Universidad de Oviedo

Bondad de ajuste a familias de localización-escala

Sergio Esteban Bruña

TRABAJO DE FIN DE GRADO

Tutorizado por: Marina Iturrate Bobes y Raúl Pérez Fernández

UNIVERSIDAD DE OVIEDO
Facultad de Ciencias
Grado en Matemáticas

Mayo de 2025

Agradecimientos

Dedicado a mi familia, en especial a mis padres, Lorenzo y Carmen, y a mi hermana, Alba, por su amor, ánimos e incondicional apoyo a lo largo de mi vida, sobre todo en los momentos más desafiantes y por ser un ejemplo de perseverancia.

A mis amigos, mostrar mi agradecimiento por formar parte del camino, escuchar mis dudas existenciales y aparecer tanto en las pequeñas victorias como en las derrotas.

Y a todas las personas que, de una forma u otra, tejieron parte de mi historia: maestros, mentores, conocidos e incluso aquellos que, sin saberlo, me enseñaron a través de los obstáculos.

Sin todos vosotros, esto no habría sido posible.

Dulce bellum inexpertis, expertus metuit.

— Píndaro

Índice general

Índice de tablas

Índice de figuras

CAPÍTULO 1

Introducción

1.1. Motivación

En el análisis estadístico, la evaluación de la bondad de ajuste de un conjunto de datos a una familia paramétrica dada es un problema fundamental, especialmente cuando se trabaja con distribuciones de localización-escala. Estas familias, que incluyen modelos ampliamente utilizados como la normal, la logística y la distribución exponencial desplazada, son de particular interés debido a su flexibilidad y aplicabilidad en diversos campos: economía (en riesgo financiero y seguros), ingeniería (en control de calidad y fiabilidad) y hidrología (para modelar eventos extremos).

El estudio de la bondad de ajuste en modelos estadísticos tiene sus raíces en los trabajos pioneros de Pearson a principios del siglo XX, con la introducción de la prueba chi-cuadrado. Sin embargo, fue en las décadas siguientes cuando Kolmogorov [20] y Smirnov [28] desarrollaron pruebas basadas en la distancia entre la función de distribución empírica y la teórica, sentando las bases para los métodos modernos de bondad de ajuste, a partir de la teoría de procesos empíricos. Sin embargo, estos avances iniciales asumían parámetros conocidos, lo que limitaba su aplicabilidad en situaciones prácticas. Cuando los parámetros de estas distribuciones deben ser estimados a partir de los datos, la validación de la adecuación del modelo se vuelve más compleja.

La necesidad de abordar este problema llevó, en la segunda mitad del siglo XX, a la adaptación de las pruebas clásicas para el caso de parámetros estimados. Autores como Lilliefors [21] y Stephens [29] realizaron contribuciones fundamentales al estudiar el comportamiento de las pruebas de Kolmogorov-Smirnov [20, 28], Cramér-von Mises [9, 31] y Anderson-Darling [2] bajo estimación de parámetros, particularmente para distribuciones normales y exponenciales.

En las últimas décadas, una nueva vertiente en el estudio de la bondad de ajuste ha cobrado relevancia: la construcción de pruebas basadas en coeficientes de asimetría y curtosis. Este enfoque, impulsado por los trabajos de Brys, Hubert y Struyf [6], ha proporcionado avances significativos al generalizar pruebas clásicas como la de Jarque-Bera [18, 19].

A pesar de estos avances, persisten desafíos y este trabajo se enfoca en plantarles cara, proponiendo un marco metodológico riguroso para evaluar la bondad de ajuste en familias de localización-escala con parámetros estimados, considerando el desempeño estadístico de las pruebas. A través de este estudio, se busca contribuir tanto al avance teórico como a la práctica aplicada.

1.2. Objetivos

Retomando lo mencionado en la sección inmediatamente anterior, el estudio teórico de los tests de bondad de ajuste y sus aplicaciones es de gran interés. Es sino por esto que este trabajo tiene como objetivo comparar la potencia de los tests de Kolmogorov-Smirnov, Cramér-von Mises y Anderson-Darling en el contexto de familias de localización-escala, evaluando su desempeño bajo distintos tamaños muestrales y diferentes métodos de estimación de parámetros. Mediante un estudio de simulación exhaustivo, se busca determinar cuál de estos estadísticos ofrece mayor sensibilidad para detectar desviaciones de la distribución teórica, proporcionando así criterios prácticos para su aplicación en problemas reales. Además, también recae dentro de los objetivos de este trabajo el proporcionar al usuario un paquete de funciones para el software estadístico R con el fin de permitir al usuario poner en práctica las modificaciones de estos tests propuestas en la presente memoria.

1.3. Estructura del trabajo

En el Capítulo 2 se exponen con carácter general los resultados más relevantes de Teoría de Probabilidades, Procesos Empíricos, Inferencia Estadística y coeficientes de asimetría y curtosis, base teórica sobre la que se cimenta el trabajo. Por su parte, el Capítulo 3 está enfocado específicamente a las familias de localización-escala, haciendo hincapié en su definición, caracterización e ilustrando con numerosos ejemplos de entre los más conocidos y utilizados en la literatura. A lo largo del Capítulo 4, el más relevante desde

el punto de vista teórico, se presentan algunos tests de bondad de ajuste basados en coeficientes de asimetría y curtosis, o en el uso de la función de distribución empírica. Más aún, para estos últimos se elabora un profundo desarrollo compuesto para hipótesis compuestas y se prueba la propiedad de libertad de parámetros dentro de cada familia de localización-escala. A continuación, en la sección inicial del Capítulo 5 se explica la metodología utilizada para la experimentación, para inmediatamente después presentar los resultados de las simulaciones realizadas mediante el método de Monte Carlo para cada objeto de estudio explicitado en los objetivos del presente trabajo. Finalmente, en el Capítulo 6 se resumen las líneas generales del trabajo y se recogen las conclusiones obtenidas, así como se recomiendan futuras líneas de investigación.

CAPÍTULO 2

Preliminares

El propósito de este capítulo es introducir de forma clara y concisa los conceptos fundamentales de probabilidad y estadística que se utilizarán en los subsiguientes capítulos, junto con algunos conceptos complementarios que facilitarán una comprensión más amplia y contextualizada del contenido. Al mismo tiempo, se fijará la notación que se utilizará a lo largo del texto para cada concepto. Para más referencias consultar Casella-Berger [7] y Hewitt E. - Stromberg K. [16], donde se propone una introducción más exhaustiva y detallada.

2.1. Teoría de la probabilidad

La teoría de la probabilidad proporciona el marco matemático fundamental para modelizar fenómenos aleatorios y cuantificar la incertidumbre, proporcionando las herramientas necesarias sobre las que se desarrolla la teoría de la Inferencia Estadística.

2.1.1. Primeras definiciones

A continuación, se introducen las definiciones más básicas utilizadas en el presente trabajo.

Definición 2.1. *Sea Ω un conjunto, una colección de subconjuntos suyos se denomina* ***σ-álgebra****, denotado por $\mathcal{A} \subseteq \mathcal{P}(\Omega)$, si se satisfacen las siguientes 3 propiedades:* (I) $\emptyset \in \mathcal{A}$*;* (II) *Si $H \in \mathcal{A}$ entonces $H^c \in \mathcal{A}$;* (III) *Si $H_1, H_2, \ldots \in \mathcal{A}$ entonces $\cup_{n\in\mathbb{N}} H_n \in \mathcal{A}$.*

Definición 2.2. *Sea Ω un conjunto y $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ una σ-álgebra sobre Ω. Entonces la dupla $(\Omega, \mathcal{A})$ recibe el nombre de* ***espacio medible****.*

Definición 2.3. *Dado un $(\Omega, \mathcal{A})$ espacio medible, decimos que $\mathbb{P} : \mathcal{A} \longrightarrow [0,1]$ es una*

medida de probabilidad *si satisface:* (I) $\mathbb{P}(B) \geq 0 \ \forall B \in \mathcal{A}$; (II) $\mathbb{P}(\Omega) = 1$; (III) *Si* $B_1, B_2, \ldots \in \mathcal{A}$ *son disjuntos dos a dos, entonces* $\mathbb{P}\left(\cup_{n\in\mathbb{N}} B_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}\left(B_n\right)$.

Definición 2.4. *La terna* $(\Omega, \mathcal{A}, \mathbb{P})$ *recibe el nombre de* ***espacio de probabilidad****.*

Definición 2.5. *Dado un espacio de probabilidad* $(\Omega, \mathcal{A}, \mathbb{P})$ *y un espacio medible* (S, Σ)*. Una* ***variable aleatoria*** *es una aplicación* $\mathbf{X} : \Omega \longrightarrow S$ *verificando que* $\forall B \in \Sigma$, $\mathbf{X}^{-1}(B) \in \mathcal{A}$*. Se llama probabilidad inducida por* $\mathbf{X}$ *a* $\mathbb{P}_{\mathbf{X}} = \mathbb{P}\left(\mathbf{X}^{-1}(B)\right)$, $\forall B \in \Sigma$ *.Además, se le dice soporte de* $\mathbf{X}$ *al conjunto de valores que puede tomar la variable aleatoria, denotado* $\mathbb{X}$*. En este trabajo, las variables aleatorias siempre serán denotadas con negrita.*

En adelante, siempre que se consideren variables aleatorias, lo serán sobre un mismo espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$, definido previamente, salvo que se exprese lo contrario.

Sea $\mathbf{Y} : \Omega \longrightarrow \mathbb{R}$ una variable aleatoria que describe los posibles resultados de un experimento, definida sobre un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$. Denotamos por $\mathbb{P}_{\mathbf{Y}}$ la medida de probabilidad inducida por $\mathbf{Y}$.

El *espacio muestral* asociado a $\mathbf{Y}$, denotado por $\mathscr{Y}$, representa el conjunto de todas sus posibles realizaciones. Sobre $\mathscr{Y}$ se considera una σ-álgebra $\mathscr{B}(\mathscr{Y}) \subseteq \mathcal{P}(\mathscr{Y})$, cuyos elementos corresponden a los subconjuntos medibles de $\mathscr{Y}$ (también denominados *eventos* o *sucesos*).

Asumimos que la distribución $\mathbb{P}_{\mathbf{Y}}$, que describe la estructura probabilística de los eventos en $\mathscr{Y}$, es desconocida. Sin embargo, suponemos disponible una familia $\mathscr{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ de medidas de probabilidad sobre $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$ tal que $\mathbb{P}_{\mathbf{Y}} \in \mathscr{P}$. Aquí, Θ es el *espacio paramétrico*, es decir, el conjunto de todos los posibles valores del parámetro ϑ que indexan las distribuciones candidatas en $\mathscr{P}$. Esta construcción constituye el punto de partida de un *experimento estadístico*.

Definición 2.6. *Un* ***modelo o experimento estadístico*** *es una terna* $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}), \mathscr{P})$ *con* $\mathscr{Y} \neq \emptyset$*, una* σ*-álgebra* $\mathscr{B}(\mathscr{Y}) \subseteq \mathcal{P}(\mathscr{Y})$ *sobre* $\mathscr{Y}$*, y una familia de medidas de probabilidad* $\mathscr{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ *sobre el espacio medible* $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$*. Si* $\Theta \subseteq \mathbb{R}^p$*, con* $p \in \mathbb{N}$*, entonces llamamos a* $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}), \mathscr{P})$ *un* ***modelo estadístico paramétrico****, donde* $\vartheta \in \Theta$ *se denomina* ***parámetro****, y* Θ *es el* ***espacio paramétrico****.*

En este trabajo, supondremos que el espacio medible de llegada es siempre $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, donde $\mathcal{B}_{\mathbb{R}}$ es la σ-álgebra de Borel en $\mathbb{R}$ (generada por los abiertos de $\mathbb{R}$), trabajaremos siempre sobre él.

2.2. Teoría de la Inferencia Estadística

La Inferencia Estadística tiene como objetivo fundamental examinar e interpretar datos muestrales con el fin de obtener conclusiones generalizadas sobre las características probabilísticas de la población. también cuantificar la precisión y fiabilidad de las estimaciones obtenidas y conclusiones alcanzadas, respectivamente.

Introduzcamos ahora el concepto de función de distribución de una variable aleatoria, que caracteriza totalmente su distribución de probabilidad.

Definición 2.7. *Sea* $\mathbf{X}$ *una variable aleatoria, entonces su* ***función de distribución*** $F_{\mathbf{X}} : \mathbb{R} \longrightarrow [0,1]$,*viene dada por:*

$$F_{\mathbf{X}} = \mathbb{P}\left(\mathbf{X} \leq x\right), \ x \in \mathbb{R}. \tag{2.1}$$

Además, la función de distribución verifica necesariamente las propiedades:
*(*I*)* $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$; *(*II*)* $F(x)$ *es una función monótona no-decreciente; (*III*)* $F(x)$ *es continua por la derecha,* $\lim_{x\to x_0^+} F(x) = F(x_0)$.

Definición 2.8. *Sea* F *una función de distribución arbitraria sobre* $\mathbb{R}$. *Entonces, llamamos a la función*

$$F^{-1} : (0,1) \longrightarrow \mathbb{R},$$
$$t \longmapsto F^{-1}(t) := \text{ínf}\{x \in \mathbb{R} : F(x) \geq t\},$$

la ***función cuantil*** *o* ***función inversa generalizada*** *(es continua por la izquierda) correspondiente a* F.

En este trabajo solo se consideran variables aleatorias continuas, es decir, aquellas para las que su función de distribución es absolutamente continua. Bajo este contexto se puede definir la función de densidad de una variable aleatoria.

Definición 2.9. *Sea* $\mathbf{X}$ *una variable aleatoria, entonces su* ***función de densidad*** $f_{\mathbf{X}} : \mathbb{R} \longrightarrow \mathbb{R}$ *viene dada por:*

$$f_{\mathbf{X}}(x) = \frac{d}{dx} F_{\mathbf{X}}(x), \ x \in \mathbb{R}\,.$$

Relacionando esta última expresión junto con la Ecuación (2.1), es posible observar que la función de densidad caracteriza totalmente la distribución de probabilidad de la variable:

$$\mathbb{P}\left(\mathbf{X} \leq x\right) = \int_{-\infty}^{x} f_{\mathbf{X}}(x)\, dx\,.$$

Ahondado ahora sobre transformaciones de variables, si $\mathbf{X}$ es una variable aleatoria, entonces para cualquier función Borel-medible $g : \mathbb{R} \longrightarrow \mathbb{R}$, (es decir, $\forall B \in \mathcal{B}_{\mathbb{R}},\ g^{-1}(B) \in \mathcal{B}_{\mathbb{R}}$), $\mathbf{Y} = g(\mathbf{X})$ es también una variable aleatoria con rango o recorrido $\mathbb{Y} = \{y : y = g(x),\ x \in \mathbb{X}\}$. Se sigue el siguiente esquema:

$$(\Omega, \mathcal{A}, \mathbb{P}) \xrightarrow{\mathbf{X}} (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mathbb{P}_{\mathbf{X}}) \xrightarrow{g} (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mathbb{P}_{g(\mathbf{X})})\,. \qquad \overset{Y=g(\mathbf{X})=g\circ\mathbf{X}}{\longrightarrow}$$

En el contexto de la estadística, es habitual trabajar con funciones estrictamente monótonas (inyectivas y derivada con signo constante). De hecho, en este trabajo se utilizarán habitualmente transformaciones lineales, en concreto, transformaciones localización-escala. Es notoria la siguiente relación:

$$\mathbb{P}\left(g\left(\mathbf{X}\right) \leq y\right) = \begin{cases} \mathbb{P}\left(\mathbf{X} \leq g^{-1}(y)\right) = F_{\mathbf{X}}(g^{-1}(y)) & \text{si g es creciente}\,. \\ \mathbb{P}\left(\mathbf{X} \geq g^{-1}(y)\right) = 1 - F_{\mathbf{X}}(g^{-1}(y)) & \text{si g es decreciente}\,. \end{cases}$$

Siguiendo el desarollo, es de vital importancia presentar el siguiente resultado, que expone el comportamiento de la función de densidad bajo transformaciones lineales.

