidad. En el Ejemplo 5.2 se ilustran los contrastes para datos censurados e hipótesis nula simple basados en la función de distribución introducidos en el Capítulo 3: las adaptaciones a datos censurados de los contrastes de Kolmogorov-Smirnov, Cramér-von Mises y Anderson-Darling.

**Ejemplo 5.1.** (Ilustración de contrastes para datos completos en el caso de una hipótesis nula simple). Se plantean los siguientes contrastes de especificación:

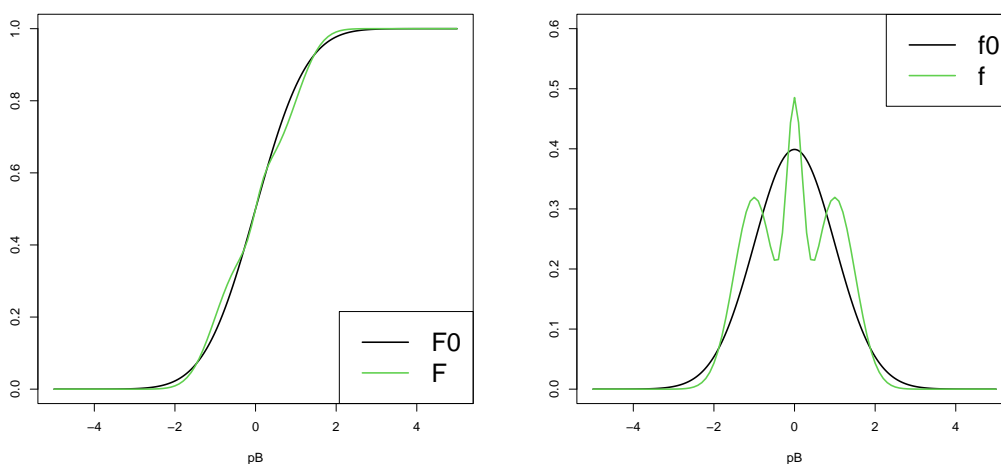
$$\begin{aligned} \text{Escenario A} & \begin{cases} H_0 : X \sim \text{Exp}(0.5), \\ H_1 : X \approx \text{Exp}(0.5), \end{cases} \\ \text{Escenario B} & \begin{cases} H_0 : X \sim N(0,1), \\ H_1 : X \approx N(0,1). \end{cases} \end{aligned}$$

Como hipótesis alternativas, para el primer contraste se generan muestras de una  $\text{Gamma}(1.2, 0.55)$  y para el segundo de una mixtura de distribuciones  $N(-1, 0.5)$ ,  $N(0, 0.2)$  y  $N(1, 0.5)$  con probabilidades 0.4, 0.2 y 0.4, respectivamente. Las gráficas de las funciones de distribución y densidad de las hipótesis nula y alternativa en cada escenario están representadas en la Figura 5.1

Para cada escenario se llevarán a cabo los contrastes de Kolmogorov-Smirnov (KS), Cramér-von Mises (CvM), Anderson-Darling (AD) y Bickel y Rosenblatt (BR) obteniendo los valores críticos de dos maneras: por el método de Montecarlo y usando las distribuciones asintóticas de los estadísticos.



(a) Escenario A



(b) Escenario B

Figura 5.1: Comparación de la hipótesis nula y la alternativa en cada escenario.

Las valores críticos asintóticos del contraste de Kolmogorov-Smirnov se consultaron en [27], los del contraste de Cramér-von Mises en [1] y los del contraste de Anderson-Darling en [9]. Los valores críticos asintóticos del contraste de Bickel y Rosenblatt se obtuvieron aproximando las fórmulas vistas en el Teorema 1.25. Para la estimación de la densidad se ha usado un kernel gaussiano y la ventana seleccionada por la regla del dedo (rule of thumb) [25, Subsección 3.4.2]. El código de R utilizado está en el Anexo I.4.

Los resultados se pueden ver en las tablas 5.2, 5.3, 5.4 y 5.5. En ellas se recogen las proporciones de rechazo tanto bajo la hipótesis nula como bajo la hipótesis alternativa de los distintos

contrastes en los dos escenarios y para varios tamaños muestrales, aproximando los valores críticos por el método de Montecarlo en las dos primeras y usando las distribuciones asintóticas de los estadísticos en las dos últimas.

| Tamaño muestral<br>$n$ | Proporción de rechazos bajo la hipótesis nula |       |       |       |             |       |       |       |
|------------------------|---|-------|-------|-------|-------------|-------|-------|-------|
|                        | Escenario A                                   |       |       |       | Escenario B |       |       |       |
|                        | KS  | CvM   | AD    | BR    | KS          | CvM   | AD    | BR    |
| 100                    | 0.042   | 0.047 | 0.052 | 0.049 | 0.051       | 0.054 | 0.057 | 0.047 |
| 200                    | 0.044   | 0.047 | 0.043 | 0.034 | 0.069       | 0.070 | 0.056 | 0.053 |
| 300                    | 0.031   | 0.047 | 0.043 | 0.041 | 0.040       | 0.048 | 0.050 | 0.058 |
| 400                    | 0.058   | 0.056 | 0.052 | 0.043 | 0.041       | 0.051 | 0.049 | 0.043 |
| 500                    | 0.054   | 0.051 | 0.050 | 0.052 | 0.051       | 0.056 | 0.057 | 0.056 |
| 600                    | 0.047   | 0.048 | 0.052 | 0.040 | 0.038       | 0.047 | 0.045 | 0.056 |
| 700                    | 0.042   | 0.049 | 0.045 | 0.061 | 0.060       | 0.060 | 0.061 | 0.064 |
| 800                    | 0.069   | 0.058 | 0.056 | 0.048 | 0.048       | 0.042 | 0.046 | 0.044 |

Tabla 5.2: Resultados obtenidos para un nivel de significación de 0.05 y  $M=1000$  muestras usando el método de Montecarlo.

Al estar aproximando los valores críticos por el método de Montecarlo, las proporciones de rechazo bajo la hipótesis nula no coinciden exactamente con el nivel de significación buscado (0.05), pero están razonablemente cerca como se puede comprobar en la Tabla 5.2.

| Tamaño muestral<br>$n$ | Proporción de rechazos bajo la hipótesis alternativa |       |       |       |             |       |       |       |
|------------------------|--|-------|-------|-------|-------------|-------|-------|-------|
|                        | Escenario A  |       |       |       | Escenario B |       |       |       |
|                        | KS   | CvM   | AD    | BR    | KS          | CvM   | AD    | BR    |
| 100                    | 0.263  | 0.269 | 0.278 | 0.247 | 0.202       | 0.142 | 0.185 | 0.440 |
| 200                    | 0.468  | 0.548 | 0.540 | 0.406 | 0.356       | 0.216 | 0.265 | 0.807 |
| 300                    | 0.585  | 0.675 | 0.709 | 0.551 | 0.459       | 0.317 | 0.444 | 0.957 |
| 400                    | 0.760  | 0.814 | 0.820 | 0.683 | 0.668       | 0.517 | 0.715 | 0.992 |
| 500                    | 0.870  | 0.908 | 0.934 | 0.835 | 0.789       | 0.648 | 0.812 | 1     |
| 600                    | 0.905  | 0.939 | 0.959 | 0.856 | 0.806       | 0.694 | 0.866 | 1     |
| 700                    | 0.947  | 0.967 | 0.981 | 0.924 | 0.933       | 0.908 | 0.977 | 1     |
| 800                    | 0.976  | 0.985 | 0.991 | 0.938 | 0.963       | 0.895 | 0.985 | 1     |

Tabla 5.3: Resultados obtenidos para un nivel de significación de 0.05 y  $M=1000$  muestras usando el método de Montecarlo.