Teorema 2.1. *[7, Cap. 3] Sea $\mathbf{X}$, $g(\mathbf{X}) = \mathbf{Y}$ variables aleatorias y $f_{\mathbf{X}}$, $f_{\mathbf{Y}}$ sus respectivas funciones de densidad. Supongamos que existe una partición $A_0, A_1, \ldots, A_k$ de $\mathbb{X}$ de forma que $\mathbb{P}(\mathbf{X} \in A_0) = 0$ y $f_{\mathbf{X}}$ es continua para cada A_i. Además, supongamos también que existen funciones $g_1, \ldots, g_k$ definidas sobre $A_1, \ldots, A_k$ respectivamente, satisfaciendo: (I) $g(x) = g_i(x)$, para $x \in A_i$; (II) g_i es monótona sobre cada A_i; (III) El conjunto $\mathbb{Y} = \{y :\ y = g_i(x)$ para algún $x \in A_i\}$ es el mismo para cada $i = 1, \ldots, k$; (IV) $g_i^{-1}(y)$ tiene derivada continua sobre $\mathbb{Y}$, para cada $i = 1, \ldots, k$;*

Entonces:

$$f_{\mathbf{Y}}(y) = \begin{cases} \displaystyle\sum_{i=1}^{k} f_{\mathbf{X}}\left(g_i^{-1}(y)\right) \left|\frac{d}{dy} g_i^{-1}(y)\right| & \text{si} \quad y \in \mathbb{Y}\,. \\ 0 & \text{en otro caso}\,. \end{cases} \tag{2.2}$$

Teorema 2.2. *Para cualquier función de distribución F en $\mathbb{R}$, se cumple que*

$$\forall 0 \leq t \leq 1 : (F \circ F^{-1})(t) \geq t. \tag{2.3}$$

La igualdad en la Ecuación (2.3) se cumple si y sólo si t pertenece al rango de F en $\bar{\mathbb{R}}$, es decir, cuando existe un argumento $x \in \bar{\mathbb{R}}$ tal que $F(x) = t$. Si F es continua, entonces tenemos que $F \circ F^{-1} = \mathrm{id}_{[0,1]}$.

A continuación se presenta un resultado muy extendido, conocido en la literatura como transformación cuantil, que relaciona la función de distribución de una variable aleatoria con la función de distribución inversa y viceversa.

Proposición 2.3. *Sea F una función de distribución arbitraria sobre $\mathbb{R}$. Entonces,* (I) *si $\mathbf{U} \equiv U(0,1)$ (distribución uniforme), entonces $\mathbf{X} = F^{-1}(\mathbf{U})$ es una variable aleatoria con función de distribución F;* (II) *si $\mathbf{X}$ es una variable aleatoria continua con distribución F, entonces $F(\mathbf{X})$ es uniformemente distribuida sobre $[0,1]$.*

Es conveniente ahora para el desarrollo del presente capítulo definir la relación de igualdad en distribución.

Definición 2.10. *Dos variables aleatorias $\mathbf{X}$ e $\mathbf{Y}$ se dicen iguales en distribución, y se denotarán $\mathbf{X} \overset{\mathcal{D}}{=} Y$, si se verifica que: $F_{\mathbf{X}}(x) = F_{\mathbf{Y}}(x)\,,\ \forall x \in \mathbb{R}\,.$*

Observación 1. *Ser igualmente distribuidas no implica necesariamente ser iguales.*

Observación 2. *En $\overset{\mathcal{D}}{=}$, $\mathcal{D}$ se utiliza de forma genérica, no significa que la variable aleatoria $\mathbf{X}$ tenga distribución $\mathcal{D}$. Para especificar que una variable aleatoria sigue cierta distribución, se utiliza la notación: $\mathbf{X} \equiv \mathcal{Q}$, siendo $\mathcal{Q}$ una distribución cualquiera.*

De la mano del Teorema 2.1, el siguiente resultado ejemplifica la situación para transformaciones de localización-escala.

Teorema 2.4. *Sea f una función de densidad y $a \in (-\infty, \infty)$, $b > 0$ parámetros cualesquiera pero fijos. Entonces la función $g(x) = \frac{1}{b} f\left(\frac{x-a}{b}\right)$ es también función de densidad.*

Demostración. Debemos comprobar que la función g es no negativa y su integral en $\mathbb{R}$ vale 1. Por ser $f(x)$ función de densidad, es no negativa, por tanto $\forall x \in \mathbb{R},\ f(x) \geq 0$, en concreto, $\frac{1}{b} f\left(\frac{x-a}{b}\right) \geq 0$ pues $b > 0$.

$$\int_{-\infty}^{\infty} g(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{b} f\left(\frac{x-a}{b}\right)\,dx\ = \int_{-\infty}^{\infty} f(y)\,dy = 1,$$

realizando el cambio de variable $y = \frac{x-a}{b}$ y teniendo en cuenta que f es función de densidad. Por tanto, tras la transformación se consigue una función de densidad. □

Para inferir propiedades de una distribución poblacional $\mathcal{D}$, es fundamental establecer un mecanismo de generación de datos. A continuación, se formaliza el concepto de muestreo aleatorio simple.

Definición 2.11. *Dada una variable aleatoria* $\mathbf{X} \equiv \mathcal{D}$*, se denomina* ***muestreo aleatorio simple*** *a partir de* $\mathbf{X}$ *al procedimiento experimental de observar de manera independiente un conjunto de variables* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *distribuidas igual que* $\mathbf{X}$ *(las variables* $\mathbf{X}_i$ *son independientes unas de otras). A* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *se le denomina* ***muestra aleatoria simple*** *de variables aleatorias independientes e idénticamente distribuidas (i.i.d.). Además, se bautiza como* ***realización muestral*** *a un conjunto de observaciones* $(x_1, \ldots, x_n)$ *proveniente de* $\mathbf{X}_1, \ldots, \mathbf{X}_n$*.*

Seguidamente se aporta la definición de estadístico en el marco de la inferencia estadística.

Definición 2.12. *Un estadístico es cualquier aplicación medible:*

$$T : (\mathbf{X}_1, \ldots, \mathbf{X}_n) \longrightarrow \mathbb{R}^p, \ p \leq n, \ p \in \mathbb{N}\,.$$

Observación 3. *Notar que un estadístico es una variable aleatoria, y permite trasladar la distribución de probabilidades muestral a un espacio de menor dimensión.*

Con visión en la definición de la función de distribución empírica, es menester desarrollar el concepto de estadístico ordenado.

Definición 2.13. *Dada una muestra aleatoria simple* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *de tamaño n procedente de una variable aleatoria* $\mathbf{X}$*, es posible ordenar sus componentes de forma no decreciente* $\mathbf{X}_{(1)} \leq \ldots \leq \mathbf{X}_{(n)}$*, se denomina* ***muestra ordenada*** *a la aplicación* $X_{(.)}(x_1, \ldots, x_n) = (x_{(1)}, \ldots, x_{(n)})$*. Se define el* ***estadístico de orden k*** *a la aplicación*

$$\mathbf{X}_{(k)} : \mathbb{R}^n \longrightarrow \mathbb{R}^n\,.$$

$$(x_1, \ldots, x_n) \longmapsto x_{(k)}\,.$$

que devuelve la k-ésima componente de la muestra ordenada.

A continuación, se introduce la función de distribución empírica, que será protagonista en los desarrollos del Capítulo 4.

Definición 2.14. *Sea una variable aleatoria* $\mathbf{X}$ *con función de distribución* $F_{\mathbf{X}}$ *y consideramos* $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ *una muestra aleatoria simple de tamaño n. Se define la* función de distribución empírica *de X asociada a* $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ *como la función* $\hat{F}_n \colon \mathbb{R} \to [0, 1]$ *que*

a cada $x \in \mathbb{R}$ *le asocia el valor:*

$$\hat{F}_n(x) = \frac{\#\{\mathbf{X}_i \leq x\}}{n} = \begin{cases} 0 & si\ x < \mathbf{X}_{(1)}, \\ \frac{i}{n} & si\ \mathbf{X}_{(i)} \leq x < \mathbf{X}_{(i+1)}\ para\ algún\ i \in \{1, \ldots, n-1\}, \\ 1 & si\ x \geq \mathbf{X}_{(n)}. \end{cases}$$

donde se ha utilizado la notación #*, que denota el cardinal de un conjunto.*

Observación 4. *La función de distribución empírica* $\hat{F}_n(\cdot)$ *es un estimador funcional de la función de distribución de una variable aleatoria,* F_X*. Además, la ley fuerte de los grandes números asegura que para cada* $x \in \mathbb{R}$ *se cumple* $\hat{F}_n(x) \overset{c.s.}{\to} F_X(x)$*.*

Antes de proseguir, se fija notación y se desarrolla sobre la esperanza y varianza asociadas a una variable aleatoria.

Definición 2.15 (Esperanza de una variable aleatoria)**.** *Sea* $\mathbf{X}$ *una variable aleatoria, donde* $\mathbb{X} \subseteq \mathbb{R}$*. La esperanza de* $\mathbf{X}$*, denotada por* $\mathbb{E}(\mathbf{X})$*, se define como:*

$$\mathbb{E}(\mathbf{X}) = \begin{cases} \int_{\mathbb{X}} x \cdot f_X(x)\, dx & si\ \mathbf{X}\ es\ absolutamente\ continua\ con\ densidad\ f_X, \\ \sum_{x_i \in \mathbb{X}} x_i \cdot \mathbb{P}(X = x_i) & si\ \mathbf{X}\ es\ discreta\ y\ \mathbb{X}\ es\ numerable, \end{cases}$$

suponiendo que la integral o la serie convergen absolutamente. En caso contrario, decimos que $\mathbb{E}(\mathbf{X})$ *no está definida.*

Definición 2.16 (Varianza y desviación estándar)**.** *Dada una variable aleatoria* $\mathbf{X}$ *con* $\mathbb{E}(\mathbf{X}^2) < \infty$*, la varianza de* $\mathbf{X}$ *se define como la esperanza del cuadrado de su desviación respecto a la media:*

$$\mathbb{V}(\mathbf{X}) = \mathbb{E}\left[(X - \mathbb{E}(\mathbf{X}))^2\right] = \mathbb{E}(\mathbf{X}^2) - [\mathbb{E}(\mathbf{X})]^2,$$

La desviación estándar de $\mathbf{X}$ *es* $sd(\mathbf{X}) = \sqrt{\mathbb{V}(\mathbf{X})}$*.*

Una vez familiarizados con estos conceptos, es conveniente explicar una de las funciones más usadas en el ámbito de la estadística, intimamente ligada a los momentos de una variable aleatoria, es la *función generadora de momentos*. Esta herramienta no solo unifica el estudio de las propiedades distribucionales, sino que también simplifica el cálculo de momentos. A continuación, se introduce su definición formal.

Definición 2.17. *Se define el momento de orden* k*, con* $k \in \mathbb{N}$*, de una variable aleatoria* $\mathbf{X}$ *como* $m_k = \mathbb{E}\left[\mathbf{X}^k\right]$*.*

Definición 2.18. *Sea* $\mathbf{X}$ *una variable aleatoria. La* ***función generadora de momentos*** *de* $\mathbf{X}$*, es* $M_{\mathbf{X}}(t) = \mathbb{E}\left(e^{t\mathbf{X}}\right)$ *si la esperanza existe para cualquier instante* t *en un entorno del cero. En caso contrario, decimos que la función generadora de momentos de la variable no existe. Si* $\mathbf{X}$ *es continua, entonces:*

$$M_{\mathbf{X}}(t) = \mathbb{E}\left(e^{t\mathbf{X}}\right) = \int_{-\infty}^{\infty} e^{t\mathbf{X}} f_{\mathbf{X}}(x)\, dx\,.$$

Se deriva el siguiente teorema que relaciona los momentos de una variable aleatoria con las derivadas de la función generadora de momentos evaluada en $t = 0$.

Teorema 2.5. *Si la variable aleatoria* $\mathbf{X}$ *tiene función generadora de momentos* $M_{\mathbf{X}}$*, entonces:*

$$\mathbb{E}\left(\mathbf{X}^n\right) = M_{\mathbf{X}}^n(0), \text{ donde } M_{\mathbf{X}}^n(0) = \frac{d^n}{dt^n} M_{\mathbf{X}}(t)\bigg|_{t=0}\,.$$

Anteriormente se definieron los momentos de una variable aleatoria, de igual manera que ahora se definen los momentos muestrales asociados a una muestra aleatoria simple.

Definición 2.19. *Dada una muestra aleatoria simple* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *de una variable aleatoria* $\mathbf{X}$*, se define el* ***momento muestral de orden r*** *como* $\overline{\mathbf{X}^r} = \sum_{i=1}^n \frac{\mathbf{X}_i^r}{n}$*. De forma análoga, se define el* ***momento muestral centrado de orden r*** *como* $\overline{\left(\mathbf{X} - \overline{\mathbf{X}}\right)^r} = \sum_{i=1}^n \frac{\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)^r}{n}$*.*

Desviando el cauce, es conveniente ahora tratar el término de variable aleatoria estandarizada, que irá ganando protagonismo en el avance de los capítulos.

Definición 2.20 (Variable aleatoria estandarizada)**.** *Sea* $\mathbf{X}$ *una variable aleatoria tal que existen* $\mathbb{E}(\mathbf{X})$ *y* $\mathbb{V}(\mathbf{X})$*. Entonces se denomina variable aleatoria estandarizada a la transformación resultante de restar a la variable original la esperanza y dividir por la desviación típica.*

$$\mathbf{Z} = \frac{\mathbf{X} - \mathbb{E}(\mathbf{X})}{\sqrt{\mathbb{V}(\mathbf{X})}}\,. \tag{2.4}$$

Siguiendo esta definición, observamos que, del mismo modo, los valores en una variable estandarizada se obtienen a partir de la variable original restando la esperanza y dividiendo por la desviación típica. Los valores estandarizados son adimensionales (sin unidades físicas), y miden la distancia en desviaciones típicas con respecto a la esperanza de la variable aleatoria original.

Introduzcamos algunas medidas estadísticas robustas, esto es, medidas que no son excesivamente afectadas o influenciadas por outliers—puntos atípicos o valores extremos—

.

Definición 2.21 (Cuantil de orden α)**.** *Un* ***cuantil de orden*** α *de una variable aleatoria* $\mathbf{X}$ *es una medida de posición que denotamos por* $C_\alpha(\mathbf{X})$ *para* $\alpha \in (0,1)$*. Se define como un valor tal que verifica:*

$$\lim_{x \to C_\alpha(\mathbf{X})^-} F_{\mathbf{X}}(x) \leq \alpha \leq F_{\mathbf{X}}(C_\alpha(\mathbf{X}))\,.$$

Además, el caso particular del cuantil de orden $\alpha = 0.5$ *se denomina como la mediana y se denota por* $Me(\mathbf{X})$*.*

Cuando la medida de posición central utilizada ha sido la mediana, una medida de variabilidad adecuada es el rango intercuartílico.

Definición 2.22 (Rango intercuartílico)**.** *El* ***rango intercuartílico (IQR)*** *de una variable aleatoria continua* $\mathbf{X}$ *es una medida de dispersión. Viene dado por la siguiente expresión:*

$$IQR(\mathbf{X}) = Q^3_{\mathbf{X}} - Q^1_{\mathbf{X}}\,, \tag{2.5}$$

donde $Q^1_{\mathbf{X}} = C_{0.25}(\mathbf{X})$ *es el primer cuartil, y* $Q^3_{\mathbf{X}} = C_{0.75}(\mathbf{X})$ *es el tercer cuartil.*

La siguiente propiedad será de gran utilidad en este trabajo, donde nos referiremos a ella como linealidad del cuantil para parámetro de escala positivo.

Proposición 2.6. *Sea* $\mathbf{X}$ *una variable aleatoria definida como* $\mathbf{X} = a + b\mathbf{Z}$ *con* $a \in \mathbb{R}$ *y* $b \in (0, \infty)$*, siendo* $\mathbf{Z}$ *otra variable aleatoria. Entonces, para cada* $\alpha \in (0,1)$ *se verifica* $C_\alpha(\mathbf{X}) = a + bC_\alpha(\mathbf{Z})$*.*

2.2.1. Métodos de construcción de estimadores

Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria $\mathbf{X} \sim F_\theta$, donde $\theta = (\theta_1, \ldots, \theta_l) \in \Theta \subseteq \mathbb{R}^l$ son parámetros desconocidos de una familia paramétrica de distribuciones. El objetivo de esta subsección es estimar dichos parámetros desconocidos por medio de los métodos de los momentos y de máxima verosimilitud.

Método de los momentos

El problema de estimación se puede resolver igualando los momentos muestrales y poblacionales, de modo que se plantea una ecuación o sistema de ecuaciones cuyas incógnitas son los valores paramétricos desconocidos. A este método se le denomina método de los momentos.

El momento muestral de orden $k \in \mathbb{N}$ se define como el estadístico:

$$\mathbf{X}^{[k]} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i^k.$$

mientras que el momento poblacional correspondiente es $\mathbb{E}(\mathbf{X}^k) = g_k(\theta_1, \ldots, \theta_l)$, donde $g_k \colon \Theta \to \mathbb{R}$ es una función conocida que depende de la familia paramétrica considerada.

Suponiendo que se quieran estimar los parámetros $\theta_1, \ldots, \theta_l \in \Theta$, se plantearía un sistema de al menos l ecuaciones como el siguiente:

$$\begin{cases} \mathbb{E}(\mathbf{X}) = \overline{\mathbf{X}}_n = g_1(\theta_1, \ldots, \theta_l) \\ \mathbb{E}(\mathbf{X}^2) = \overline{\mathbf{X}}_n^{[2]} = g_2(\theta_1, \ldots, \theta_l) \\ \vdots \qquad\qquad\qquad\qquad\qquad \vdots \\ \mathbb{E}(\mathbf{X}^l) = \overline{\mathbf{X}}_n^{[l]} = g_l(\theta_1, \ldots, \theta_l) \end{cases}$$

Es necesario despejar por lo menos $\theta_1, \ldots, \theta_l$ del sistema anterior para obtener las estimaciones de los l parámetros.

En el caso de que los parámetros sean $\mathbb{E}(\mathbf{X}) = \mu$, $\mathbb{V}(\mathbf{X}) = \sigma^2$, aplicar el método de los momentos resulta en los estimadores clásicos de la media y desviación típica muestrales. Notar que, debido a que la varianza muestral es un estimador sesgado de la varianza poblacional, es común utilizar la medida cuasivarianza muestral.

Método de máxima verosimilitud

El método de máxima verosimilitud es una técnica estadística para estimar los parámetros desconocidos de un modelo, eligiendo los valores que hacen más probables los datos observados. Para este método es estrictamente necesario contar con la definición de función de verosimilitud, que se desarrolla a continuación.

Definición 2.23. *Sea* $\mathbf{X}$ *una variable aleatoria cuya distribución depende de cierto parámetro* $\theta \in \Theta$ *y su función de densidad es* $f(x;\theta)$*. Si* $(x_1, \ldots, x_n)$ *es una muestra de observaciones independientes de* $\mathbf{X}$*, se define la función de verosimilitud de* $(x_1, \ldots, x_n)$ *como* $L(x_1, \ldots, x_n; \theta) : \Theta \to [0, \infty)$ *tal que a cada* $\theta \in \Theta$ *le asocia el valor:*

$$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

El método de máxima verosimilitud se basa en asignar como el valor de θ estimado para $(x_1, \ldots, x_n)$ el $\hat{\theta} \in \Theta$ que hace máxima la función de verosimilitud.

Definición 2.24. *Se denomina estimador máximo-verosímil de θ basado en $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ al estimador $\hat{\theta}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ tal que para cada posible realización muestral $(x_1, \ldots, x_n)$ se tiene que:*

$$L(x_1, \ldots, x_n; \hat{\theta}(x_1, \ldots, x_n)) = \max_{\theta \in \Theta} L(x_1, \ldots, x_n; \theta).$$

Observación 5. *Siempre que sea posible, es habitual maximizar el logaritmo de la función de verosimilitud en lugar de la propia función de verosimilitud para obtener estimadores máximo verosímiles, pues al aplicar la transformación por el logaritmo se conservan los valores extremos, es continua y creciente. Además, generalmente resulta menos costoso computacionalmente y más sencillo analíticamente.*

2.3. Coeficientes de asimetría y curtosis

2.3.1. Coeficientes de asimetría: definición y propiedades

La noción de simetría para una variable aleatoria $\mathbf{X}$ es habitual en el contexto de la estadística.

Definición 2.25. *Una variable aleatoria $\mathbf{X}$ se dice simétrica respecto a x_0 si la distribución de $\mathbf{X} - x_0$ coincide con la de $x_0 - \mathbf{X}$, esto es*

$$\mathbb{P}(\mathbf{X} \leq x_0 - x) = \mathbb{P}(\mathbf{X} \geq x_0 + x), \ \forall x \in \mathbb{R}\,. \tag{2.6}$$

Equivalentemente, si F es la función de distribución de la variable aleatoria $\mathbf{X}$, y no existe ningún punto $x \in \mathbb{R}$ tal que $\mathbb{P}(\mathbf{X} = x) > 0$, se caracteriza la simetría respecto a x_0 como $F(x_0 - x) = 1 - F(x_0 + x), \ \forall x \in \mathbb{R}$. Adicionalmente, si $\mathbf{X}$ es continua con función de densidad f, entonces la simetría respecto a x_0 también se caracteriza como $f(x_0 - x) = f(x_0 + x), \ \forall x \in \mathbb{R}$. Por tanto, si se dice que una variable aleatoria $\mathbf{X}$ es simétrica, necesariamente existe $x_0 \in \mathbb{R}$ respecto al cual $\mathbf{X}$ es simétrica. La mediana (si es única) y la media (si existe) coinciden, tomando además el valor donde ocurre la simetría.

A continuación se introducen unas propiedades deseables para medir la asimetría de una variable aleatoria, provocando la formalización de la noción de coeficiente de asimetría. En este contexto, se dirá que un coeficiente de asimetría es cualquier función $\gamma : \mathcal{F} \longrightarrow \mathbb{R}$ que asigna a cualquier variable aleatoria una medida de su asimetría ($\mathcal{F}$ denota un conjunto o familia de variables aleatorias). Entre otras propiedades, se requiere típicamente que los coeficientes de asimetría verifiquen las siguientes propiedades:

(I) $\gamma(\mathbf{X}) = 0$, si $\mathbf{X}$ es simétrica;

(II) $\gamma(b\mathbf{X} + a) = \gamma(\mathbf{X}), a \in \mathbb{R}$ y $b \in (0, \infty)$, invariante a transformaciones localización-escala;

(III) $\gamma(-\mathbf{X}) = -\gamma(\mathbf{X})$.

Observación 6. *La condición (*I*) no requiere que $\gamma(\mathbf{X}) = 0$ si y sólo si $\mathbf{X}$ es simétrica; realmente es factible encontrar distribuciones asimétricas para las que los valores de los coeficientes de asimetría más conocidos son iguales a cero.*

El coeficiente de asimetría más prominente, conocido como coeficiente de asimetría de Fisher o de momentos, se atribuye típicamente a Pearson [25] (y también a Charlier [8] y Edgeworth [13]) y se define como sigue.

Definición 2.26. *El coeficiente de asimetría -o skewness- de Fisher de una variable aleatoria $\mathbf{X}$ se conoce como el tercer momento de la variable estandarizada:*

$$\gamma_M(X) = skew\,(\mathbf{X}) = \mathbb{E}\left[\left(\frac{\mathbf{X} - \mu}{\sigma}\right)^3\right] . \tag{2.7}$$

La distribución de $\mathbf{X}$ se dice asimétrica positiva, asimétrica negativa o no asimétrica dependiendo de si el valor de $skew\,(\mathbf{X})$ es positivo, negativo o cero, respectivamente.

Notar que, a lo largo de los años, en la literatura se han desarrollado numerosos coeficientes de asimetría, entre ellos: coeficiente de asimetría no paramétrico $\gamma_{NP}(X)$ (atribuido a Yule [32]), coeficiente de asimetría de Bowley $\gamma_B(X)$ [4] (elude la necesidad de existencia de momentos finitos de las variables aleatorias), etc.

A continuación se presentan resultados relativos a coeficientes de asimetría, algunos de ellos específicos para el coeficiente de asimetría de Fisher y otros con carácter general (como la invarianza a transformaciones de localización-escala).

Proposición 2.7. *Sea $\mathbf{X}$ una variable aleatoria simétrica respecto a λ, entonces $\gamma_M(X) = skew\,(\mathbf{X}) = 0$.*

Demostración. Por hipótesis $\lambda - \mathbf{X} \overset{\mathcal{D}}{=} \mathbf{X} - \lambda$, junto con linealidad y simetría, tenemos que:

$$\text{skew}(\mathbf{X}) = \frac{\mathbb{E}\left[(\mathbf{X} - \lambda)^3\right]}{\sigma^3} = \frac{\mathbb{E}\left[(\lambda - \mathbf{X})^3\right]}{\sigma^3} = -\frac{\mathbb{E}\left[(\mathbf{X} - \lambda)^3\right]}{\sigma^3} \implies \text{skew}(\mathbf{X}) = 0\,. \quad \square$$

Observación 7. *No es cierto el recíproco. Una distribución no simétrica puede tener coeficiente de asimetría 0.*

En el trabajo que nos ocupa es notable el interés por las familias de localización-escala, la siguiente proposición desmuestra la invarianza de los coeficientes de asimetría bajo transformaciones de localización-escala.

Proposición 2.8. *El coeficiente de asimetría de una variable aleatoria* $\mathbf{Z}$ *es invariante frente a transformaciones de localización-escala* $a + b\mathbf{Z},\ a \in \mathbb{R},\ b \in (0, \infty)$.

Teorema 2.9. *Es posible expresar el coeficiente de asimetría de Fisher de una variable aleatoria* $\mathbf{X}$ *en términos de los tres primeros momentos de* $\mathbf{X}$*, siempre que estos existan:*

$$skew\,(\mathbf{X}) = \frac{\mathbb{E}\left(\mathbf{X}^3\right) - 3\mu\mathbb{E}\left(\mathbf{X}^2\right) + 2\mu^3}{\sigma^3} = \frac{\mathbb{E}\left(\mathbf{X}^3\right) - 3\mu\sigma^2 - \mu^3}{\sigma^3}\,. \tag{2.8}$$

Demostración. Notar que $(\mathbf{X} - \mu)^3 = \mathbf{X}^3 - 3\mathbf{X}^2\mu + 3\mathbf{X}\mu^2 - \mu^3$, y simplemente utilizando la linealidad de la esperanza obtenemos la primera igualdad. La segunda igualdad viene de sustituir la relación $\mathbb{E}\left(\mathbf{X}^2\right) = \sigma^2 + \mu^2$. □

2.3.2. Coeficientes de curtosis: definición y propiedades

El coeficiente de curtosis es una medida estadística que describe la forma de la distribución de probabilidad de una variable aleatoria, especialmente en lo que respecta a la concentración de sus valores en torno a la media y la presencia de colas extremas. Este, junto con los coeficientes de asimetría anteriormente descritos, proporcionan información muy valiosa sobre las distribuciones de estudio.

Definición 2.27. *La curtosis -o kurtosis- de una variable aleatoria* $\mathbf{X}$ *es el cuarto momento de la variable estandarizada:*

$$kurt\,(\mathbf{X}) = \mathbb{E}\left[\left(\frac{\mathbf{X} - \mu}{\sigma}\right)^4\right]\,. \tag{2.9}$$

Por ejemplificar la manera de cálculo de este coeficiente se presenta el siguiente teorema, que reporta una expresión para calcular la curtosis de una variable aleatoria en términos de los cuatro primeros momentos.

Teorema 2.10. *Es posible expresar la curtosis de una variable aleatoria* $\mathbf{X}$ *en términos de los cuatro primeros momentos de* $\mathbf{X}$*, siempre que estos existan.*

$$\begin{aligned} kurt(\mathbf{X}) &= \frac{\mathbb{E}(\mathbf{X}^4) - 4\mu\mathbb{E}(\mathbf{X}^3) + 6\mu^2\mathbb{E}(\mathbf{X}^2) - 3\mu^4}{\sigma^4} \\ &= \frac{\mathbb{E}(\mathbf{X}^4) - 4\mu\mathbb{E}(\mathbf{X}^3) + 6\mu^2\sigma^2 + 3\mu^4}{\sigma^4}\,. \end{aligned} \tag{2.10}$$

Demostración. De igual manera que en el Teorema 2.8 se obtiene una expresión de $(\mathbf{X}-\mu)^4$ utilizando el Teorema del binomio. Luego se aplica la linealidad del operador esperanza para conseguir la primera igualdad, y se sustituye la relación $\mathbb{E}\left(\mathbf{X}^2\right) = \sigma^2 + \mu^2$ para alcanzar la segunda. □

2.4. Teoría sobre procesos empíricos

En esta sección se recogen de manera preparatoria algunos resultados sobre la teoría de procesos empíricos y medidas, basados en la introducción propuesta por T. Dickhaus en [11]. Partimos de una muestra aleatoria simple $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ de tamaño n procedente de una variable aleatoria $\mathbf{Y}$.

Lema 2.11. *[16] Sea F una función real, no decreciente definida sobre $\mathbb{R}$. Entonces, F tiene límites laterales finitos (a izquierda y derecha) en todos los puntos de $\mathbb{R}$, y es continua excepto en un conjunto numerable de puntos de $\mathbb{R}$.*

Teorema 2.12 (Glivenko-Cantelli)**.** *Se cumple que*

$$\left\|\hat{F}_n - F\right\|_\infty = \sup_{y\in\mathbb{R}} \left|\hat{F}_n(y) - F(y)\right| \to 0 \quad \mathbb{P}\text{-casi seguro cuando } n \to \infty.$$

Demostración. Su prueba se deriva del Lema 2.11 siguiendo la argumentación de Einmahl y de Haan desarrollada en [14]. □

Definición 2.28 (Proceso empírico (reducido))**.** *Bajo las asunciones generales del marco teórico de esta sección, llamamos* ***proceso empírico*** *asociado a $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ a la función aleatoria $\sqrt{n}\left(\hat{F}_n - F\right)(\cdot)$. En el caso especial donde la variable aleatoria verifica $Y \sim U(0,1)$, escribiremos $U_1, \ldots, U_n$ en lugar de $Y_1, \ldots, Y_n$ al referirnos a la m.a.s. En este último caso, definimos además para $0 \leq u \leq 1$ las cantidades*

$$\hat{G}_n(u) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i \leq u\}}, \tag{2.11}$$

$$\mathbb{U}_n(u) = \sqrt{n}\left(\hat{G}_n(u) - u\right), \tag{2.12}$$

y llamamos a la función aleatoria $\mathbb{U}_n(\cdot)$ ***proceso empírico reducido****.*

Lema 2.13. *Sea $n \in \mathbb{N}$, consideramos $U \sim U(0,1)$ y $U_1, \ldots, U_n$ una m.a.s. de variables i.i.d., definimos $Y_i = F^{-1}(U_i)$ para una función de distribución F dada sobre $\mathbb{R}$. Entonces, se cumplen las siguientes afirmaciones:*

(I) $\forall y \in \mathbb{R} : \hat{F}_n(y) = \hat{G}_n(F(y))$
(II) $\forall y \in \mathbb{R} : \sqrt{n}\left(\hat{F}_n(y) - F(y)\right) = \mathbb{U}_n(F(y))$.
(III) Si F *es continua, entonces* $\mathbb{U}_n(u) = \sqrt{n}\left(\hat{F}_n(F^{-1}(u)) - u\right)$ *para todo* $u \in [0,1]$.