En la Tabla 5.3 se aprecia que los contrastes sobre la función de distribución son más potentes en el escenario A (siendo el de Anderson-Darling el más potente), mientras que el contraste sobre

la función de densidad es más potente en el escenario B. Esto se puede explicar intuitivamente viendo las gráficas de las funciones de distribución y de densidad de la hipótesis nula y de la alternativa en cada escenario, que se representan en la Figura 5.1. Por un lado, en el escenario B las funciones de distribución son muy parecidas, mientras que las funciones de densidad se diferencian bastante. Por otro lado, en el escenario A, aunque no esté tan claro como en el B, los distintos estadísticos sobre la función de distribución son más sensibles a la discrepancia entre las gráficas de las funciones que el estadístico de Bickel y Rosenblatt sobre la función de densidad.

| Tamaño muestral<br>$n$ | Proporción de rechazos bajo la hipótesis nula |       |       |    |             |       |       |       |
|------------------------|---|-------|-------|----|-------------|-------|-------|-------|
|                        | Escenario A                                   |       |       |    | Escenario B |       |       |       |
|                        | KS  | CvM   | AD    | BR | KS          | CvM   | AD    | BR    |
| 100                    | 0.035   | 0.044 | 0.049 | 1  | 0.045       | 0.053 | 0.051 | 0.016 |
| 200                    | 0.040   | 0.037 | 0.037 | 1  | 0.066       | 0.062 | 0.057 | 0.024 |
| 300                    | 0.032   | 0.047 | 0.041 | 1  | 0.047       | 0.049 | 0.050 | 0.024 |
| 400                    | 0.050   | 0.057 | 0.057 | 1  | 0.035       | 0.041 | 0.044 | 0.023 |
| 500                    | 0.045   | 0.051 | 0.045 | 1  | 0.047       | 0.055 | 0.057 | 0.024 |
| 600                    | 0.046   | 0.053 | 0.048 | 1  | 0.045       | 0.058 | 0.054 | 0.023 |
| 700                    | 0.045   | 0.049 | 0.038 | 1  | 0.048       | 0.048 | 0.048 | 0.032 |
| 800                    | 0.050   | 0.047 | 0.047 | 1  | 0.053       | 0.052 | 0.051 | 0.021 |

Tabla 5.4: Resultados obtenidos para un nivel de significación de 0.05 y  $M=1000$  muestras usando las distribuciones asintóticas de los estadísticos.

| Tamaño muestral<br>$n$ | Proporción de rechazos bajo la hipótesis alternativa |       |       |    |             |       |       |       |
|------------------------|--|-------|-------|----|-------------|-------|-------|-------|
|                        | Escenario A  |       |       |    | Escenario B |       |       |       |
|                        | KS   | CvM   | AD    | BR | KS          | CvM   | AD    | BR    |
| 100                    | 0.236  | 0.266 | 0.269 | 1  | 0.183       | 0.138 | 0.168 | 0.321 |
| 200                    | 0.446  | 0.508 | 0.518 | 1  | 0.312       | 0.202 | 0.275 | 0.692 |
| 300                    | 0.598  | 0.675 | 0.703 | 1  | 0.475       | 0.326 | 0.446 | 0.890 |
| 400                    | 0.730  | 0.814 | 0.832 | 1  | 0.645       | 0.479 | 0.667 | 0.980 |
| 500                    | 0.852  | 0.903 | 0.925 | 1  | 0.772       | 0.642 | 0.817 | 0.999 |
| 600                    | 0.903  | 0.941 | 0.956 | 1  | 0.839       | 0.736 | 0.889 | 1     |
| 700                    | 0.951  | 0.966 | 0.978 | 1  | 0.918       | 0.847 | 0.964 | 1     |
| 800                    | 0.968  | 0.982 | 0.990 | 1  | 0.965       | 0.934 | 0.988 | 1     |

Tabla 5.5: Resultados obtenidos para un nivel de significación de 0.05 y  $M=1000$  muestras usando las distribuciones asintóticas de los estadísticos.

Es necesario destacar que las aproximaciones obtenidas utilizando la distribución asintótica

del estadístico de Bickel y Rosenblatt no son buenas, puesto que la convergencia se produce de forma muy lenta y para estos tamaños muestrales no es adecuado su uso. En particular, las proporciones de rechazo bajo la hipótesis nula no se parecen al nivel de significación que se trató de obtener, pues están alrededor de 0.02 para el escenario A y valen 1 para el escenario B, en lugar de acercarse a 0.05 como se pretendía (esto se refleja en la Tabla 5.4). Por tanto, las proporciones de rechazo recogidas en la Tabla 5.5 para el estadístico de Bickel y Rosenblatt no son fiables.

**Ejemplo 5.2.** (Contrastes basados en la función de distribución para datos con censura aleatoria por la derecha en el caso de una hipótesis nula simple). Se plantea el siguiente contraste de especificación:

$$\begin{cases} H_0 : F(x) = 1 - e^{0.5x}, \\ H_1 : F(x) \neq 1 - e^{0.5x}. \end{cases}$$

Esto es, la hipótesis nula afirma que la muestra proviene de una población  $Exp(0.5)$ . Para generar muestras con censura aleatoria por la derecha, que en este ejemplo tendrán todas un porcentaje similar de observaciones censuradas para simplificar las cuentas, se generan muestras de la variable de interés  $X$  que sigue una  $Exp(0.5)$  y de la variable de censura  $C$  a partir de una  $Exp(0.25)$ . De esta forma, al obtener la muestra censurada  $\{(T_i, \delta_i)\}_{i=1}^n$ , aproximadamente un tercio de las observaciones son tiempos de censura y el resto tiempos de vida. Esto se debe a que al seguir ambas variables aleatorias sendas distribuciones exponenciales, se pueden elegir los parámetros de forma que se obtenga la proporción de observaciones censuradas que se desee. En efecto, si  $X \sim Exp(\lambda_1)$  y  $C \sim Exp(\lambda_2)$ , entonces la probabilidad de que una observación quede censurada es

$$\begin{aligned} \mathbb{P}(C < X) &= \int_0^\infty g(t)(1 - F(t)) dt = \int_0^\infty \lambda_2 e^{-\lambda_2 t} e^{-\lambda_1 t} dt \\ &= \lambda_2 \int_0^\infty e^{-(\lambda_1 + \lambda_2)t} dt = \lambda_2 \left[ -\frac{1}{\lambda_1 + \lambda_2} e^{-(\lambda_1 + \lambda_2)t} \right]_0^\infty \\ &= \lambda_2 \left( 0 + \frac{1}{\lambda_1 + \lambda_2} \right) = \frac{\lambda_2}{\lambda_1 + \lambda_2}. \end{aligned}$$

De este modo, para  $\lambda_1 = \frac{1}{2}$  y  $\lambda_2 = \frac{1}{4}$  se tiene que  $\mathbb{P}(C < X) = \frac{1/4}{1/2+1/4} = \frac{1/4}{3/4} = \frac{1}{3}$ . En cuanto a la hipótesis alternativa, en este ejemplo se generan muestras de una distribución  $Exp(0.6)$  para la variable de interés y de una  $Exp(0.3)$  para la variable de censura, obteniendo de nuevo una muestra con aproximadamente un tercio de las observaciones censuradas. Las funciones de distribución y de densidad de la variable de interés para la hipótesis nula y la alternativa se comparan en la Figura 5.6.

En este ejemplo se realizan los contrastes de Kolmogorov-Smirnov (KS), Cramér-von Mises (CvM) y Anderson-Darling (AD) adaptados a datos censurados. En la Tabla 5.7 se muestran las proporciones de rechazo bajo la hipótesis nula y bajo la alternativa de cada uno de los

contrastes para varios tamaños muestrales. Se aprecia que el contraste de Cramér-von Mises es menos potente que los otros en este escenario y para estos tamaños muestrales, lo que puede deberse a que el fenómeno de la censura complica a este contraste la detección de la discrepancia existente entre las hipótesis nula y alternativa, mientras que los otros contrastes siguen siendo sensibles a ella. Además, las proporciones de rechazo bajo la hipótesis nula están cerca del nivel de significación de 0.05 que se pretendía obtener. El código de R utilizado está en el Anexo I.5.

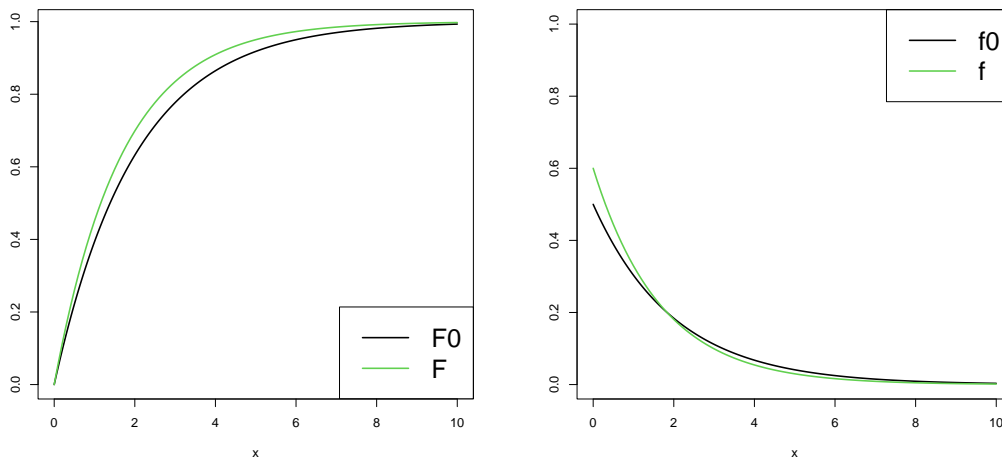


Tabla 5.6: Comparación de la hipótesis nula y la alternativa.

| Tamaño muestral $n$ | Prop. rechazos bajo $H_0$ |       |       | Prop. rechazos bajo $H_1$ |       |       |
|---------------------|---------------------------|-------|-------|---------------------------|-------|-------|
|                     | KS                        | CvM   | AD    | KS                        | CvM   | AD    |
| 100                 | 0.057                     | 0.056 | 0.057 | 0.285                     | 0.172 | 0.291 |
| 200                 | 0.068                     | 0.041 | 0.064 | 0.484                     | 0.192 | 0.512 |
| 300                 | 0.047                     | 0.069 | 0.051 | 0.647                     | 0.371 | 0.684 |
| 400                 | 0.052                     | 0.063 | 0.056 | 0.746                     | 0.388 | 0.765 |
| 500                 | 0.042                     | 0.045 | 0.040 | 0.815                     | 0.486 | 0.840 |
| 600                 | 0.041                     | 0.063 | 0.046 | 0.889                     | 0.539 | 0.926 |
| 700                 | 0.046                     | 0.068 | 0.032 | 0.926                     | 0.574 | 0.931 |
| 800                 | 0.037                     | 0.039 | 0.045 | 0.957                     | 0.604 | 0.976 |

Figura 5.7: Resultados obtenidos para un nivel de significación de 0.05 y  $M=1000$  muestras usando el método de Montecarlo.

# Anexo I

## Notación y código

### I.1. Notación

$X_n = o_p(b_n) \Leftrightarrow$  para todo  $\epsilon > 0$ , existen  $M > 0$  y  $N > 0$  tales que

$$\mathbb{P}\left(\left|\frac{X_n}{b_n}\right| > M\right) < \epsilon, \text{ para todo } n > N.$$

$$X_n = O_p(b_n) \Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{X_n}{b_n}\right| \geq \epsilon\right) = 0, \text{ para todo } \epsilon > 0.$$

### I.2. Código del ejemplo 1.19

```
#Comparación de la potencia de los estadísticos sobre la función de distribución
#Kolmogorov-Smirnov
#Cramér-von Mises
#Anderson-Darling

set.seed(2025) #Semilla para que los resultados sean reproducibles
x<-seq(0,1,length.out=100000); #Puntos en los que se van a dibujar las gráficas

#####
g0<-function(x) pbeta(x,0.2,0.8) #Función de distribución F0
f0<-function(x) dbeta(x,0.2,0.8) #Función de densidad f0
r0<-function(n) rbeta(n,0.2,0.8) #Función para generar una muestra aleatoria simple
# de tamaño n de la distribución F0
```

```

#Lo mismo para las alternativas:
g1<-function(x) pbeta(x-0.01,0.2,0.85)
f1<-function(x) dbeta(x-0.01,0.2,0.85)
r1<-function(n) rbeta(n,0.2,0.85)+0.01

g2<-function(x) pbeta(x,0.5,1)
f2<-function(x) dbeta(x,0.5,1)
r2<-function(n) rbeta(n,0.5,1)

g3<-function(x) pbeta(x,0.1,0.4)
f3<-function(x) dbeta(x,0.1,0.4)
r3<-function(n) rbeta(n,0.1,0.4)
#####

#Gráficas comparando las alternativas con la nula

#Escenario A
windows()
plot(x,g0(x),type='l',lwd=2,col=1,xlim=c(0,1),ylab="")
curve(g1(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='bottomright',legend=c('F0','F'),col=c(1,3),lwd=2,cex=2)

windows()
plot(x,f0(x),type='l',lwd=2,col=1,ylim=c(0,10),ylab="")
curve(f1(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='topright',legend=c('f0','f'),col=c(1,3),lwd=2,cex=2)

#Escenario B
windows()
plot(x,g0(x),type='l',lwd=2,col=1,xlim=c(0,1),ylab="")
curve(g2(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='bottomright',legend=c('F0','F'),col=c(1,3),lwd=2,cex=2)

windows()
plot(x,f0(x),type='l',lwd=2,col=1,ylim=c(0,10),ylab="")
curve(f2(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='topright',legend=c('f0','f'),col=c(1,3),lwd=2,cex=2)

```

```
#Escenario C
windows()
plot(x,g0(x),type='l',lwd=2,col=1,xlim=c(0,1),ylab="")
curve(g3(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='bottomright',legend=c('F0','F'),col=c(1,3),lwd=2,cex=2)

windows()
plot(x,f0(x),type='l',lwd=2,col=1,ylim=c(0,10),ylab="")
curve(f3(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='topright',legend=c('f0','f'),col=c(1,3),lwd=2,cex=2)

#####
N=c(seq(10,25,5),seq(50,125,25)); #Tamaños muestrales
a=c(0.05); #Niveles de significación
ejs=c(1,2,3); #Escenarios que se simulan
#(para no repetir todas las cuentas si sólo interesa ver alguno de ellos)
#####

#Se inicializan los valores críticos, p-valores y potencias
if(TRUE){
  nf=length(N);
  nc=length(a);

  qD<-matrix(0,nf,nc);
  nD<-matrix(0,nf,nc);
  qC<-matrix(0,nf,nc);
  nC<-matrix(0,nf,nc);
  qA<-matrix(0,nf,nc);
  nA<-matrix(0,nf,nc);

  if(sum(ejs==1)>0){
    PD1<-matrix(0,nf,nc);
    PC1<-matrix(0,nf,nc);
    PA1<-matrix(0,nf,nc);
  }

  if(sum(ejs==2)>0){
    PD2<-matrix(0,nf,nc);
```

```
PC2<-matrix(0,nf,nc);
PA2<-matrix(0,nf,nc);
}

if(sum(ejs==3)>0){
  PD3<-matrix(0,nf,nc);
  PC3<-matrix(0,nf,nc);
  PA3<-matrix(0,nf,nc);
}
}

#Para cada tamaño muestral
for (i in 1:nf){
  n=N[i]
  B=10000; #Número de muestras aleatorias que se generan

  #Se inicializan los vectores que almacenarán
  #los estadísticos evaluados en cada muestra
  D0=rep(0,B);
  C0=rep(0,B);
  A0=rep(0,B);

  D=rep(0,B);
  D1=rep(0,B);
  D2=rep(0,B);
  D3=rep(0,B);

  C=rep(0,B);
  C1=rep(0,B);
  C2=rep(0,B);
  C3=rep(0,B);

  A=rep(0,B);
  A1=rep(0,B);
  A2=rep(0,B);
  A3=rep(0,B);

  #Para cada muestra
```

```

for (b in 1:B){

  #Se generan dos muestras de la distribución nula
  #para verificar el nivel de significación de los contrastes
  x0<-r0(n);
  y<-r0(n);
  x0j<-sort(x0);
  xj<-sort(y);
  F0=g0(x0j);
  F=g0(xj);
  j=1:n;
  jn=j/n;
  j1n=(j-1)/n;

  #Evaluación del estadístico de KS sobre las muestras
  D01=jn-F0;
  D02=F0-j1n;
  d1=jn-F;
  d2=F-j1n;
  D0[b]=max(D01,D02);
  D[b]=max(d1,d2);

  #Evaluación del estadístico de CvM
  C0[b]=1/(12*n**2)+1/n*sum((F0-(2*j-1)/(2*n))**2);
  C[b]=1/(12*n**2)+1/n*sum((F-(2*j-1)/(2*n))**2);

  #Evaluación del estadístico de AD
  A0[b]=-1-1/(n**2)*sum((2*j-1)*log(F0)+(2*n+1-2*j)*log(1-F0));
  A[b]=-1-1/(n**2)*sum((2*j-1)*log(F)+(2*n+1-2*j)*log(1-F));

  #Se generan muestras aleatorias de las alternativas
  #y se evalúan los estadísticos sobre ellas para los escenarios escogidos

  #Escenario A
  if( sum(ejs==1)>0){
    y1<-r1(n);
    xj1<-sort(y1);
    F1=g0(xj1);
  }
}