Demostración. Para demostrar el apartado (I), calculamos directamente:

$$\hat{G}_n(F(y)) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i \leq F(y)\}} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \leq y\}} = \hat{F}_n(y), \tag{2.13}$$

donde el lado izquierdo de 2.13 se sigue de la Proposición 2.3. El apartado (II) es consecuencia inmediata del apartado (I). Para demostrar (III), sustituimos $y = F^{-1}(u)$ en (b) y recordamos del Teorema 2.2 que $F \circ F^{-1} = \mathrm{id}_{[0,1]}$ cuando F es continua. □

Observación 8. *(a) La función* F*, aunque típicamente desconocida en el contexto estadístico, es una función determinista fija. Por tanto, el apartado (II) del Lema 2.13 muestra que el comportamiento estocástico de un proceso empírico (general) ya está determinado por el del proceso empírico reducido.*

(b) Si solo se está interesado en enunciados probabilísticos sobre variables aleatorias i.i.d. $Y_1, \ldots, Y_n$ *con función de distribución* F *de* Y_1*, entonces la Proposición 2.3 implica que podemos asumir* $Y_i = F^{-1}(U_i)$ *para* $1 \leq i \leq n$ *sin pérdida de generalidad.*

A continuación se exponen definiciones y resultados sobre la teoría de procesos estocásticos, que establecen la base para trabajar en los capítulos subsecuentes.

Definición 2.29. *Un* ***proceso estocástico*** *es una colección indexada* $(X_t)_{t\in\mathcal{T}}$ *de variables aleatorias, donde* $\mathcal{T}$ *denota un conjunto de índices (arbitrario).*

Definición 2.30. *Un proceso estocástico real* $(X_t)_{t\in\mathcal{T}}$ *se llama* ***proceso gaussiano*** *si para cualquier* $m \in \mathbb{N}$ *y cualquier* m*-tupla* $(t_1, \ldots, t_m) \subseteq \mathcal{T}$*, el vector aleatorio* $(X_{t_1}, \ldots, X_{t_m})^\top$ *tiene distribución normal conjunta.*

A partir de estas definiciones, es posible introducir seguidamente los conceptos de movimiento y puente browniano.

Definición 2.31. *Un proceso estocástico real* $(B_t)_{t\geq 0}$*, definido en algún espacio de probabilidad* $(\Omega, \mathcal{F}, \mathbb{P})$*, se llama* ***movimiento browniano*** *si cumple las siguientes condiciones:*

(I) *El proceso* $(B_t)_{t\geq 0}$ *es gaussiano con* $\mathbb{E}[B_t] = 0$ *para todo* $t \geq 0$ *y* $Cov(B_s, B_t) = min\{s,t\}$ *para todo* $s, t \geq 0$.

(II) *La trayectoria* $t \mapsto B_t(\omega)$ *es continua para* $\mathbb{P}$*-casi todo* $\omega \in \Omega$.

Definición 2.32. *Si* $(B_t)_{0\leq t\leq 1}$ *es un movimiento browniano, entonces el proceso* $(B_t^0)_{0\leq t\leq 1}$*, definido por* $B_t^0 = B_t - tB_1$ *para* $0 \leq t \leq 1$*, se llama* ***puente browniano****.*

Se concluye la sección exponiendo resultados teóricos sobre los conceptos anteriormente introducidos, relacionándolos con los procesos empíricos, para más adelante derivar conclusiones relativas a estos.

Lema 2.14. *Un puente browniano* $(B_t^0)_{0\leq t\leq 1}$ *es un proceso gaussiano centrado con* $Cov(B_s^0, B_t^0) = min\{s,t\} - st$ *para todo* $0 \leq s,t \leq 1$*. En particular,* $B_0^0 = B_1^0 = 0$ *casi seguramente.*

Teorema 2.15. *Todas las distribuciones marginales finito-dimensionales asociadas al proceso empírico reducido* $(\mathbb{U}_n(u))_{0\leq u\leq 1}$ *convergen débilmente a las distribuciones marginales correspondientes de un puente browniano* $(B_u^0)_{0\leq u\leq 1}$ *cuando* $n \to \infty$*.*

Por último, el extendido corolario de Donsker, que proporciona convergencia en distribución para el proceso empírico reducido.

Corolario 2.16. *[12, Donsker] El proceso empírico reducido* $(\mathbb{U}_n(u))_{0\leq u\leq 1}$ *converge en distribución a* $(B_u^0)_{0\leq u\leq 1}$ *cuando* $n \to \infty$*.*

CAPÍTULO 3

Familias de localización-escala

En el análisis estadístico y en la teoría de probabilidades, las familias de distribuciones juegan un papel fundamental en la modelización de datos y fenómenos aleatorios. Entre las familias más notorias se encuentran las de localización-escala, que modelizan variaciones en la dispersión y posición de una distribución sin alterar su estructura esencial. Antes de proceder con la definición formal de estas familias, es crucial resaltar su relevancia y las aplicaciones en las que son frecuentemente utilizadas. De forma general, estas familias son altamente relevantes en: modelos de regresión, análisis de supervivencia, teoría de la decisión, procesamiento de señales... Más particularmente, se encuentra gran utilidad a estas familias en contextos donde los cambios en la media y la dispersión son los principales factores de variación, como en problemas de estimación, robustez estadística y bondad de ajuste. Precisamente estos últimos son los problemas que nos ocupan en este trabajo.

3.1. Definición

A continuación se introduce el concepto formal de familia de localización-escala para una distribución de probabilidad $\mathcal{Q}$ genérica.

Definición 3.1 (Familia de localización-escala (LoScF))**.** *Sea $\mathcal{Q}$ una distribución de probabilidad. Se denomina* ***familia de localización-escala*** *(LoScF, abreviado del inglés Location-Scale Family) generada por $\mathcal{Q}$ al conjunto de todas las distribuciones de probabilidad de aquellas variables aleatorias $\mathbf{X}$ para las que existe una variable aleatoria $\mathbf{Z}$ con distribución $\mathcal{Q}$ y dos parámetros $a \in \mathbb{R}$ y $b \in (0, \infty)$ tales que $\mathbf{X} \overset{\mathcal{D}}{=} a + b\mathbf{Z}$. Es una familia 2-paramétrica donde al parámetro a se le conoce como el parámetro de localización y al parámetro b se le conoce como parámetro de escala.*

Observación 9. *Bajo el marco de la definición anterior, se bautiza como* ***distribución***

generadora *a la distribución $\mathcal{Q}$ con respecto a la cual se parametriza la familia de localización-escala (la distribución subyacente de la LoScF). Típicamente se hablará de familia de localización-escala sin mencionar explícitamente la distribución generadora.*

Definición 3.2 (Familia de localización (LoF))**.** *En las condiciones de la definición de LoScF, cuando el parámetro $b = 1$, la familia 1-paramétrica resultante se denomina* ***familia de localización*** *(Location Family) asociada a la distribución dada. Es una subfamilia de la familia localización-escala.*

Definición 3.3 (Familia de escala (ScF))**.** *En las condiciones de la definición de LoScF, cuando el parámetro $a = 0$, la familia 1-paramétrica resultante se denomina* ***familia de escala*** *(Scale Family) asociada a la distribución dada. Es una subfamilia de la familia localización-escala.*

Como motivación a las familias de localización-escala, aportamos luz sobre este tipo de transformaciones con ejemplos cotidianos introducidos en [27, Cap. 5]. Las transformaciones de escala se presentan de manera natural cuando se modifican las unidades físicas de medida (por ejemplo, al cambiar de metros a pies). Por otro lado, las transformaciones de localización se producen cuando se altera el punto de referencia cero (como al medir distancias o tiempos). No obstante, es relevante señalar que las transformaciones de localización y escala pueden coincidir simultáneamente en ciertos cambios de unidades físicas, como sucede al convertir de grados Celsius a Fahrenheit.

3.2. Caracterización y propiedades

Es posible considerar que dos distribuciones de probabilidad relacionadas por una transformación localización-escala gobiernan la misma cantidad aleatoria subyacente, pero en unidades físicas diferentes. Esta relación es lo suficientemente importante como para darle nombre.

Definición 3.4 (Pertenencia a la misma LoScF)**.** *Sean $\mathcal{P}$ y $\mathcal{Q}$ distribuciones de probabilidad relacionadas por una transformación localización-escala con funciones de distribución F y G respectivamente. Entonces $\mathcal{P}$ y $\mathcal{Q}$ son del mismo tipo o pertenecen a la misma familia de localización-escala si existen constantes $a \in \mathbb{R}$ y $b \in (0, \infty)$ de modo que:*

$$F(x) = G\left(\frac{x-a}{b}\right), \ x \in \mathbb{R}. \tag{3.1}$$

Teorema 3.1. *Ser del mismo tipo es una relación de equivalencia sobre la colección de distribuciones de probabilidad en* $\mathbb{R}$*. En efecto, si* $\mathcal{P}$*,* $\mathcal{Q}$ *y* $\mathcal{R}$ *son distribuciones de probabilidad en* $\mathbb{R}$*, entonces se verifican:*

1. *Reflexividad:* $\mathcal{P}$ *es del mismo tipo que* $\mathcal{P}$*.*
2. *Simetría: si* $\mathcal{P}$ *es del mismo tipo que* $\mathcal{Q}$*, entonces* $\mathcal{Q}$ *es del mismo tipo que* $\mathcal{P}$*.*
3. *Transitividad: si* $\mathcal{P}$ *es del mismo tipo que* $\mathcal{Q}$ *y* $\mathcal{Q}$ *es del mismo tipo que* $\mathcal{R}$*, entonces* $\mathcal{P}$ *es del mismo tipo que* $\mathcal{R}$*.*

Demostración. Sean F, G y H las funciones de distribución asociadas a $\mathcal{P}$, $\mathcal{Q}$ y $\mathcal{R}$ respectivamente, comprobemos que se verifican las 3 propiedades.

1. Trivial tomando $a = 0$ y $b = 1$.
2. Por ser $\mathcal{P}$ del mismo tipo que $\mathcal{Q}$ existen $a \in \mathbb{R}$ y $b \in (0, \infty)$ de modo que $F(x) = G\left(\frac{x-a}{b}\right),\ x \in \mathbb{R}$. Por tanto, $G(x) = F\left(a + bx\right) = F\left(\frac{x-\left(-\frac{a}{b}\right)}{\frac{1}{b}}\right)$.
3. Por ser $\mathcal{P}$ del mismo tipo que $\mathcal{Q}$ y $\mathcal{Q}$ del mismo tipo que $\mathcal{R}$, existen $a, c \in \mathbb{R}$ y $b, d \in (0, \infty)$ de modo que $F(x) = G\left(\frac{x-a}{b}\right)$ y $G(x) = H\left(\frac{x-c}{d}\right),\ x \in \mathbb{R}$. Entonces $F\left(x\right) = H\left(\frac{x-(a+bc)}{bd}\right),\ x \in \mathbb{R}$. □

Por consiguiente, hay una partición de la colección de distribuciones de probabilidad sobre $\mathbb{R}$, cuyas clases de equivalencia son exclusivas, donde todas las distribuciones en cada clase verifican la relación ser del mismo tipo. Aún más, las LoScF son trivialmente cerradas bajo transformaciones localización-escala. Claramente, los resultados no dependen de los representantes de la clase elegidos.

Observación 10. *En adelante, al hacer referencia a una familia de localización-escala generada por una distribución de probabilidad cualquiera* $\mathcal{Q}$*, se trabajará indistintamente con* $\mathbf{X} \overset{\mathcal{D}}{=} a + b\mathbf{Z}$ *con* $a \in \mathbb{R}$ *y* $b \in (0, \infty)$*, sobreentendiendo que es el conjunto o familia de distribuciones de las variables aleatorias* $\mathbf{X}$ *(con* $\mathbf{Z} \equiv \mathcal{Q}$*), es decir, la LoScF.*

A continuación, desarrollamos distintas relaciones para determinar la distribución de $\mathbf{X}$ a partir de las correspondientes funciones de probabilidad para $\mathbf{Z}$, bajo el contexto de una familia de localización-escala $\mathbf{X} \overset{\mathcal{D}}{=} a + b\mathbf{Z}$ con $a \in \mathbb{R}$ y $b \in (0, \infty)$.

Proposición 3.2 (Función de distribución de la LoScF)**.** *[27, Cap. 5] Sea* $\mathcal{Q}$ *una distribución de probabilidad arbitraria y* $\mathbf{Z}$ *una variable aleatoria definida sobre* $(\Omega, \mathcal{A}, \mathbb{P})$ *con*

distribución $\mathcal{Q}$. Consideremos la familia de distribuciones de localización-escala (LoScF) generada por $\mathcal{Q}$, es decir, $\mathbf{X} \overset{\mathcal{D}}{=} a + b\mathbf{Z}$ con $a \in \mathbb{R}$ y $b \in (0, \infty)$. Si $\mathbf{Z}$ tiene función de distribución $G_{\mathbf{Z}}$, entonces $\mathbf{X}$ tiene función de distribución $F_{\mathbf{X}}$, dada por:

$$F_{\mathbf{X}}(x) = G_{\mathbf{Z}}\left(\frac{x-a}{b}\right), \ x \in \mathbb{R}\,. \tag{3.2}$$

Demostración. Esto se deduce fácilmente de:
$\forall x \in \mathbb{R}$,

$$F_{\mathbf{X}}(x) = \mathbb{P}(\mathbf{X} \leq x) = \mathbb{P}\left(a + b \cdot \mathbf{Z} \leq x\right) = \mathbb{P}\left(\mathbf{Z} \leq \frac{x-a}{b}\right) = G_{\mathbf{Z}}\left(\frac{x-a}{b}\right). \qquad \square$$

Sin embargo, para la distribución de densidad (Probability Density Function) de una variable aleatoria continua, obtenemos resultados ligeramente diferentes respecto de la Ecuación (3.2), de la Proposición inmediatamente anterior. Esto no es sorprendente, pues la función a considerar tiene significados diferentes.

Proposición 3.3 (Función de densidad (PDF) de la LoScF)**.** *[27, Cap. 5] Sea una LoScF generada por la distribución $\mathcal{Q}$ arbitraria, $\mathbf{X} \overset{\mathcal{D}}{=} a + b\mathbf{Z}$ con $a \in \mathbb{R}$ y $b \in (0, \infty)$. Entonces, si $\mathbf{Z}$ es una variable aleatoria continua con función de densidad de probabilidad $g_{\mathbf{Z}}$, entonces $\mathbf{X}$ también es una variable aleatoria continua con función de densidad de probabilidad $f_{\mathbf{X}}$, dada por:*

$$f_{\mathbf{X}}(x) = \frac{1}{b} \cdot g_{\mathbf{Z}}\left(\frac{x-a}{b}\right), \ x \in \mathbb{R}\,. \tag{3.3}$$

Demostración. Notar primero que $0 = \mathbb{P}\left(\mathbf{X} = x\right) = \mathbb{P}\left(\mathbf{Z} = \frac{x-a}{b}\right)$. Además, como $\mathbf{Z}$ es continua, también lo es $\mathbf{X}$. La Ecuación (3.3) es consecuencia directa de la Proposición 3.2, del Teorema 2.1 y de que la función de densidad de probabilidad se consigue a partir de la derivada de la función de distribución, pues $f_{\mathbf{X}} = F'_{\mathbf{X}}$, $g_{\mathbf{Z}} = G'_{\mathbf{Z}}$. $\square$

En cuanto a los grafos de las funciones de densidad, notar lo siguiente:

- Para la familia localización asociada a $g_{\mathbf{Z}}$, el grafo de $f_{\mathbf{X}}$ se obtiene mediante una traslación del de $g_{\mathbf{Z}}$, a unidades a la derecha si $a > 0$ o $-a$ unidades a la izquierda si $a < 0$.

- Para la familia escala asociada a $g_{\mathbf{Z}}$, si $b > 1$ el grafo de $f_{\mathbf{X}}$ se obtiene a partir del de $g_{\mathbf{Z}}$, estirando horizontalmente y comprimiendo verticalmente mediante el factor b. Si $0 < b < 1$ entonces se obtiene comprimiendo horizontalmente y estirando verticalmente mediante el factor b.

A continuación, se presenta la relación entre las funciones cuantil dentro de una familia de localización-escala. En particular, si z es un cuantil de orden p para $\mathbf{Z}$, entonces $x = a + bz$ es un cuantil de orden p para $\mathbf{X}$.