```

```

D11=jn-F1;
D12=F1-j1n;
D1[b]=max(D11,D12);

C1[b]=1/(12*n**2)+1/n*sum((F1-(2*j-1)/(2*n))**2);

A1[b]=-1-1/(n**2)*sum((2*j-1)*log(F1)+(2*n+1-2*j)*log(1-F1));
}

#Escenario B
if(sum(ejs==2)>0){
  y2<-r2(n);
  xj2<-sort(y2);
  F2=g0(xj2);

  D21=jn-F2;
  D22=F2-j1n;
  D2[b]=max(D21,D22);

  C2[b]=1/(12*n**2)+1/n*sum((F2-(2*j-1)/(2*n))**2);

  A2[b]=-1-1/(n**2)*sum((2*j-1)*log(F2)+(2*n+1-2*j)*log(1-F2));
}

#Escenario C
if(sum(ejs==3)>0){
  y3<-r3(n);
  xj3<-sort(y3);
  F3=g0(xj3);

  D31=jn-F3;
  D32=F3-j1n;
  D3[b]=max(D31,D32);

  C3[b]=1/(12*n**2)+1/n*sum((F3-(2*j-1)/(2*n))**2);

  A3[b]=-1-1/(n**2)*sum((2*j-1)*log(F3)+(2*n+1-2*j)*log(1-F3));
}

```

```

}
}

#Para los distintos niveles de significación escogidos
#se obtienen los valores críticos del contraste
#y las potencias en los escenarios seleccionados
for (k in 1:nc){
  alpha=a[k];

  #Valores críticos
  qD[i,k]<-quantile(D0,1-alpha);
  nD[i,k]<-sum(D>quantile(D0,1-alpha))/B;
  qC[i,k]<-quantile(C0,1-alpha);
  nC[i,k]<-sum(C>quantile(C0,1-alpha))/B;
  qA[i,k]<-quantile(A0,1-alpha);
  nA[i,k]<-sum(A>quantile(A0,1-alpha))/B;

  #Escenario A
  if( sum(ejs==1)>0){
    PD1[i,k]<-sum(D1>quantile(D0,1-alpha))/B;
    PC1[i,k]<-sum(C1>quantile(C0,1-alpha))/B;
    PA1[i,k]<-sum(A1>quantile(A0,1-alpha))/B;
  }

  #Escenario B
  if( sum(ejs==2)>0){
    PD2[i,k]<-sum(D2>quantile(D0,1-alpha))/B;
    PC2[i,k]<-sum(C2>quantile(C0,1-alpha))/B;
    PA2[i,k]<-sum(A2>quantile(A0,1-alpha))/B;
  }

  #Escenario C
  if( sum(ejs==3)>0){
    PD3[i,k]<-sum(D3>quantile(D0,1-alpha))/B;
    PC3[i,k]<-sum(C3>quantile(C0,1-alpha))/B;
    PA3[i,k]<-sum(A3>quantile(A0,1-alpha))/B;
  }
}
}

```

```
}

#Potencias escenario A
View(cbind(PD1,PC1,PA1))

#Potencias escenario B
View(cbind(PD2,PC2,PA2))

#Potencias escenario C
View(cbind(PD3,PC3,PA3))

#Valores críticos
qD
qC
qA

#Niveles de significación observados
nD
nC
nA
```

### I.3. Código del ejemplo 1.28

```
#Potencia del estadístico sobre la función de densidad
#Bickel y Rosenblatt

set.seed(2025) #Semilla para que los resultados sean reproducibles
r<-seq(-3,3,length.out=100000); #Puntos en los que se van a dibujar las gráficas

#####
g0<-function(x) pnorm(x,0,1) #Función de distribución F0
f0<-function(x) dnorm(x,0,1) #Función de densidad f0
r0<-function(n) rnorm(n,0,1) #Función para generar una muestra aleatoria simple
# de tamaño n de la distribución F0

#Lo mismo para las alternativas:
```

```

probs<-c(0.4,0.2,0.4)
mus <- c(-1,0,1)
sds <- sqrt(c(0.5,0.2,0.5))
f1<-function(x) probs[1]*dnorm(x,mus[1],sds[1]**2)
  +probs[2]*dnorm(x,mus[2],sds[2]**2)+probs[3]*dnorm(x,mus[3],sds[3]**2)
g1<-Vectorize(function(x) integrate(f1,-Inf,x)$value)
r1<-function(n) {components <- sample(1:3,prob=probs,size=n,replace=TRUE)
  samples <- sds[components]*rnorm(n)+mus[components]}

g2<-function(x) pnorm(x,0,1.2)
f2<-function(x) dnorm(x,0,1.2)
r2<-function(n) rnorm(n,0,1.2)

g3<-function(x) pnorm(x,0.2,1)
f3<-function(x) dnorm(x,0.2,1)
r3<-function(n) rnorm(n,0.2,1)

#####

#Gráficas comparando las alternativas con la nula

#Escenario A
windows()
plot(r,g0(r),type='l',lwd=2,col=1,ylab="")
curve(g1(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='bottomright',legend=c('F0','F'),col=c(1,3),lwd=2,cex=2)

windows()
plot(r,f0(r),type='l',lwd=2,col=1,ylab="",ylim=c(0,0.6))
curve(f1(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='topright',legend=c('f0','f'),col=c(1,3),lwd=2,cex=2)

#Escenario B
windows()
plot(r,g0(r),type='l',lwd=2,col=1,ylab="")
curve(g2(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='bottomright',legend=c('F0','F'),col=c(1,3),lwd=2,cex=2)

```

```

windows()
plot(r,f0(r),type='l',lwd=2,col=1,ylab="")
curve(f2(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='topright',legend=c('f0','f'),col=c(1,3),lwd=2,cex=2)

#Escenario C
windows()
plot(r,g0(r),type='l',lwd=2,col=1,ylab="")
curve(g3(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='bottomright',legend=c('F0','F'),col=c(1,3),lwd=2,cex=2)

windows()
plot(r,f0(r),type='l',lwd=2,col=1,ylab="")
curve(f3(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='topright',legend=c('f0','f'),col=c(1,3),lwd=2,cex=2)

#####
N=seq(100,800,100); #Tamaños muestrales
alpha=0.05; #Nivel de significación
#####

#Se inicializan los vectores que almacenarán
#los valores críticos, p-valores y potencias
if(TRUE){
  nf=length(N);
  qR<-rep(0,nf);
  nR<-rep(0,nf);
  PR1<-rep(0,nf);
  PR2<-rep(0,nf);
  PR3<-rep(0,nf);
}

#Función que calcula el estadístico a partir de una muestra
dbr<-function(s) {

  #Se estima la densidad
  d<-density(s)
  fn<-function(x) (x>min(d$x) & x<max(d$x))*approxfun(d,rule=2)(x)