Proposición 3.4 (Función cuantil de la LoScF)**.** *[27, Cap. 5] Sea una LoScF generada por la distribución $\mathcal{Q}$ arbitraria, $\mathbf{X} \overset{\mathcal{D}}{=} a + b\mathbf{Z}$ con $a \in \mathbb{R}$ y $b \in (0, \infty)$. Si $G_{\mathbf{Z}}^{-1}$ y $F_{\mathbf{X}}^{-1}$ son las funciones cuantil de $\mathbf{Z}$ y $\mathbf{X}$ respectivamente, entonces: $F_{\mathbf{X}}^{-1}(p) = a + bG_{\mathbf{Z}}^{-1}(p)$, con $p \in (0, 1)$.*

Demostración. Este último resultado se sigue directamente de la Ecuación (3.2). Como se puede comprobar, en las condiciones del enunciado, $p \in (0, 1)$:

$$\begin{aligned} 1.\ F_{\mathbf{X}}^{-1}(p) &= \text{ínf}\left\{x \in \mathbb{R} \mid F_{\mathbf{X}}(x) \geq p\right\} = \text{ínf}\left\{x \in \mathbb{R} \mid G_{\mathbf{Z}}\left(\frac{x-a}{b}\right) \geq p\right\} \\ &= \text{ínf}\left\{bx + a \in \mathbb{R} \mid G_{\mathbf{Z}}(x) \geq p\right\} = \text{ínf}\, b\left\{x \in \mathbb{R} \mid G_{\mathbf{Z}}(x) \geq p\right\} + a \\ &= a + bG_{\mathbf{Z}}^{-1}(p)\,. \end{aligned}$$

□

En aras de completar la caracterización de las familias localización-escala, se presenta el siguiente teorema, en el que se establece una relación entre las funciones de densidad de miembros de una familia de localización-escala.

Teorema 3.5. *[7, Cap. 3] Sea f una función de densidad, $a \in \mathbb{R}$, $b \in (0, \infty)$ parámetros cualesquiera pero fijos. Entonces, $\mathbf{X}$ es una variable aleatoria con función de densidad $\frac{1}{b}f\left(\frac{x-a}{b}\right)$ si y solo si existe una variable aleatoria $\mathbf{Z}$ con función de densidad f y $X \overset{\mathcal{D}}{=} a + b\mathbf{Z}$.*

Demostración. $\boxed{\Longleftarrow}$ Definimos $g(x) = bx + a$ y sea $\mathbf{Z} = g(\mathbf{X})$. Basta notar que g es una función monótona, $g^{-1}(x) = \frac{x-a}{b}$ y $\left|\frac{d}{dx}g^{-1}(x)\right| = \frac{1}{b}$. Entonces por el teorema de cambio de variable, la función de densidad de $\mathbf{X}$ es:

$$f_{\mathbf{X}}(x) = f_{\mathbf{Z}}\left(g^{-1}(x)\right)\frac{1}{b} = f_{\mathbf{Z}}\left(\frac{x-a}{b}\right)\frac{1}{b}\,.$$

$\boxed{\Longrightarrow}$ Definimos $g(x) = \frac{x-a}{b}$, y la variable aleatoria $\mathbf{Z} = g(\mathbf{X})$. Notar que $g^{-1}(z) = bz + a$, $\left|\frac{d}{dz}g^{-1}(z)\right| = b$. Entonces, por el teorema de cambio de variable, la función de densidad de $\mathbf{Z}$ es:

$$f_{\mathbf{Z}}(z) = f_{\mathbf{X}}\left(g^{-1}(z)\right)\left|\frac{d}{dz}g^{-1}(z)\right| = \frac{1}{b}f\left(\frac{(bx+a)-a}{b}\right)b = f(z)\,.$$

Tambien,

$$b\mathbf{Z} + a = bg(\mathbf{X}) + a = b\left(\frac{\mathbf{X}-a}{b}\right) + a = \mathbf{X}\,.$$

□

Este resultado, junto con el Teorema 3.1 y la Definición 3.1, reafirma una proposición que hasta el momento puede haber permanecido bajo la manta—*implícita*—: *es posible obtener cualquier miembro de una familia de localización-escala a partir de otro miembro cualquiera mediante una transformación localización-escala* (utilizando las propiedades de transitividad y simetría de la relación de equivalencia). Como ya se había comentado tras el Teorema 3.1, la familia es trivialmente cerrada bajo transformaciones localización-escala.

Tras dilucidar esto, concluimos por tanto que es posible hacer todo tipo de cálculos para una distribución generadora cualquiera y posteriormente adaptarlos a transformaciones localización-escala de la variable aleatoria generadora asociada. En este contexto, se presenta el siguiente teorema, que relaciona la media, varianza y desviación típica de una LoScF general.

Teorema 3.6. *Suponiendo, como antes, una familia de localización-escala* $\mathbf{X} \overset{\mathcal{D}}{=} a + b\mathbf{Z}$ *con* $a \in \mathbb{R},\ b \in (0, \infty)$. *Si existen* $\mathbb{E}(\mathbf{Z})$ *y* $\mathbb{V}(\mathbf{Z})$ *se tiene que:*

(I) $\mathbb{E}(\mathbf{X}) = a + b \cdot \mathbb{E}(\mathbf{Z})$, (II) $\mathbb{V}(\mathbf{X}) = b^2 \cdot \mathbb{V}(\mathbf{Z})$, (III) $sd(\mathbf{X}) = b \cdot sd(\mathbf{Z})$.

Demostración. Trivial utilizando propiedades de la esperanza y varianza. □

Observación 11. *En las condiciones del teorema anterior, notar que si existen* $\mathbb{E}(\mathbf{Z})$ *y* $\mathbb{V}(\mathbf{Z})$ *trivialmente existen también* $\mathbb{E}(a + b\mathbf{Z})$ *y* $\mathbb{V}(a + b\mathbf{Z})$ *para todos los parámetros* $a \in \mathbb{R},\ b \in (0, \infty)$. *Notar también que basta la existencia de la esperanza (varianza) de un miembro de la familia para tener asegurada la existencia de la esperanza (varianza) de los demás miembros. Análogamente, es suficiente la no existencia de la esperanza (varianza) de un miembro para derivar la no existencia para cualquier otro miembro de la familia.*

Con la mirada puesta en el horizonte más amplio y en consecuencia de los resultados anteriores, es posible considerar como distribución generadora de la familia a una distribución cualquiera perteneciente a la misma familia. Sin embargo, en aras de facilitar la interpretación, parece conveniente fijar de forma general unos requisitos comunes a cumplir por la distribución generadora, para cada familia. Motivados por esto, introducimos las siguientes definiciones.

Definición 3.5 (Distribución estandarizada en una LoScF)**.** *Sea una LoScF generada por la distribución* $\mathcal{Q}$ *arbitraria,* $\mathbf{X} \overset{\mathcal{D}}{=} a + b\mathbf{Z}$ *con* $a \in \mathbb{R}$ *y* $b \in (0, \infty)$. *Si existen* $\mathbb{E}(\mathbf{X})$ *y* $\mathbb{V}(\mathbf{X})$, *entonces se denomina distribución estandarizada a aquella que sigue la variable aleatoria estandarizada, es decir, la resultante de restar a la variable original la esperanza y dividir*

por la desviación típica. Se denota por $\mathcal{Q}_{0,1}$.

$$\mathbf{Z} = \frac{\mathbf{X} - \mathbb{E}(\mathbf{X})}{sd(\mathbf{X})} \equiv \mathcal{Q}_{0,1}\,. \tag{3.4}$$

En contraposición, también estudiaremos distribuciones para las que no existen la esperanza y/o varianza, como la distribución de Cauchy. Para estos casos utilizaremos la estandarización en el sentido de Cauchy, introducida a continuación.

Definición 3.6 (Distribución estandarizada en el sentido de Cauchy en una LoScF)**.** *Sea una LoScF generada por la distribución* $\mathcal{Q}$ *arbitraria,* $\mathbf{X} \overset{\mathcal{D}}{=} a + b\mathbf{Z}$ *con* $a \in \mathbb{R}$ *y* $b \in (0, \infty)$*. Si no existen* $\mathbb{E}(\mathbf{X})$ *y/o* $\mathbb{V}(\mathbf{X})$*, entonces se denomina distribución estandarizada en el sentido de Cauchy a aquella que sigue la variable aleatoria estandarizada en el sentido de Cauchy, es decir, la resultante de restar a la variable original la mediana y dividir entre la mitad del rango intercuartílico. Se denota por* $\mathcal{Q}_{0,1}^{Cauchy}$.

$$\mathbf{Z} = \frac{\mathbf{X} - Me(\mathbf{X})}{IQR(\mathbf{X})/2} \equiv \mathcal{Q}_{0,1}^{Cauchy}\,. \tag{3.5}$$

Observación 12. *Algunas LoScF ya han sido estudiadas anteriormente, y existe la costumbre de bautizar a una de las distribuciones de la familia como "distribución estándar", es el representante canónico más reconocido por su uso generalizado en la literatura para cada distribución y habitualmente coincide con la distribución estandarizada. No obstante, siempre se explicitará con qué trabajamos en estas situaciones. Notar la diferencia con la definición de distribución estandarizada.*

Ejemplo 3.1. *En la distribución uniforme no coincide la distribución estandarizada con la distribución que sigue la variable aleatoria estándar. En esta familia, la distribución que sigue la variable aleatoria estándar es* $\mathcal{U}(0,1)$*; mientras que la distribución estandarizada es* $\mathcal{U}\left(-\sqrt{3}, \sqrt{3}\right)$.

En adelante, consecuentemente, al construir una familia de localización-escala, se partirá de una distribución generadora estandarizada, siempre que existan la esperanza y la varianza. En caso contrario, se tomará como base una distribución generadora estandarizada en el sentido de Cauchy. En cada situación específica, si es requerido, se indicará explícitamente el tipo de estandarización empleada.

Lema 3.7. *Para toda LoScF generada por una distribución estandarizada* $\mathcal{Q}_{0,1}$ *arbitraria,* $\mathbf{X} \overset{\mathcal{D}}{=} \mu + \sigma\mathbf{Z}$ *con* $\mu \in \mathbb{R}$ *y* $\sigma \in (0, \infty)$*, el parámetro de localización* μ *se corresponde con el valor de la esperanza* $\mathbb{E}(\mathbf{X})$*; y el parámetro de escala* σ *se corresponde con el valor de la desviación típica* $sd(\mathbf{X})$.

Demostración. Trivial a partir del Teorema 3.6, la propia definición de distribución estandarizada y la construcción de las LoScF. □

Es ahora conveniente introducir una notación para discernir, dentro de las funciones de distribución, entre los miembros de una familia localización-escala. Además, esta notación permitirá descargar las expresiones y facilitar su comprensión.

Observación 13. *Sea una LoScF generada por una distribución estandarizada $\mathcal{Q}_{0,1}$ (o estandarizada en el sentido de Cauchy $\mathcal{Q}_{0,1}^{Cauchy}$) arbitraria, $\mathbf{X} \overset{\mathcal{D}}{=} \mu + \sigma\mathbf{Z}$ con $\mu \in \mathbb{R}$ y $\sigma \in (0, \infty)$. En lo sucesivo denotaremos indistintamente $F_{\mathbf{X}}(x)$ como $F_{\mu,\sigma}(x)$, $f_{\mathbf{X}}(x)$ como $f_{\mu,\sigma}(x)$, y $F_{\mathbf{X}}^{-1}(x)$ como $F_{\mu,\sigma}^{-1}(x)$. En su caso, para la variable aleatoria generadora asociada a la distribución estandarizada $\mathcal{Q}_{0,1}$ (o estandarizada en el sentido de Cauchy $\mathcal{Q}_{0,1}^{Cauchy}$) se utilizarán indistintamente $F_{0,1}(x)$, $f_{0,1}(x)$ y $F_{0,1}^{-1}(x)$. Se generalizará su uso en el Capítulo 4.*

Ejemplo 3.2. *La familia de distribuciones $\{\mathcal{N}(\mu, \sigma^2) : \ \mu \in \mathbb{R}, \ \sigma \in (0, \infty)\}$ es una LoScF generada por la distribución normal estandarizada $\mathcal{N}(0, 1)$.*

Añadir que, como las familias localización-escala (LoScF) esencialmente se corresponden con cambios de unidades, no sorprende que los valores estandarizados se mantengan invariantes bajo transformaciones localización-escala. Se sigue el siguiente teorema con carácter general, pues se verifica para cualquier distribución generadora de la LoScF.

Teorema 3.8. *Sea una LoScF generada por la distribución $\mathcal{Q}$ arbitraria, $\mathbf{X} \overset{\mathcal{D}}{=} a + b\mathbf{Z}$ con $a \in \mathbb{R}$ y $b \in (0, \infty)$, de forma que existen $\mathbb{E}(\mathbf{Z})$ y $\mathbb{V}(\mathbf{Z})$. Los valores estandarizados de las variables aleatorias $\mathbf{Z}$ y $\mathbf{X}$ coinciden siempre, es decir, son iguales.*

Demostración.

$$\frac{\mathbf{X} - \mathbb{E}(\mathbf{X})}{\mathrm{sd}(\mathbf{X})} = \frac{a + b \cdot \mathbf{Z} - (a + b \cdot \mathbb{E}(\mathbf{Z}))}{b \cdot \mathrm{sd}(\mathbf{Z})} = \frac{\mathbf{Z} - \mathbb{E}(\mathbf{Z})}{\mathrm{sd}(\mathbf{Z})}, \tag{3.6}$$

donde se ha utilizado el Teorema 3.6. □

Corolario 3.9. *Los coeficientes de asimetría y curtosis, si existen, son invariantes en familias de localización-escala, pero no caracterizan unívocamente.*

$$\text{(I) } skew(\mathbf{X}) = skew(\mathbf{Z}), \qquad \text{(II) } kurt(\mathbf{X}) = kurt(\mathbf{Z}). \tag{3.7}$$

Demostración. Recordar que los coeficientes de asimetría y curtosis de una variable aleatoria se corresponden, respectivamente, con el tercer y cuarto momento de la variable estandarizada. □

Teorema 3.10. *Sea una LoScF generada por la distribución $\mathcal{Q}$ arbitraria, $\mathbf{X} \overset{\mathcal{D}}{=} a+b\mathbf{Z}$ con $a \in \mathbb{R}$ y $b \in (0,\infty)$, de forma que no existen $\mathbb{E}(\mathbf{Z})$ y/o $\mathbb{V}(\mathbf{Z})$. Los valores estandarizados en el sentido de Cauchy de las variables aleatorias $\mathbf{Z}$ y $\mathbf{X}$ coinciden siempre, es decir, son iguales.*

Demostración.

$$\frac{\mathbf{X} - Me(\mathbf{X})}{\mathrm{IQR}(\mathbf{X})/2} = \frac{a + b\cdot\mathbf{Z} - (a + b\cdot Me(\mathbf{Z}))}{b\cdot\mathrm{IQR}(\mathbf{Z})/2} = \frac{\mathbf{Z} - Me(\mathbf{Z})}{\mathrm{IQR}(\mathbf{Z})/2}, \tag{3.8}$$

donde se ha utilizado las definiciones de mediana, rango intercuartílico y la propiedad de linealidad del cuantil para parámetros de escala positivos. □

Al respecto de los dos teoremas previos, no cabe mayor especulación: hemos conseguido probar que los coeficientes de asimetría y curtosis, cuando estos coeficientes existen, se mantienen constantes para cada familia de localización-escala (recordemos que estos coeficientes son invariantes a transformaciones de localización-escala).

Para terminar, procedemos aportando una relación que se verifica para cualquier distribución generadora de la LoScF.

Proposición 3.11. *Sea una LoScF generada por la distribución $\mathcal{Q}$ arbitraria, $\mathbf{X} \overset{\mathcal{D}}{=} a+b\mathbf{Z}$ con $a \in \mathbb{R}$ y $b \in (0,\infty)$. Es posible representar los momentos de $\mathbf{X}$ centrados, si existen, a partir de los de $\mathbf{Z}$.*

Demostración. Utilizando el Teorema del binomio, tenemos la siguiente cadena de igualdades $\mathbf{X}^n = (a+b\mathbf{Z})^n = \sum_{k=0}^{n}\binom{n}{k}b^k a^{n-k}\mathbf{Z}^k$. Por tanto, basta utilizar que la esperanza es un operador lineal, y concluimos la siguiente relación:

$$\mathbb{E}(\mathbf{X}^n) = \sum_{k=0}^{n}\binom{n}{k}b^k a^{n-k}\mathbb{E}\left(\mathbf{Z}^k\right),\ n \in \mathbb{N}. \tag{3.9}$$

Notar que $\mathbf{Z}^0 = 1$ es una variable degenerada. □

Corolario 3.12. *Los momentos de $\mathbf{X}$ (parámetro de localización a), si existen, tienen una representación simple en términos de los momentos de $\mathbf{Z}$.*

Demostración. Notar que $(\mathbf{X}-a)^n = b^n\mathbf{Z}^n$, por tanto conseguimos la relación:

$$\mathbb{E}\left[(\mathbf{X}-a)^n\right] = b^n\mathbb{E}(\mathbf{Z}^n),\ n \in \mathbb{N}. \tag{3.10}$$

Simplemente se ha aplicado que la esperanza es un operador lineal. □

Es posible hallar una relación entre las funciones generadoras de momentos dentro de una LoScF, como se expone en el siguiente resultado.

Propiedades 3.13. *Sea una LoScF generada por la distribución $\mathcal{Q}$ arbitraria, dada por $\mathbf{X} \overset{\mathcal{D}}{=} a + b\mathbf{Z}$ con $a \in \mathbb{R}$ y $b \in (0, \infty)$. Si $\mathbf{Z}$ tiene como función generadora de momentos a $M_{\mathbf{Z}}$ entonces $\mathbf{X}$ tiene función generadora de momentos $M_{\mathbf{X}}$, dada por:*

$$M_{\mathbf{X}}(t) = \mathbb{E}\left(e^{t\mathbf{X}}\right) = \mathbb{E}\left[e^{t(a+b\mathbf{Z})}\right] = e^{ta}\mathbb{E}\left(e^{tb\mathbf{Z}}\right) = e^{ta}M_{\mathbf{Z}}(bt)\ . \tag{3.11}$$

En vista de lo expuesto anteriormente, se han desarrollado diversas recetas para identificar, en su caso, a qué LoScF puede pertenecer una variable aleatoria. Es importante recordar que los coeficientes de asimetría y curtosis, en virtud del Corolario 3.9, no caracterizan a las LoScF. Asimismo, los desarrollos presentados permiten un análisis más sencillo de cada LoScF. Para ello, es suficiente, por ejemplo, calcular los coeficientes de asimetría y curtosis para la variable aleatoria estándar de la distribución (pues coinciden con su tercer y cuarto momento, respectivamente), utilizar la función generadora de momentos para la variable estándar, o incluso hacerlo para cualquier miembro de la familia arbitrario, ya que los resultados son invariantes. Motivados por estas numerosas posibilidades de estudio, introducimos la siguiente sección de ejemplos, donde se pondrán de manifiesto algunas de las recetas aquí indicadas.

Observación 14. *Respecto a la estimación de parámetros por máxima verosimilitud, es común no poder obtener una expresión cerrada para los parámetros de la familia de localización-escala. Sin embargo, estos pueden estimarse mediante métodos numéricos a partir de la función de verosimilitud o de log-verosimilitud. En este trabajo, cuando se presenta esta situación, se emplea el algoritmo de Nelder-Mead [24, 26] para obtener una aproximación numérica de los parámetros.*

3.3. Ejemplos

Con el fin de ilustrar las familias de localización-escala, en esta sección se introducen numerosos ejemplos de familias de localización-escala, a partir de diversas distribuciones de probabilidad iniciales distintas.

Como se comentó en la sección anterior, dada una LoScF $\mathbf{X} \overset{\mathcal{D}}{=} a + b\mathbf{Z}$ con $a \in \mathbb{R}$ y $b \in (0, \infty)$ es posible hacer todo tipo de cálculos para la variable aleatoria estandarizada $\mathbf{Z}$ y posteriormente adaptarlos a la variable aleatoria $\mathbf{X}$. Esta es, si no, la principal estrategia

que seguiremos en esta sección, siempre que la naturaleza de la distribución de probabilidad lo permita. Sin embargo, al ser los resultados que caracterizan las LoScF invariantes para cualquier representante, seleccionaremos inicialmente aquel que a priori simplifique las cuentas, para luego, gracias a la relación de equivalencia, seleccionar como representante, una vez la familia ya está construida, a la variable estandarizada.

3.3.1. Distribución normal: LoScF

La distribución normal, o gaussiana, juega un papel fundamental en el marco de la probabilidad y estadística, en gran parte debido al reconocido Teorema del Límite Central, que establece una conexión clave entre ambos campos. Además de ser una de las distribuciones más estudiadas y utilizadas, presenta propiedades especialmente convenientes.

Definición 3.7 (Distribución normal estándar)**.** *Una variable aleatoria* $\mathbf{Z}$ *sigue la distribución normal estándar, denotado por* $\mathbf{Z} \equiv \mathcal{N}(0,1)$*, si su función de densidad viene dada por:*

$$\phi(z) = f_{\mathbf{Z}}(z) = f_{0,1}(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}, \ z \in \mathbb{R}\,.$$

Observación 15. *Notar que la variable aleatoria* $\mathbf{Z}$ *con distribución normal estándar verifica* $\mathbb{E}(\mathbf{Z}) = 0, \ \mathbb{V}(\mathbf{Z}) = 1$*.*

Propiedades 3.14. *La función de densidad de* $\mathbf{Z}$ *verifica las siguientes propiedades: (*I*)* ϕ *simétrica respecto a* $z = 0$*; (*II*)* $\phi(z) \longrightarrow 0$ *cuando sucede que* $z \longrightarrow \pm\infty$*; (*III*)* $\frac{d}{dz}\phi(z) = -z\phi(z)$ *.*

Definición 3.8. *Supongamos que* $\mathbf{Z}$ *sigue la distribución normal estándar. Entonces, para* $a \in \mathbb{R}, \ b \in (0,\infty)$*, la variable aleatoria* $\mathbf{X} = a + b\mathbf{Z}$ *sigue una distribución normal con parámetro de localización* a *y parámetro de escala* b*.*

Más aún, la teoría construida en la primera sección del capítulo, concretamente la Proposición 3.3, nos permite derivar una expresión de la función de densidad de $\mathbf{X}$ a partir de la estándar.

$$f_{a,b}(x) = \frac{1}{b} f_{0,1}\left(\frac{x-a}{b}\right) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{(x-a)}{b}\right)^2}, \ x \in \mathbb{R}\,.$$

Observación 16. *Notar que ahora la variable aleatoria* $\mathbf{X}$ *con distribución normal verifica* $\mathbb{E}(\mathbf{X}) = a, \ \mathbb{V}(\mathbf{X}) = b^2$*.*

Poseemos las herramientas necesarias para calcular los coeficientes de asimetría y curtosis correspondientes a la LoScF generada por la distribución normal.

Propiedades 3.15. *En la LoScF generada por la distribución normal se verifica:*

$$\text{(I) } skew(\mathbf{Z}) = 0, \qquad \text{(II) } kurt(\mathbf{Z}) = 3\,. \tag{3.12}$$

El hecho de que skew$(\mathbf{Z}) = 0$ es algo inmediato derivado de la simetría de la distribución normal estándar respecto a su esperanza $x = 0$. En virtud del Corolario 3.9, conseguimos la relación para toda la familia.

Propiedades 3.16 (Estimadores máxima verosimilitud LoScF normal)**.** *Sea una LoScF normal con parámetros* (μ, σ) *y una muestra* $\mathbf{x} = (x_1, \ldots, x_n)$*. Para la LoScF normal, los estimadores de máxima verosimilitud (MLE) se obtienen con las siguientes expresiones.*

$$\hat{\mu}_{MLE} = \sum_{i=1}^{n} \frac{x_i}{n}.$$

$$\hat{\sigma}_{MLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

3.3.2. Distribución uniforme: LoScF

Sigamos con una de las distribuciones de probabilidad más simples y reconocidas, la distribución uniforme continua. En particular, las distribuciones uniformes continuas son las herramientas básicas para simular otras distribuciones de probabilidad. La distribución uniforme continua es un caso fundamental de la familia de localización-escala donde los parámetros determinan los límites del soporte.

Definición 3.9 (Distribución uniforme general)**.** *Una variable aleatoria* $\mathbf{Z}$ *sigue la distribución uniforme, denotado por* $\mathbf{Z} \equiv \mathcal{U}(a, b)$*, si su función de densidad viene dada por:*

$$f_{\mathbf{Z}}(u) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(u) \quad \textit{donde} \quad \mathbb{1}_{[a,b]}(u) = \begin{cases} 1 & \textit{si} \quad u \in [a,b], \\ 0 & \textit{en otro caso.} \end{cases}$$

Observación 17. *Notar que la variable aleatoria* $\mathbf{Z}$ *con distribución uniforme general verifica* $\mathbb{E}(\mathbf{Z}) = \frac{a+b}{2}$, $\mathbb{V}(\mathbf{Z}) = \frac{(b-a)^2}{12}$.

Definición 3.10 (Distribución uniforme estándar)**.** *Una variable aleatoria* $\mathbf{Z}$ *sigue la distribución uniforme estándar si* $\mathbf{Z} \equiv \mathcal{U}(0, 1)$*, es decir, su función de densidad viene dada por* $f_{\mathbf{Z}}(u) = \mathbb{1}_{[0,1]}(u)$.

Observación 18. *Notar que la variable aleatoria* $\mathbf{Z}$ *con distribución uniforme estándar verifica* $\mathbb{E}(\mathbf{Z}) = \frac{1}{2},\ \mathbb{V}(\mathbf{Z}) = \frac{1}{12}$.

Definición 3.11 (Distribución uniforme estandarizada)**.** *Una variable aleatoria* $\mathbf{Z}$ *sigue la distribución uniforme estandarizada si* $\mathbf{Z} \equiv \mathcal{U}(-\sqrt{3}, \sqrt{3})$*, es decir, su función de densidad se deduce sustituyendo en la fórmula de la distribución general.*

Definición 3.12. *Supongamos que* $\mathbf{Z}$ *sigue la distribución uniforme estandarizada. Entonces, para* $k \in \mathbb{R},\ w \in (0, \infty)$*, la variable aleatoria* $\mathbf{X} = k + w\mathbf{Z}$ *sigue una distribución uniforme con parámetro de localización* k *y parámetro de escala* w*.*

De hecho, en virtud de la Proposición 3.3, conseguimos, a partir de la de $\mathbf{Z}$, la expresión de la función de densidad de $\mathbf{X}$ a partir de la estandarizada:

$$\begin{aligned} f_{\mathbf{X}}(x) = f_{k,w}(x) &= \frac{1}{w} f_{0,1}\left(\frac{x-k}{w}\right) = \frac{1}{w}\frac{1}{2\sqrt{3}}\mathbb{1}_{[-\sqrt{3},\sqrt{3}]}\left(\frac{x-k}{w}\right) \\ &= \frac{1}{w}\frac{1}{2\sqrt{3}}\mathbb{1}_{[-\sqrt{3}w+k,\sqrt{3}w+k]}(x). \end{aligned}$$

Observación 19. *Notar que ahora la variable aleatoria* $\mathbf{X}$ *con distribución uniforme verifica* $\mathbb{E}(\mathbf{X}) = k,\ \mathbb{V}(\mathbf{X}) = w^2$.

Fácilmente, es posible determinar los coeficientes de asimetría y curtosis de la LoScF generada por la distribución uniforme.

Propiedades 3.17. *En la LoScF generada por la distribución uniforme se verifica:*

$$\text{(I) } skew\,(\mathbf{Z}) = 0, \qquad \text{(II) } kurt\,(\mathbf{Z}) = \frac{9}{5}. \tag{3.13}$$

No sorprende que $\text{skew}(\mathbf{Z}) = 0$, pues la distribución es simétrica respecto a la esperanza, la probabilidad está distribuida de manera uniforme. Por consiguiente, echando mano ahora del Corolario 3.9, concluimos la caracterización de la LoScF uniforme.

Propiedades 3.18 (Estimadores máxima verosimilitud LoScF uniforme)**.** *Sea una LoScF uniforme con parámetros* (μ, σ)*, consideramos una muestra* $\mathbf{x} = (x_1, \ldots, x_n)$ *en el soporte. Los estimadores de máxima verosimilitud (MLE) de la familia se obtienen a partir de las siguientes expresiones.*

$$\begin{aligned} \hat{\mu}_{MLE} &= \frac{1}{2}(min(x_1, \ldots, x_n) + max(x_1, \ldots, x_n)). \\ \hat{\sigma}_{MLE} &= \sqrt{\frac{1}{12}}(max(x_1, \ldots, x_n) - min(x_1, \ldots, x_n)). \end{aligned}$$

3.3.3. Distribución Laplace: LoScF

La distribución de Laplace, también conocida como distribución doble exponencial, es un ejemplo clásico de familia de localización-escala; se construye a partir de dos variables exponenciales.

Definición 3.13 (Distribución de Laplace general)**.** *Una variable aleatoria* $\mathbf{X}$ *sigue una distribución de Laplace*(μ, b) *si su función de densidad de probabilidad es:*

$$f(x|\mu, b) = \begin{cases} \frac{1}{2b} \exp\left(-\frac{\mu - x}{b}\right) & si\ x < \mu, \\ \frac{1}{2b} \exp\left(-\frac{x - \mu}{b}\right) & si\ x \geq \mu. \end{cases}$$

donde $\mu \in \mathbb{R}$ *es el parámetro de localización y* $b > 0$ *es el parámetro de escala.*

Propiedades 3.19. *Bajo las condiciones de la anterior definición, la variable aleatoria tiene soporte* $\mathbb{X} = (-\infty, \infty)$*,* $E[\mathbf{X}] = \mu$ *y* $Var(\mathbf{X}) = 2b^2$*.*

Por tanto, es posible construir la familia de localización-escala Laplace directamente a partir de una variable aleatoria que siga la distribución Laplace estandarizada, definida a continuación:

Definición 3.14. *Una variable aleatoria* $\mathbf{Z}$ *sigue la* ***distribución de Laplace estandarizada*** *si* $\mathbf{Z} \equiv Laplace(0, \frac{1}{\sqrt{2}})$*. Su función de densidad se deduce a partir de la definición previa.*

Observación 20. *Notar que la variable aleatoria* $\mathbf{Z}$ *con distribución de Laplace estandarizada verifica trivialmente* $\mathbb{E}(\mathbf{Z}) = 0,\ \mathbb{V}(\mathbf{Z}) = 1.$

Definición 3.15. *Supongamos que* $\mathbf{Z}$ *sigue la distribución de Laplace estandarizada. Entonces, para* $a \in \mathbb{R},\ b \in (0, \infty)$*, la variable aleatoria* $\mathbf{X} = a + b\mathbf{Z}$ *sigue una distribución de Laplace y respecto a la familia tiene parámetro de localización* a *y parámetro de escala* b*.*

Más aún, la teoría construida en la primera sección del capítulo, concretamente la Proposición 3.3, nos permite derivar una expresión de la función de densidad de $\mathbf{X}$ a partir de la estandarizada.

$$f_{a,b}(x) = \frac{1}{b} f_{0,1}\left(\frac{x - a}{b}\right) = \begin{cases} \frac{1}{b}\frac{\sqrt{2}}{2} \exp\left(-\frac{(-x+a)}{b}\sqrt{2}\right) & \text{si } x < a \\ \frac{1}{b}\frac{\sqrt{2}}{2} \exp\left(-\frac{(x-a)}{b}\sqrt{2}\right) & \text{si } x \geq a \end{cases}, \ x \in \mathbb{R}\,.$$

Observación 21. *Notar que ahora la variable aleatoria* $\mathbf{X}$ *con distribución de Laplace verifica* $\mathbb{E}(\mathbf{X}) = a,\ \mathbb{V}(\mathbf{X}) = b^2$.

Inmediatamente tras estas consideraciones previas, poseemos las herramientas necesarias para conseguir los coeficientes de asimetría y curtosis de la LoScF generada por la distribución de Laplace.