```

```
#Se calcula la esperanza de la estimación bajo la nula
kf0<-Vectorize(function(x) integrate(
  function(x,y) dnorm(x-y,0,d$bw)*f0(y),-Inf,Inf,x)$value)

#Se evalúa el estadístico
valor<-integrate(function(x) (fn(x)-kf0(x))**2,-Inf,Inf)$value
return(valor)
}

#Para cada tamaño muestral
for (i in 1:nf){
  n=N[i]
  B=1000; #Número de muestras aleatorias que se generan

  #Se inicializan los vectores que almacenarán los estadísticos
  #evaluados en cada muestra
  R0=rep(0,B);
  R=rep(0,B);
  R1=rep(0,B);
  R2=rep(0,B);
  R3=rep(0,B);

  for (b in 1:B){

    #Se genera una muestra de la distribución nula
    #que servirá para obtener los valores críticos
    R0[b]<-dbr(r0(n))

    #Se genera otra muestra de la nula que servirá
    #para verificar el nivel de significación
    R[b]<-dbr(r0(n))

    #Se generan muestras aleatorias de las alternativas
    #y se evalúa el estadístico sobre ellas para los escenarios escogidos
    R1[b]<-dbr(r1(n))
    R2[b]<-dbr(r2(n))
    R3[b]<-dbr(r3(n))
```

```

}

#Se obtienen los valores críticos del contraste y las potencias

#Valores críticos
qR[i]<-quantile(R0,1-alpha);

#Niveles
nR[i]<-sum(R>qR[i])/B;

#Potencias
PR1[i]<-sum(R1>qR[i])/B;
PR2[i]<-sum(R2>qR[i])/B;
PR3[i]<-sum(R3>qR[i])/B;
}

#Potencias
View(cbind(PR1,PR2,PR3))

#Valores críticos
qR

#Niveles de significación observados
nR

```

#### I.4. Código del ejemplo 5.1

```

set.seed(2025)#Semilla para que los resultados ssean reproducibles

###Escenario A
#Hipótesis nula
f0A<-function(x) dexp(x,0.5) #Función de densidad
FOA<-function(x) pexp(x,0.5) #Función de distribución
r0A<-function(n) rexp(n,0.5) #Función que genera muestras i.i.d.
#Hipótesis alternativa

```

```

f1A<-function(x) dgamma(x,1.2,0.55) #Función de densidad
F1A<-function(x) pgamma(x,1.2,0.55) #Función de distribución
r1A<-function(n) rgamma(n,1.2,0.55) #Función que genera muestras i.i.d.

#Representación gráfica de las funciones de distribución
pA<-seq(0,10,0.001) #Puntos sobre los que se dibuja
windows()
plot(pA,F0A(pA),type='l',lwd=2,col=1,ylab="")
curve(F1A(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='bottomright',legend=c('F0','F'),col=c(1,3),lwd=2,cex=2)
#Gráfica de las funciones de densidad
windows()
plot(pA,f0A(pA),type='l',lwd=2,col=1,ylab="",ylim=c(0,1))
curve(f1A(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='topright',legend=c('f0','f'),col=c(1,3),lwd=2,cex=2)

###Escenario B
#Hipótesis nula
f0B<-function(x) dnorm(x,0,1) #Función de densidad
F0B<-function(x) pnorm(x,0,1) #Función de distribución
r0B<-function(n) rnorm(n,0,1) #Función que genera muestras i.i.d.
#Hipótesis alternativa
#Al ser una mixtura de distribuciones normales se definen
probs<-c(0.4,0.2,0.4) #las probabilidades de que una observación
#provenga de cada distribución
mus<-c(-1,0,1) #las medias de las distribuciones
sds<-c(0.5,0.2,0.5) #las desviaciones típicas de las distribuciones
#Función de densidad
f1B<-function(x) probs[1]*dnorm(x,mus[1],sds[1])+
  probs[2]*dnorm(x,mus[2],sds[2])+probs[3]*dnorm(x,mus[3],sds[3])
#Función de distribución
F1B<-Vectorize(function(x) integrate(f1B,-Inf,x)$value)
#Función que genera muestras i.i.d.
r1B<-function(n) {components <- sample(1:3,prob=probs,size=n,replace=TRUE)
  samples <- sds[components]*rnorm(n)+mus[components]}

#Gráfica distribuciones

```

```

pB<-seq(-5,5,0.01) #Puntos en los que se dibuja
windows()
plot(pB,F0B(pB),type='l',lwd=2,col=1,ylab="")
curve(F1B(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='bottomright',legend=c('F0','F'),col=c(1,3),lwd=2,cex=2)
#Gráfica densidades
windows()
plot(pB,f0B(pB),type='l',lwd=2,col=1,ylab="",ylim=c(0,0.6))
curve(f1B(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='topright',legend=c('f0','f'),col=c(1,3),lwd=2,cex=2)

###Definición de funciones que calculan estadísticos
#Estadísticos sobre la función de distribución en el escenario A
DFA<-function(s) {
  sj<-sort(s) #Ordena la muestra
  Fj=FOA(sj) #Hace la transformación integral
  j=1:n
  jn=j/n
  j1n=jn-1/n
  D=rep(0,3)
  D[1]=max(jn-Fj,Fj-j1n) #Estadístico de Kolmogorov-Smirnov
  D[2]=1/(12*n**2)+1/n*sum((Fj-(2*j-1)/(2*n))**2) #Cramér-von Mises
  D[3]=-1-1/(n**2)*sum((2*j-1)*log(Fj)+(2*n+1-2*j)*log(1-Fj)) #Anderson-Darling
  return(D)}

#Estadístico de Bickel y Rosenblatt en el escenario A
DBRA<-function(s) {
  d<-density(s) #Estimación no paramétrica de la densidad
  #Se convierte en una función de R
  fn<-function(x) (x>min(d$x) & x<max(d$x))*approxfun(d,rule=2)(x)
  #Esperanza de la estimación
  kf0<-Vectorize(function(x) integrate(
    function(x,y) dnorm(x-y,0,d$bw)*f0A(y),-Inf,Inf,x)$value)
  valor<-integrate(function(x) (fn(x)-kf0(x))**2,-Inf,Inf)$value #Estadístico
  return(valor)
}

```

```

#Estadísticos sobre la función de distribución en el escenario B
DFB<-function(s) {
  sj<-sort(s) #Ordena la muestra
  Fj=F0B(sj) #Hace la transformación integral
  j=1:n
  jn=j/n
  j1n=jn-1/n
  D=rep(0,3)
  D[1]=max(jn-Fj,Fj-j1n) #Estadístico de Kolmogorov-Smirnov
  D[2]=1/(12*n**2)+1/n*sum((Fj-(2*j-1)/(2*n))**2) #Cramér-von Mises
  D[3]=-1-1/(n**2)*sum((2*j-1)*log(Fj)+(2*n+1-2*j)*log(1-Fj)) #Anderson-Darling
  return(D)}

#Estadístico de Bickel y Rosenblatt en el escenario B
DBRB<-function(s) {
  d<-density(s) #Estimación no paramétrica de la densidad
  #Se convierte en una función de R
  fn<-function(x) (x>min(d$x) & x<max(d$x))*approxfun(d,rule=2)(x)
  #Esperanza de la estimación
  kf0<-Vectorize(function(x) integrate(
    function(x,y) dnorm(x-y,0,d$bw)*f0B(y),-Inf,Inf,x)$value)
  valor<-integrate(function(x) (fn(x)-kf0(x))**2,-Inf,Inf)$value #Estadístico
  return(valor)
}

###Simulación
N=seq(100,800,100) #Tamaños muestrales
alpha=0.05 #Nivel de significación

#Inicialización de los vectores que almacenarán los valores críticos,
#niveles de significación observados y potencias observadas
if(TRUE){
  nf=length(N); #Número de muestras

  qKSA<-rep(0,nf) #Valores críticos estimados para el estadístico de KS
  #en el escenario A usando el método de Montecarlo
  nKSA<-rep(0,nf) #Niveles de significación observados para el estadístico de KS
  #en el escenario A usando el método de Montecarlo