Propiedades 3.20. *En la LoScF generada por la distribución de Laplace se verifica:*

$$\text{(I) } skew\,(\mathbf{Z}) = 0, \qquad \text{(II) } kurt\,(\mathbf{Z}) = 6\,. \tag{3.14}$$

El hecho de que skew$(\mathbf{Z}) = 0$ es algo inmediato derivado de la simetría de la distribución de Laplace estandarizada respecto a su esperanza $x = 0$. En virtud del Corolario 3.9, conseguimos la relación para toda la LoScF de Laplace.

Respecto a la estimación de parámetros por máxima verosimilitud, a pesar de que existe una expresión cerrada para la distribución de Laplace $\hat{\mu}_{\text{MLE}} = \text{Me}\{x_1, \ldots, x_n\}$, $\hat{\sigma}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{N} |x_i - \hat{\mu}|$, no se han conseguido derivar expresiones para los parámetros de la familia de localización-escala construida a partir de la distribución estandarizada. Sin embargo, estos se estimarán mediante métodos numéricos a partir de la función de verosimilitud o de log-verosimilitud, empleando el algoritmo de Nelder-Mead [24, 26] para obtener una aproximación numérica de los parámetros.

3.3.4. Distribución exponencial: LoScF

La distribución exponencial constituye un pilar de los modelos de duración y procesos estocásticos continuos, definida por su tasa constante de decaimiento λ. Como distribución clave en procesos de Markov, aparece naturalmente como tiempo entre eventos en procesos de Poisson. Su propiedad de falta de memoria la hace particularmente útil en aplicaciones de fiabilidad y análisis de supervivencia.

Definición 3.16 (Distribución exponencial general)**.** *Una variable aleatoria* $\mathbf{X}$ *sigue una distribución exponencial con parámetro* $\lambda > 0$ *si su función de densidad de probabilidad es:*

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & si\ x \geq 0, \\ 0 & si\ x < 0. \end{cases}$$

donde $\lambda > 0$ *es el parámetro de escala.*

Propiedades 3.21. *Para* $\mathbf{X} \sim Exp(\lambda)$ *se cumple que* $\mathbb{X} = [0, \infty)$, $\mathbb{E}[\mathbf{X}] = \frac{1}{\lambda}$, $Var(\mathbf{X}) = \frac{1}{\lambda^2}$.

Notar que esta vez no es posible obtener directamente la variable aleatoria estandarizada (limitados por el soporte), sin embargo, esto no constituye ningún problema ni nos limita a la hora de construir la familia de localización-escala. Para generar la familia de localización-escala se parte de $\mathbf{Y} \sim \text{Exp}(1)$, considerando los parámetros $a \in \mathbb{R}$ y $b > 0$, se tiene que $\mathbf{X} = a + b\mathbf{Y}$ sigue una distribución exponencial desplazada.

Convenientemente, utilizando ahora la relación de equivalencia del Teorema 3.1, la LoScF ya generada no depende del representante elegido, por ello seleccionamos $\mathbf{Z} = Y - 1$ como nuevo representante, pues verifica $\mathbb{E}[\mathbf{Z}] = 0$, $\mathbb{V}(\mathbf{Z}) = 1$, es la variable aleatoria estandarizada de la familia. Al retomar la variable estandarizada como representante, volvemos a tener consistencia con la notación.

Definición 3.17. *Supongamos que* $\mathbf{Z}$ *sigue la distribución exponencial estandarizada. Entonces, para* $a \in \mathbb{R}$, $b \in (0, \infty)$, *la variable aleatoria* $\mathbf{X} = a + b\mathbf{Z}$ *sigue una distribución exponencial desplazada y respecto a la familia tiene parámetro de localización* a *y parámetro de escala* b.

Más aún, la teoría construida en la primera sección del capítulo, concretamente la Proposición 3.3, nos permite derivar una expresión de la función de densidad de $\mathbf{X}$ a partir de la estandarizada.

$$f_{0,1}(x) = \begin{cases} e^{-(x+1)} & \text{si } x \geq -1, \\ 0 & \text{si } x < -1. \end{cases} \quad x \in \mathbb{R}\,.$$

$$f_{a,b}(x) = \frac{1}{b} f_{0,1}\left(\frac{x-a}{b}\right) = \begin{cases} \frac{1}{b} e^{-(\frac{x-a}{b}+1)} & \text{si } x \geq a-b, \\ 0 & \text{si } x < a-b. \end{cases} \quad x \in \mathbb{R}\,.$$

Observación 22. *Notar que ahora la variable aleatoria* $\mathbf{X}$ *con distribución exponencial desplazada verifica* $\mathbb{E}(\mathbf{X}) = a$, $\mathbb{V}(\mathbf{X}) = b^2$.

Ahora, poseemos las herramientas necesarias para conseguir los coeficientes de asimetría y curtosis de la LoScF generada por la distribución exponencial.

Propiedades 3.22. *Para la exponencial desplazada estandarizada* ($\mu = 0$, $\sigma = 1$)*:*

$$(1)\ \mathit{skew}(\mathbf{Z}) = 2, \quad (2)\ \mathit{kurt}(\mathbf{Z}) = 9. \tag{3.15}$$

En virtud del Corolario 3.9, conseguimos la relación para toda la LoScF exponencial.

Observación 23. *La asimetría positiva constante refleja la cola derecha con más peso que la cola izquierda, característica de esta distribución.*

Propiedades 3.23 (Estimadores máxima verosimilitud LoScF exponencial)**.** *Sea una LoScF exponencial con parámetros* (μ, σ)*, consideramos una muestra* $\mathbf{x} = (x_1, \ldots, x_n)$ *en el soporte. Los estimadores de máxima verosimilitud (MLE) de la familia se obtienen a partir de las siguientes expresiones.*

$$\hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i = mean(x_1, \ldots, x_n).$$

$$\hat{\sigma}_{MLE} = \overline{\mathbf{x}} - \text{mín}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i - \text{mín}(\mathbf{x}).$$

3.3.5. Distribución logística: LoScF

La distribución logística es otro ejemplo importante de familias de localización-escala, ampliamente utilizada en modelos de regresión logística y en problemas de crecimiento. Su densidad de probabilidad, función de distribución acumulativa y propiedades se describen a continuación.

Definición 3.18 (Distribución logística general)**.** *Una variable aleatoria* $\mathbf{X}$ *sigue una distribución logística* (μ, s) *si su función de densidad de probabilidad es:*

$$f(x|\mu, s) = \frac{e^{-\frac{x-\mu}{s}}}{s\left(1 + e^{-\frac{x-\mu}{s}}\right)^2}, \quad x \in \mathbb{R}.$$

donde $\mu \in \mathbb{R}$ *es el parámetro de localización y* $s > 0$ *es el parámetro de escala.*

Propiedades 3.24. *Bajo las condiciones de la anterior definición, la variable aleatoria tiene soporte* $\mathbb{X} = (-\infty, \infty)$*,* $\mathbb{E}[\mathbf{X}] = \mu$ *y* $Var(\mathbf{X}) = \frac{s^2\pi^2}{3}$*.*

Por tanto, es posible construir la familia de localización-escala logística directamente a partir de una variable aleatoria que siga la distribución logística estandarizada, definida a continuación.

Definición 3.19. *Una variable aleatoria* $\mathbf{Z}$ *sigue la* ***distribución logística estandarizada*** *si* $\mathbf{Z} \equiv Logística(0, \frac{\sqrt{3}}{\pi})$*. Su función de densidad se deduce a partir de la definición previa. Coincide con la distribución logística estándar.*

Observación 24. *Notar que la variable aleatoria* $\mathbf{Z}$ *con distribución logística estandarizada verifica trivialmente* $\mathbb{E}(\mathbf{Z}) = 0,\ \mathbb{V}(\mathbf{Z}) = 1.$

Definición 3.20. *Supongamos que* $\mathbf{Z}$ *sigue la distribución logística estandarizada. Entonces, para* $a \in \mathbb{R}, b \in (0, \infty)$*, la variable aleatoria* $\mathbf{X} = a + b\mathbf{Z}$ *sigue una distribución logística con parámetro de localización* a *y parámetro de escala* b*.*

Más aún, la teoría construida en la primera sección del capítulo, concretamente la Proposición 3.3, nos permite derivar una expresión de la función de densidad de $\mathbf{X}$ a partir de la estándar:

$$f_{a,b}(x) = \frac{1}{b} f_{0,1}\left(\frac{x-a}{b}\right) = \frac{1}{b} \frac{e^{-\frac{x-a}{b}\frac{\pi}{\sqrt{3}}}}{\frac{\sqrt{3}}{\pi}\left(1 + e^{-\frac{x-a}{b}\frac{\pi}{\sqrt{3}}}\right)^2}, \quad x \in \mathbb{R}.$$

Observación 25. *Notar que ahora la variable aleatoria* $\mathbf{X}$ *con distribución logística verifica* $\mathbb{E}(\mathbf{X}) = a,\ \mathbb{V}(\mathbf{X}) = b^2$*.*

Inmediatamente tras estas consideraciones previas, poseemos las herramientas necesarias para determinar los coeficientes de asimetría y curtosis de la LoScF generada por la distribución logística.

Propiedades 3.25. *En la LoScF generada por la distribución logística se verifica:*

$$(1)\ \mathit{skew}(\mathbf{Z}) = 0, \qquad (2)\ \mathit{kurt}(\mathbf{Z}) = \frac{21}{5}. \tag{3.16}$$

En virtud del Corolario 3.9, conseguimos la relación para toda la LoScF logística.

Respecto a la estimación de parámetros por máxima verosimilitud, no se ha obtenido una expresión cerrada para los parámetros de la familia de localización-escala. Sin embargo, estos serán estimados mediante métodos numéricos a partir de la función de verosimilitud o de log-verosimilitud empleando el algoritmo de Nelder-Mead [24, 26] para obtener una aproximación numérica de los parámetros.

3.3.6. Representaciones gráficas

A modo de ejemplo, se presenta en la Figura 3.1 un conjunto de representaciones gráficas de funciones de distribución pertenecientes a familias de localización-escala (LoScF), específicamente las distribuciones normal, logística, exponencial y de Laplace, caracterizadas por distintos parámetros de localización y escala. Resalta la similitud en el patrón de distribución de probabilidad entre las funciones de distribución normal, logística y de Laplace.

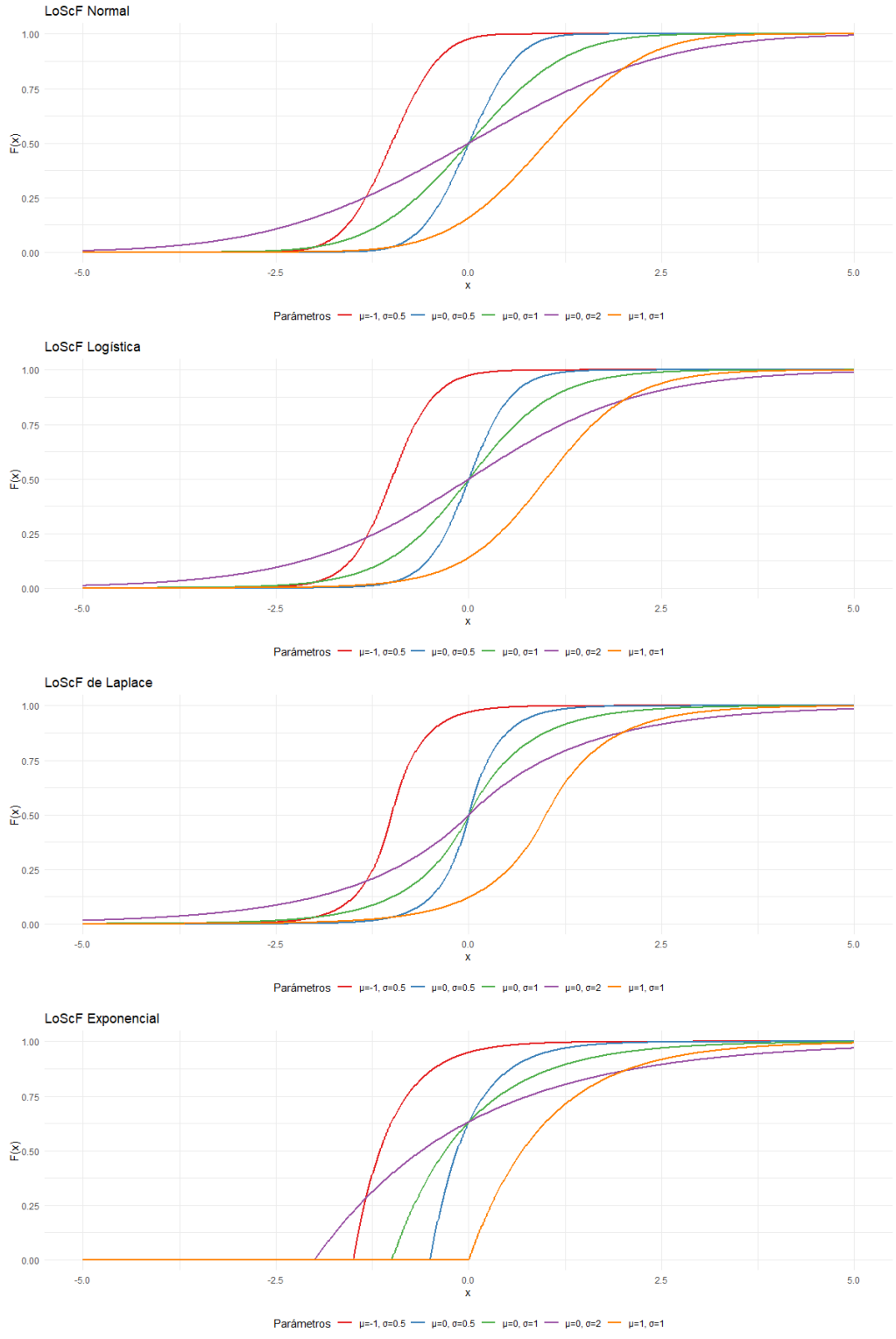

Figura 3.1: Función de distribución de las LoScF normal, logística, Laplace y exponencial con distintos parámetros de localización-escala.

CAPÍTULO 4

Tests de bondad de ajuste a familias de localización-escala

Este capítulo aborda de manera rigurosa el problema de los contrastes de bondad de ajuste (Goodness-of-Fit) aplicados a familias de localización-escala. Se parte con una introducción general a los contrastes de hipótesis; a continuación, se contextualizan las pruebas de significación a específicamente el desafío que nos ocupa: bondad de ajuste. Posteriormente, se desarrolla un marco teórico tanto para los tests clásicos de bondad de ajuste basados en la función de distribución empírica ECDF (Kolmogorov-Smirnov, Anderson-Darling y Cramér-von Mises) como para los tests de bondad de ajuste basados en coeficientes de asimetría y/o curtosis. También se aportan algoritmos para la realización de estos tests.

4.1. Preámbulo sobre contrastes de hipótesis: bondad de ajuste

Recordando el Capítulo 2 y siguiendo la definición 2.6, consideramos un experimento estadístico, la terna $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}), \mathscr{P})$ con $\mathscr{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ asociada a una variable aleatoria $\mathbf{Y}$ con $\mathbb{P}_\mathbf{Y}$ la medida de probabilidad asociada.

Realmente el objetivo de la inferencia estadística es derivar afirmaciones sobre la verdadera, pero desconocida distribución $\mathbb{P}_Y$ o, equivalentemente, sobre el verdadero, pero desconocido e inobservable valor de ϑ, basándose en los datos observados, denotando esta vez una observación por y. Formalmente, numerosos y diferentes tipos de problemas de inferencia estadística pueden especificarse como problemas de decisión estadística; sin embargo, nos centraremos en los problemas de contraste, que constituyen una clase importante de problemas de decisión estadística. El objetivo de estos es determinar, de acuerdo

con la información muestral recogida, si determinada hipótesis acerca de alguna característica poblacional debe ser rechazada o no. Para ello se analizará si existen indicios suficientes para cuestionar su validez.