```

```
naKSA<-rep(0,nf) #Niveles de significación observados para el estadístico de KS
#en el escenario A usando los valores críticos asintóticos
PKSA<-rep(0,nf) #Potencia observada para el estadístico de KS en el escenario A
#usando el método de Montecarlo
PaKSA<-rep(0,nf) #Potencia observada para el estadístico de KS en el escenario A
#usando los valores críticos asintóticos
```

```
#Análogo para el resto de contrastes
```

```
qCMA<-rep(0,nf)
nCMA<-rep(0,nf)
naCMA<-rep(0,nf)
PCMA<-rep(0,nf)
PaCMA<-rep(0,nf)
```

```
qADA<-rep(0,nf)
nADA<-rep(0,nf)
naADA<-rep(0,nf)
PADA<-rep(0,nf)
PaADA<-rep(0,nf)
```

```
qBRA<-rep(0,nf)
nBRA<-rep(0,nf)
naBRA<-rep(0,nf)
PBRA<-rep(0,nf)
PaBRA<-rep(0,nf)
```

```
#Lo mismo para el escenario B
```

```
qKSB<-rep(0,nf)
nKSB<-rep(0,nf)
naKSB<-rep(0,nf)
PKSB<-rep(0,nf)
PaKSB<-rep(0,nf)
```

```
qCMB<-rep(0,nf)
nCMB<-rep(0,nf)
naCMB<-rep(0,nf)
PCMB<-rep(0,nf)
PaCMB<-rep(0,nf)
```

```
qADB<-rep(0,nf)
nADB<-rep(0,nf)
naADB<-rep(0,nf)
PADB<-rep(0,nf)
PaADB<-rep(0,nf)

qBRB<-rep(0,nf)
nBRB<-rep(0,nf)
naBRB<-rep(0,nf)
PBRB<-rep(0,nf)
PaBRB<-rep(0,nf)
}

for (i in 1:nf){ #Para cada tamaño muestral
  n=N[i]
  B=10000 #Número de muestras que se generan

  #Se inicializan vectores que van a acumular las evaluaciones de
  #los distintos estadísticos para este tamaño muestral

  #Escenario A
  #Sobre muestras de la hipótesis nula para aproximar
  #los valores críticos por Montecarlo
  KSOA=rep(0,B)
  CMOA=rep(0,B)
  ADOA=rep(0,B)
  BROA=rep(0,B)

  #Sobre otras muestras de la nula para verificar los niveles de significación
  KSaA=rep(0,B)
  CMaA=rep(0,B)
  ADaA=rep(0,B)
  BRaA=rep(0,B)

  #Sobre muestras de la alternativa para obtener aproximaciones de las potencias
  KS1A=rep(0,B)
  CM1A=rep(0,B)
```

```
AD1A=rep(0,B)
BR1A=rep(0,B)

#Lo mismo para el escenario B
KSOB=rep(0,B)
CMOB=rep(0,B)
ADOB=rep(0,B)
BROB=rep(0,B)

KSaB=rep(0,B)
CMaB=rep(0,B)
ADaB=rep(0,B)
BRaB=rep(0,B)

KS1B=rep(0,B)
CM1B=rep(0,B)
AD1B=rep(0,B)
BR1B=rep(0,B)

for (b in 1:B){

  x0A<-r0A(n) #Se genera una muestra de la nula
  DOA<-DFA(x0A) #Se evalúan para esta muestra los estadísticos
  #de los contrastes sobre la distribución
  #Se almacenan en los vectores correspondientes
  KSOA[b]<-DOA[1]
  CMOA[b]<-DOA[2]
  ADOA[b]<-DOA[3]
  BROA[b]<-DBRA(x0A) #Se evalúa para esta muestra el estadístico del
  #contraste sobre la densidad

  #Lo mismo para otra muestra de la nula
  xaA<-r0A(n)
  DaA<-DFA(xaA)
  KSaA[b]<-DaA[1]
  CMaA[b]<-DaA[2]
  ADaA[b]<-DaA[3]
```

```

BRaA[b]<-DBRA(xaA)
#Se usa la distribución asintótica del estadístico de BR para aproximar
#el nivel de significación del contraste
#Se calculan los coeficientes normalizantes
h<-density(xaA)$bw #Parámetro ventana por la regla del dedo
muk<-density(xaA,give.Rkern = TRUE)/(n*h)
convk<-Vectorize(function(u) integrate(
function(u,v) dnorm(u+v,0,h)*dnorm(v,0,h),-Inf,Inf,u)$value)
intconvk<-integrate(function(u) convk(u)**2,-Inf,Inf)$value
sigmakf0<-sqrt(2*intconvk*integrate(function(x) f0A(x)**2,-Inf,Inf)$value)
qalpha<-muk+qnorm(1-alpha)*sigmakf0/(n*sqrt(h)) #Valor crítico asintótico
#para un nivel alpha
naBRA[i]<-naBRA[i]+(BRaA[b]>qalpha)/B #Actualización de
#la aproximación del nivel de significación para este tamaño muestral

#Lo mismo para la alternativa
x1A<-r1A(n)
D1A<-DFA(x1A)
KS1A[b]<-D1A[1]
CM1A[b]<-D1A[2]
AD1A[b]<-D1A[3]
BR1A[b]<-DBRA(x1A)
#Se usa la distribución asintótica del estadístico de BR
#para aproximar la potencia del contraste
#Mismas cuentas que para el nivel de significación
h<-density(x1A)$bw
muk<-density(x1A,give.Rkern = TRUE)/(n*h)
convk<-Vectorize(function(u) integrate(
function(u,v) dnorm(u+v,0,h)*dnorm(v,0,h),-Inf,Inf,u)$value)
intconvk<-integrate(function(u) convk(u)**2,-Inf,Inf)$value
sigmakf0<-sqrt(2*intconvk*integrate(function(x) f0A(x)**2,-Inf,Inf)$value)
qalpha<-muk+qnorm(1-alpha)*sigmakf0/(n*sqrt(h))
PaBRA[i]<-PaBRA[i]+(BR1A[b]>qalpha)/B

#Lo mismo para el escenario B
x0B<-r0B(n)
DOB<-DFB(x0B)

```

```

KSOB[b]<-DOB[1]
CMOB[b]<-DOB[2]
ADOB[b]<-DOB[3]
BROB[b]<-DBRB(xOB)

xaB<-rOB(n)
DaB<-DFB(xaB)
KSaB[b]<-DaB[1]
CMaB[b]<-DaB[2]
ADaB[b]<-DaB[3]
BRaB[b]<-DBRB(xaB)
h<-density(xaB)$bw
muk<-density(xaB,give.Rkern = TRUE)/(n*h)
convk<-Vectorize(function(u) integrate(
function(u,v) dnorm(u+v,0,h)*dnorm(v,0,h),-Inf,Inf,u)$value)
intconvk<-integrate(function(u) convk(u)**2,-Inf,Inf)$value
sigmakf0<-sqrt(2*intconvk*integrate(function(x) fOB(x)**2,-Inf,Inf)$value)
qalpha<-muk+qnorm(1-alpha)*sigmakf0/(n*sqrt(h))
naBRB[i]<-naBRB[i]+(BRaB[b]>qalpha)/B

x1B<-r1B(n)
D1B<-DFB(x1B)
KS1B[b]<-D1B[1]
CM1B[b]<-D1B[2]
AD1B[b]<-D1B[3]
BR1B[b]<-DBRB(x1B)
h<-density(x1B)$bw
muk<-density(x1B,give.Rkern = TRUE)/(n*h)
convk<-Vectorize(function(u) integrate(
function(u,v) dnorm(u+v,0,h)*dnorm(v,0,h),-Inf,Inf,u)$value)
intconvk<-integrate(function(u) convk(u)**2,-Inf,Inf)$value
sigmakf0<-sqrt(2*intconvk*integrate(function(x) fOB(x)**2,-Inf,Inf)$value)
qalpha<-muk+qnorm(1-alpha)*sigmakf0/(n*sqrt(h))
PaBRB[i]<-PaBRB[i]+(BR1B[b]>qalpha)/B
}

#Escenario A
#Aproximación de los valores críticos por Montecarlo

```

```
qKSA[i]<-quantile(KSOA,1-alpha)
qCMA[i]<-quantile(CMOA,1-alpha)
qADA[i]<-quantile(ADOA,1-alpha)
qBRA[i]<-quantile(BROA,1-alpha)

#Aproximación de los niveles de significación por Montecarlo
nKSA[i]<-sum(KSaA>qKSA[i])/B
nCMA[i]<-sum(CMaA>qCMA[i])/B
nADA[i]<-sum(ADaA>qADA[i])/B
nBRA[i]<-sum(BRaA>qBRA[i])/B

#Aproximación de las potencias por Montecarlo
PKSA[i]<-sum(KS1A>qKSA[i])/B
PCMA[i]<-sum(CM1A>qCMA[i])/B
PADA[i]<-sum(AD1A>qADA[i])/B
PBRA[i]<-sum(BR1A>qBRA[i])/B

#Aproximación del nivel de significación de los contrastes sobre la
#distribución usando las distribuciones asintóticas
naKSA[i]<-sum(KSaA>1.36/sqrt(n))/B
naCMA[i]<-sum(CMaA>0.46136/n)/B
naADA[i]<-sum(ADaA>2.492/n)/B

#Aproximación de la potencia de los contrastes sobre la
#distribución usando las distribuciones asintóticas
PaKSA[i]<-sum(KS1A>1.36/sqrt(n))/B
PaCMA[i]<-sum(CM1A>0.46136/n)/B
PaADA[i]<-sum(AD1A>2.492/n)/B

#Lo mismo para el escenario B
qKSB[i]<-quantile(KSOB,1-alpha)
qCMB[i]<-quantile(CMOB,1-alpha)
qADB[i]<-quantile(ADOB,1-alpha)
qBRB[i]<-quantile(BROB,1-alpha)

nKSB[i]<-sum(KSaB>qKSB[i])/B
nCMB[i]<-sum(CMaB>qCMB[i])/B
```

```

nADB[i]<-sum(ADaB>qADB[i])/B
nBRB[i]<-sum(BRaB>qBRB[i])/B

PKSB[i]<-sum(KS1B>qKSB[i])/B
PCMB[i]<-sum(CM1B>qCMB[i])/B
PADB[i]<-sum(AD1B>qADB[i])/B
PBRB[i]<-sum(BR1B>qBRB[i])/B

#Asintótico
naKSB[i]<-sum(KSaB>1.36/sqrt(n))/B
naCMB[i]<-sum(CMaB>0.46136/n)/B
naADB[i]<-sum(ADaB>2.492/n)/B

PaKSB[i]<-sum(KS1B>1.36/sqrt(n))/B
PaCMB[i]<-sum(CM1B>0.46136/n)/B
PaADB[i]<-sum(AD1B>2.492/n)/B
}

#Escenario A
#Aproximación de las potencias por el método de Montecarlo
View(cbind(PKSA,PCMA,PADA,PBRA))
#Aproximación de los niveles de significación por el método de Montecarlo
View(cbind(nKSA,nCMA,nADA,nBRA))

#Aproximación de las potencias usando las distribuciones asintóticas
View(cbind(PaKSA,PaCMA,PaADA,PaBRA))
#Aproximación de los niveles de significación usando las distribuciones asintóticas
View(cbind(naKSA,naCMA,naADA,naBRA))

#Lo mismo para el escenario B
View(cbind(PKSB,PCMB,PADB,PBRB))
View(cbind(nKSB,nCMB,nADB,nBRB))

View(cbind(PaKSB,PaCMB,PaADB,PaBRB))
View(cbind(naKSB,naCMB,naADB,naBRB))

```

## I.5. Código del ejemplo 5.2

```
#Potencia de los contrastes basados en la función de distribución
#para muestras con censura aleatoria por la derecha
#Kolmogorov-Smirnov
#Cramér-von Mises
#Anderson-Darling

library(survival)
set.seed(2025)#Semilla para que los resultados ssean reproducibles

#Hipótesis nula
f0<-function(x) dexp(x,0.5) #Función de densidad
F0<-function(x) pexp(x,0.5) #Función de distribución
r0Y<-function(n) rexp(n,0.5) #Función que genera muestras i.i.d.
#de la variable tiempo de vida
r0C<-function(n) rexp(n,0.25) #Función que genera muestras i.i.d. de la variable
#tiempo de censura, escogiendo el parámetro de forma que se censuren
#aproximadamente un tercio de las observaciones
#Función que genera muestras i.i.d. de la variable tiempo observado
r0Z<-function(n) {y<-r0Y(n) #Se genera una muestra de tiempos de vida
  c<-r0C(n) #Se genera una muestra de tiempos de censura
  #Se inicializan los vectores que almacenarán los tiempos observados
  #y los indicadores de si el tiempo observado es tiempo de vida o de censura
  z<-rep(0,n)
  d<-rep(0,n)
  for(i in 1:n){ #Para cada individuo
    z[i]<-min(y[i],c[i]) #Se apunta el tiempo observado
    d[i]<-1-(c[i]<y[i])*1 #Se apunta si es un tiempo de vida (1) o censura (0)
  }
  zd<-Surv(z,d) #Se guarda en un solo vector con el formato adecuado
  #(Para los tiempos de vida almacena el valor numérico y para los
  #de censura pone el valor y un símbolo '+' )
  return(zd)}

#Hipótesis alternativa
f1<-function(x) dexp(x,0.6) #Función de densidad
F1<-function(x) pexp(x,0.6) #Función de distribución
```

```

r1Y<-function(n) rexp(n,0.6) #Función que genera muestras i.i.d. de la variable
#tiempo de vida
r1C<-function(n) rexp(n,0.3) #Función que genera muestras i.i.d. de la variable
#tiempo de censura, escogiendo el parámetro de forma que se censuren
#aproximadamente un tercio de las observaciones
#Función que genera muestras i.i.d. de la variable tiempo observado
r1Z<-function(n) {y<-r1Y(n) #Se genera una muestra de tiempos de vida
  c<-r1C(n) #Se genera una muestra de tiempos de censura
  #Se inicializan los vectores que almacenarán los tiempos observados
  #y los indicadores de si el tiempo observado es tiempo de vida o de censura
  z<-rep(0,n)
  d<-rep(0,n)
  for(i in 1:n){ #Para cada individuo
    z[i]<-min(y[i],c[i]) #Se apunta el tiempo observado
    d[i]<-1-(c[i]<y[i])*1 #Se apunta si es un tiempo de vida (1) o censura (0)
  }
  zd<-Surv(z,d) #Se guarda en un solo vector con el formato adecuado
  #(Para los tiempos de vida almacena el valor numérico y para los
  #de censura pone el valor y un símbolo '+')
  return(zd)}

#Representación gráfica de las funciones de distribución
p0<-seq(0,10,0.001) #Puntos sobre los que se dibuja
windows()
plot(p0,F0(p0),type='l',lwd=2,col=1,ylab="")
curve(F1(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='bottomright',legend=c('F0','F'),col=c(1,3),lwd=2,cex=2)
#Gráfica de las funciones de densidad
windows()
plot(p0,f0(p0),type='l',lwd=2,col=1,ylab="",ylim=c(0,1))
curve(f1(x),type='l',lwd=2,col=3,add=TRUE)
legend(x='topright',legend=c('f0','f'),col=c(1,3),lwd=2,cex=2)

###Definición de la funciones que calcula los estadísticos
DF<-function(s) { #Dada una muestra con censura aleatoria por la derecha
  km<-survfit(s~1) #Se calcula el estimador de Kaplan-Meier para la muestra
  zj<-summary(km)$time #Tiempos de vida ordenados

```

```

d=length(zj) #Número de tiempos de vida observados

F0j=F0(zj) #Se hace la transformación integral sobre los tiempos de vida ordenados
F0j1<-c(0,F0j[1:(d-1)]) #Se desplaza un lugar a la derecha
Fnj<-1-summary(km)$surv #Evaluación del estimador Kaplan-Meier de la
#función de distribución sobre los tiempos de vida observados ordenados
Fnj1<-c(0,Fnj[1:(d-1)]) #Se desplaza un lugar a la derecha

D=rep(0,3)
D[1]=max(Fnj-F0j,F0j-Fnj1) #Estadístico de Kolmogorov-Smirnov
D[2]=d/3+d*sum(c(0,Fnj)*(c(F0(zj),1)
-c(0,F0j))*(c(0,Fnj)-(c(F0(zj),1)-c(0,F0j)))) #Cramér-von Mises
D[3]=-d+d**sum((Fnj1**2-Fnj**2)*log(F0j)-
((1-Fnj1)**2-(1-Fnj)**2)*log(1-F0j)) #Anderson-Darling
return(D)}

###Simulación
N=seq(100,800,100) #Tamaños muestrales
alpha=0.05 #Nivel de significación

#Inicialización de los vectores que almacenarán los valores críticos,
#niveles de significación observados y potencias observadas
if(TRUE){
  nf=length(N); #Número de muestras

  qKS<-rep(0,nf) #Valores críticos estimados para el estadístico de KS
  #usando el método de Montecarlo
  nKS<-rep(0,nf) #Niveles de significación observados para el estadístico de KS
  #usando el método de Montecarlo
  PKS<-rep(0,nf) #Potencia observada para el estadístico de KS
  #usando el método de Montecarlo

  #Análogo para el resto de contrastes
  qCM<-rep(0,nf)
  nCM<-rep(0,nf)
  PCM<-rep(0,nf)

  qAD<-rep(0,nf)

```

```
nAD<-rep(0,nf)
PAD<-rep(0,nf)
}

for (i in 1:nf){ #Para cada tamaño muestral
  n=N[i]
  B=1000 #Número de muestras que se generan

  #Se inicializan vectores que van a acumular las evaluaciones de
  #los distintos estadísticos para este tamaño muestral

  #Sobre muestras de la hipótesis nula para aproximar
  #los valores críticos por Montecarlo
  KS0=rep(0,B)
  CM0=rep(0,B)
  AD0=rep(0,B)

  #Sobre otras muestras de la nula para verificar los niveles de significación
  KSa=rep(0,B)
  CMa=rep(0,B)
  ADa=rep(0,B)

  #Sobre muestras de la alternativa para obtener aproximaciones de las potencias
  KS1=rep(0,B)
  CM1=rep(0,B)
  AD1=rep(0,B)

  for (b in 1:B){

    x0<-r0Z(n) #Se genera una muestra censurada de la hipótesis nula
    D0<-DF(x0) #Se evalúan para esta muestra los estadísticos
    #de los contrastes
    #Se almacenan en los vectores correspondientes
    KS0[b]<-D0[1]
    CM0[b]<-D0[2]
    AD0[b]<-D0[3]

    #Lo mismo para otra muestra de la nula
```

```
xa<-rOZ(n)
Da<-DF(xa)
KSa[b]<-Da[1]
CMa[b]<-Da[2]
ADa[b]<-Da[3]

#Lo mismo para la alternativa
x1<-r1Z(n)
D1<-DF(x1)
KS1[b]<-D1[1]
CM1[b]<-D1[2]
AD1[b]<-D1[3]
}

#Aproximación de los valores críticos
qKS[i]<-quantile(KS0,1-alpha)
qCM[i]<-quantile(CM0,1-alpha)
qAD[i]<-quantile(AD0,1-alpha)

#Aproximación de los niveles de significación
nKS[i]<-sum(KSa>qKS[i])/B
nCM[i]<-sum(CMa>qCM[i])/B
nAD[i]<-sum(ADa>qAD[i])/B

#Aproximación de las potencias
PKS[i]<-sum(KS1>qKS[i])/B
PCM[i]<-sum(CM1>qCM[i])/B
PAD[i]<-sum(AD1>qAD[i])/B
}

#Aproximación de las potencias
View(cbind(PKS,PCM,PAD))
#Aproximación de los niveles de significación
View(cbind(nKS,nCM,nAD))
```



# Bibliografía

- [1] Anderson, T. W. y Darling, D. A. (1952). *Asymptotic theory of certain goodness of fit criteria based on stochastic processes*, Annals of Mathematical Statistics, 23, 193–212.
- [2] Anderson, T. W. y Darling, D. A. (1954). *A Test of Goodness of Fit*, Journal of the American Statistical Association, 49, 765–769.
- [3] Bagkavos, D. e Ioannides, D. (2012). *Smooth confidence intervals for the survival function under random right censoring*, Electronic Journal of Statistics, 6, 843-860.
- [4] Bagkavos, D., Ioannides, D. y Kalamatianou, A. (2013). *A goodness of fit test for the survival function under random right censoring*, Electronic Journal of Statistics, 7, 2550-2576.
- [5] Bickel, P. J. y Rosenblatt, M. (1973). *On Some Global Measures of the Deviations of Density Function Estimates*, Ann. Statist., 1 (6), 1071 - 1095.
- [6] Birnbaum, Z. W. (1953). *Distribution-free Tests of fit for Continuous Distribution Functions*, Annals of Mathematical Statistics, 24, 1-8.
- [7] Bothma, E., Allison, J. S. y Visagie, I. J. H. (2024). *On the effect of the Kaplan-Meier estimator's assumed tail behavior on goodness-of-fit testing*, Communications in Statistics - Simulation and Computation, 1–27.
- [8] Csörgő, M., Gombay, E. y Horváth, L. (1991). *Central Limit Theorems for  $L_p$  Distances of Kernel Estimators of Densities Under Random Censorship*, The Annals of Statistics, 19(4), 1813–1831.
- [9] D'Agostino, R.B. y Stephens, M.A., eds. (1986). *Goodness-of-Fit Techniques*, Marcel Dekker, Inc.
- [10] Efron, B. (1981). *Censored data and the bootstrap*, Journal of the American Statistical Association 76 (374), 312–9.
- [11] Efron, B. y Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall.

- 
- [12] Fan, Y. (1994). *Testing the Goodness of Fit of a Parametric Density Function by Kernel Method*, *Econometric Theory*, 10(2), 316–356.
- [13] Gill, R.D. (1980). *Censoring and Stochastic Integrals*, Mathematical Centre Tracts 124, Matematisch Centrum.
- [14] Kaplan, E. L. y Meier, P. (1958). *Nonparametric Estimation from Incomplete Observations*, *Journal of the American Statistical Association*, 53(282), 457–481.
- [15] Klein, J. P. y Moeschberger, M. L. (2003). *Survival Analysis. Techniques for Censored and Truncated Data*, Springer.
- [16] Kolmogorov, A. (1933). *Sulla determinazione empirica di una legge di distribuzione*, *Giornale dell'Istituto italiano degli actuari*, 4, 83-91.
- [17] Kvam, P.H. y Vidakovic B. (2007) *Nonparametric Statistics with Applications to Science and Engineering*, John Wiley and Sons.
- [18] Lilliefors, H. W. (1967). *On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown*, *Journal of the American Statistical Association*, 62(318), 399–402.
- [19] Lilliefors, H. W. (1969). *On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown*, *Journal of the American Statistical Association*, 64(325), 387–389.
- [20] Masey, F. (1951). *The Kolmogorov-Smirnov test for goodness of fit*, *Journal of the American Statistical Association*, 46, 68-78.
- [21] Moore, D.F. (2016). *Applied Survival Analysis Using R*, Springer.
- [22] Nadaraya, E. A. (1964). *Some New Estimates for Distribution Functions*, *Theory of Probability and Its Applications*, 9(3), 497-500.
- [23] Nikulin, M., Lemesko, B., Chimitova, E.V. y Tsvinskaya, A. (2011). *Nonparametric Goodness-of-Fit Tests for Censored Data*.
- [24] Rosenblatt, M. (1975). *A Quadratic Measure of Deviation of Two-Dimensional Density Estimates and A Test of Independence*, *The Annals of Statistics*, 3(1), 1–14.
- [25] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall/CRC.
- [26] Simonoff, J. S. (1996). *Smoothing Methods in Statistics*, Springer.
- [27] Smirnov, N. (1948). *Table for Estimating the Goodness of Fit of Empirical Distributions*, *Annals of Mathematical Statistics*, 19, 279-281.

- 
- [28] Stute, W. y Wang, J.-L. (1993). *The Strong Law under Random Censorship*, The Annals of Statistics, 21(3), 1591–1607.
- [29] Vélez Ibarrola, R. y García Pérez, A. (2012). *Principios de Inferencia Estadística*, UNED.
- [30] Wand, M. P. y Jones, M.C. (1995). *Kernel Smoothing*, Chapman and Hall/CRC.
- [31] Wasserman, L. (2006). *All of Statistics. A Concise Course in Statistical Inference*, Springer.
- [32] Watson, G. N. (1922). *A Treatise on the Theory of Bessel Functions*, Cambridge University Press.