Definición 4.1 (Contraste Estadístico)**.** *Sean dos subconjuntos no vacíos y disjuntos $\mathscr{P}_0$ y $\mathscr{P}_1$ de $\mathscr{P}$, tales que $\mathscr{P}_0 \cup \mathscr{P}_1 = \mathscr{P}$. El objetivo es decidir, basándose en los datos observados, si $\mathbb{P}_Y \in \mathscr{P}_0$ o si $\mathbb{P}_Y \in \mathscr{P}_1$ es verdadero. Si existe una correspondencia uno a uno entre los elementos de $\mathscr{P}$ y el valor de ϑ, es posible preguntar equivalentemente si $\vartheta \in \Theta_0$ o si $\vartheta \in \Theta_1$ es verdadero, dónde los subconjuntos no vacíos y disjuntos Θ_0 y Θ_1 de Θ corresponden a $\mathscr{P}_0$ y $\mathscr{P}_1$ en el sentido de que $\mathbb{P}_Y \in \mathscr{P}_0$ si y sólo si $\vartheta \in \Theta_0$. En este último caso, se define la hipótesis nula H_0 y la hipótesis alternativa H_1 como*

$$H_0 : \vartheta \in \Theta_0 \iff \mathbb{P}_Y \in \mathscr{P}_0 \qquad H_1 : \vartheta \in \Theta_1 \iff \mathbb{P}_Y \in \mathscr{P}_1.$$

A menudo, se interpretan directamente H_0 y H_1 como subconjuntos de Θ, es decir, se consideran conjuntos H_0 y H_1 tales que $H_0 \cup H_1 = \Theta$ y $H_0 \cap H_1 = \emptyset$.

*Un **contraste estadístico (no aleatorizado)** φ es una aplicación medible*

$$\varphi : (\mathscr{Y}, \mathscr{B}(\mathscr{Y})) \to (\{0,1\}, 2^{\{0,1\}})$$

con la convención de que

$$\varphi(y) = 1 \iff \textit{Rechazo de } H_0 \qquad \varphi(y) = 0 \iff \textit{No rechazo de } H_0.$$

Definición 4.2. *En las condiciones anteriores, el subconjunto $\{y \in \mathscr{Y} : \varphi(y) = 1\}$ se denomina región de rechazo o, de manera equivalente, región crítica de φ, abreviado como $\{\varphi = 1\}$. Su complementario $\{y \in \mathscr{Y} : \varphi(y) = 0\}$ se llama región de aceptación de φ, abreviado como $\{\varphi = 0\} = \{\varphi = 1\}^c$.*

La aleatoriedad de los datos implica la posibilidad de cometer un error al realizar el contraste. Se dice que ocurre un *error de tipo I* cuando se toma la decisión a favor de H_1, aunque en realidad H_0 sea verdadera. Análogamente, un *error de tipo II* ocurre si no se rechaza H_0, aunque H_1 sea verdadera. Típicamente, no es posible minimizar ambas probabilidades de error (probabilidad de error de tipo I y probabilidad de error de tipo II) simultáneamente para un modelo estadístico dado y hipótesis fijas H_0 y H_1. El enfoque clásico (Neyman-Pearson) del problema de contraste estadístico trata estos dos errores de manera asimétrica. El error de tipo I se considera más grave, y su probabilidad se acota mediante una constante predefinida $\alpha \in (0,1)$, llamada **nivel de significación**. La elección fijada en este trabajo es $\alpha = 0.05$.

Definición 4.3. *Se llama* ***función potencia del test*** φ *a la probabilidad de rechazar* H_0 *cuando el parámetro es* ϑ*, es decir,* $Pot(\vartheta) = E_\vartheta(\varphi) = P_\vartheta(RC)$*. Si* $\vartheta \in \Theta_0$ *(es decir,* ϑ *pertenece al conjunto de parámetros bajo la hipótesis nula* H_0*), entonces* $E_\vartheta(\varphi)$ *representa la probabilidad de cometer un error de tipo I. En cambio, si* $\vartheta \in \Theta_1$ *(es decir,* ϑ *pertenece al conjunto de parámetros bajo la hipótesis alternativa), entonces* $E_\vartheta(\varphi)$ *representa la probabilidad de rechazar correctamente* H_0*; es decir, el poder del test en ese punto.*

Observación 26. *En la clase de todos los contrastes* φ *para los cuales la probabilidad de error de tipo I no excede* α*, se busca entonces el mejor contraste en términos de minimizar la probabilidad de error de tipo II. Equivalentemente, esto significa que se busca maximizar la potencia del contraste bajo la restricción de mantener el nivel de significación* α*, siempre que nos encontremos bajo* H_0*.*

Tras esta introducción, hacemos frente a los problemas que nos ocupan: bondad de ajuste a LoScF, se busca comprobar la compatibilidad de los datos con un modelo fijo único (simple) o con un modelo dentro de una clase dada (compuesto).

Sea $\mathbf{X}_1, \ldots, \mathbf{X}_n$ una m.a.s. i.i.d. de una distribución desconocida F. El caso de bondad de ajuste simple surge si queremos inferir si esta muestra proviene de cierta distribución hipotética totalmente especificada F_0, pudiendo plantear el problema como la siguiente prueba de hipótesis:

$$H_0 \,:\, F = F_0 \quad \text{vs.} \quad H_1 \,:\, F \neq F_0 \,.$$

El problema de bondad de ajuste compuesto surge al querer probar si la distribución de la muestra pertenece a cierta clase $\mathcal{F}$ de funciones de distribución. En este caso, el problema de bondad de ajuste compuesto se reduce a testear:

$$H_0 :\; F \in \underbrace{\{F_\theta :\; \theta \in \Theta\}}_{\mathcal{F}} \qquad vs. \qquad H_1 :\; F \notin \{F_\theta :\; \theta \in \Theta\}.$$

donde $\Theta \subseteq \mathbb{R}^p$ con $p \in \mathbb{N}$.

Cuando la hipótesis nula es compuesta (se corresponde a distribuciones en una clase paramétrica conocida) típicamente hace que el análisis formal de los procedimientos de prueba sea mucho más desafiante. Sin embargo, en general, este es el tipo de pruebas que se desea realizar en la práctica.

4.2. GoF a partir de la distribución empírica (ECDF)

En el análisis estadístico, la función de distribución empírica juega un papel fundamental como herramienta para estimar la función de distribución acumulativa de una población a partir de una muestra de datos. Su estudio está estrechamente ligado al proceso empírico, una construcción estocástica que permite analizar la convergencia de la ECDF hacia la CDF teórica. Entre los resultados más importantes en este contexto se encuentra el teorema de Glivenko-Cantelli, que establece la convergencia uniforme casi segura de la ECDF a la CDF subyacente, sentando las bases para pruebas de bondad de ajuste.

Las pruebas de bondad de ajuste basadas en la ECDF, como las de Kolmogorov-Smirnov, Cramér-von Mises y Anderson-Darling, aprovechan las propiedades del proceso empírico para evaluar si una muestra dada proviene de una distribución específica. Estas metodologías son particularmente útiles por su capacidad para detectar discrepancias globales entre la distribución empírica y la teórica, sin depender de agrupamientos arbitrarios de los datos. En esta sección estudiamos las pruebas de bondad de ajuste para el caso simple y, más a fondo, el compuesto.

Definición 4.4. *[10, Cap. 14] Sean Q_1 y $Q_2 \in \mathscr{P}$ con funciones de distribución correspondientes F_1 y F_2. Entonces, llamamos:*

(I) *ρ_{KS}, dado por*

$$\rho_{KS}(Q_1, Q_2) = \sup_{y \in \mathbb{R}} |F_1(y) - F_2(y)|,$$

la métrica de Kolmogorov-Smirnov *(o métrica uniforme) en $\mathscr{P} \times \mathscr{P}$.*

(II) *ρ_{CvM}, dado por*

$$\rho_{CvM}(Q_1, Q_2) = \int_{\mathbb{R}} [F_1(y) - F_2(y)]^2 \, dF_2(y),$$

la distancia de Cramér-von Mises *en $\mathscr{P} \times \mathscr{P}$.*

(III) *ρ_{AD}, dado por*

$$\rho_{AD}(Q_1, Q_2) = \int_{\mathbb{R}} \left[\frac{F_1(y) - F_2(y)}{F_2(t)(1 - F_2(t))} \right]^2 dF_2(y),$$

la distancia de Anderson-Darling *en $\mathscr{P} \times \mathscr{P}$.*

Observación 27. *En adelante, bajo las condiciones de la definición anterior, se escribirá $\rho(F_1, F_2)$ en lugar de $\rho(Q_1, Q_2)$.*

En [11, Cap. 3] se presenta un desarrollo teórico exhaustivo sobre la construcción de estos tests, incluyendo la derivación de resultados relativos a su distribución teórica en contextos de bondad de ajuste simple, así como la justificación de las simplificaciones aplicadas a las expresiones de los estadísticos. Si bien este marco teórico resulta fundamental, su tratamiento detallado escapa al alcance del presente trabajo. Para nuestros propósitos, en esa materia es suficiente con considerar la teoría de procesos empíricos desarrollada en el Capítulo 2, que garantiza la comprensión de estos resultados; solo se exponen algunos de los resultados más relevantes a continuación.

Teorema 4.1. *Sean $Y_1, \ldots, Y_n$ variables aleatorias i.i.d. de valor real con cdf continua F de Y_1, $\hat{F}_n$ la cdf empírica correspondiente a $Y_1, \ldots, Y_n$, y sea $\mathbb{U}_n$ un proceso empírico reducido, donde $U_i := F(Y_i)$ para todo $1 \leq i \leq n$. Entonces se cumple:*

$$\sqrt{n} \cdot \rho_{KS}(\hat{F}_n,\, F) = \sup_{0 \leq u \leq 1} |\mathbb{U}_n(u)|,$$

$$n \cdot \rho_{CvM}(\hat{F}_n,\, F) = \int_0^1 \mathbb{U}_n^2(u) du.$$

Corolario 4.2. *Bajo los supuestos del Teorema 4.1, se cumple que*

$$\sqrt{n} \cdot \rho_{KS}(\hat{F}_n,\, F) \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} |B_t^0|,$$

$$n \cdot \rho_{CvM}(\hat{F}_n,\, F) \xrightarrow{\mathcal{D}} \int_0^1 \left[B_t^0\right]^2 dt =: W^2,$$

donde $(B_t^0)_{0 \leq t \leq 1}$ denota un puente browniano.

4.2.1. Tests para hipótesis nula simple

Como marco general se considera $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ una m.a.s. i.i.d. de tamaño n procedente de una variable aleatoria $\mathbf{Y}$ con función de distribución desconocida F y sea F_0 la función de distribución bajo H_0.

La idea básica surge de las propiedades de la distribución empírica expuestas en el Capítulo 2. Bajo H_0, para n suficientemente grande, $\hat{F}_n$ está *uniformemente* cerca de F_0, según el Teorema de Glivenko-Cantelli, bajo H_0, tenemos $\sup_t |\hat{F}_n(t) - F_0(t)| \xrightarrow{\text{c.s.}} 0\,, n \to \infty$.

Por lo tanto, cualquier medida de discrepancia entre $\hat{F}_n$ y F_0 puede usarse como un estadístico de prueba razonable. Un buen estadístico debe ser relativamente fácil de calcular y caracterizar, y dar lugar a una prueba que sea potente contra la mayoría de las distribuciones alternativas. A continuación, se presentan algunos importantes, los más conocidos y extendidos.

Prueba de Kolmogorov-Smirnov

Basado en la norma del supremo y diseñado para detectar la **máxima discrepancia** entre las distribuciones empírica y teórica, llamamos estadístico de Kolmogorov-Smirnov [20, 28] a

$$\begin{aligned} D_n := \rho_{KS}(\hat{F}_n, F_0) &= \sup_{y\in\mathbb{R}} \left|\hat{F}_n(y) - F_0(y)\right| \\ &= \max_{1\leq i\leq n} \left\{ \frac{i}{n} - F_0\left(Y(i)\right),\, F_0\left(Y_{(i)}\right) - \frac{i-1}{n} \right\}. \end{aligned}$$

Si la variable aleatoria sigue efectivamente la distribución F_0 bajo la hipótesis nula H_0, la distancia entre la función empírica F_n y F_0 será reducida, lo que se traducirá en un valor pequeño del estadístico D_n. Por el contrario, si la variable aleatoria no se ajusta a F_0 (situación en la que H_0 debería rechazarse), la discrepancia entre F_n y F_0 será significativa, generando así un valor elevado de D_n. De esta forma definimos φ_{KS} como

$$\varphi_{KS}(y) = 1 \Longleftrightarrow D_n(y) > c_n^{KS}(\alpha) \longrightarrow RC_{KS} = \{\vec{y}\,|\, D_n(\vec{y}) > c_n^{KS}(\alpha)\},$$

definiendo así la prueba de Kolmogorov-Smirnov al nivel de significación α.

Notar que, si F_0 es continua, entonces φ_{KS} es de libre distribución, es decir, la distribución de D_n bajo H_0 es independiente de la H_0 que se formule, por tanto, el valor crítico $c_n^{KS}(\alpha)$ solo depende de n y α.

Observación 28. *La distribución del estadístico está tabulada en la Tabla 1 de Anderson y Darling [1].*

Observación 29. *El test de Kolmogorov-Smirnov tiene poca sensibilidad en las colas de la distribución, siendo su punto fuerte la identificación de diferencias en la región central de los datos. Muestra potencia menor en muestras pequeñas ($n < 30$).*

Prueba de Cramér-von Mises

Basado en la norma L_2 y desarrollado para medir la **discrepancia cuadrática integrada** entre distribuciones (dando igual peso a todas las regiones de la distribución), llamamos estadístico de Cramér-von Mises [9, 31] a

$$\omega_n^2 := \rho_{CvM}(\hat{F}_n, F_0) = \int_{\mathbb{R}} \left[\hat{F}_n(z) - F_0(z)\right]^2 dF_0(z).$$

Siguiendo un razonamiento análogo al test de Kolmogorov-Smirnov, definimos φ_{CvM} como

$$\varphi_{CvM}(y) = 1 \Longleftrightarrow \omega_n^2(y) > c_n^{CvM}(\alpha) \longrightarrow RC_{CvM} = \{\vec{y}\,|\,\omega_n^2(\vec{y}) > c_n^{CvM}(\alpha)\},$$

definiendo así la prueba de Cramér-von Mises al nivel de significación α.

Observación 30. *Notar que, bajo H_0 si F_0 es continua, entonces φ_{CvM} es libre de distribución.*

Lema 4.3. *Se cumple que*

$$W_n^2 := n \cdot \omega_n^2 = \int_0^1 \mathbb{U}_n^2(u)du = \sum_{k=1}^{n} \left(U_{k:n} - \frac{2k-1}{2n} \right)^2 + \frac{1}{12n},$$

donde $U_{1:n} < \ldots < U_{n:n}$ son los estadísticos de orden de las variables aleatorias uniformemente distribuidas $U_1, \ldots, U_n$ que definen el proceso empírico reducido $\mathbb{U}_n$, donde $\mathbb{U}_n(u) = \sqrt{n}\left(\hat{G}_n(u) - u\right)$ para $0 \leq u \leq 1$.

Por tanto, una manera simplificada de representar este estadístico viene de la mano de la expresión

$$W_n^2 = n \cdot \omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(F_0(Y_{(i)}) - \frac{2i-1}{2n} \right)^2.$$

Observación 31. *El test de Cramér-von Mises es óptimo para detectar diferencias en la distribución acumulada, aunque da poca importancia a las colas de la distribución y presenta mayor complejidad computacional.*

Prueba de Anderson-Darling

Continuando con la filosofía del estadístico de Cramér-von Mises, se presenta el estadístico de Anderson-Darling [2], también basado en la norma L_2 pero introduciendo un peso para intentar subsanar la sensibilidad a las colas de la distribución. El estadístico de este test es:

$$\begin{aligned} A_n^2 = n \cdot \rho_{AD}(\hat{F}_n, F_0) &= n \int_{-\infty}^{\infty} \frac{[F_n(z) - F_0(z)]^2}{F_0(z)[1 - F_0(z)]} \, dF_0(z) \\ &= -n - \sum_{i=1}^{n} \frac{2i-1}{n} \left[\log\left(F_0(Y_{(i)})\right) + \log\left(1 - F_0(Y_{(n-i+1)})\right) \right]. \end{aligned}$$

Siguiendo un razonamiento análogo a los test anteriores, definimos φ_{AD} como

$$\varphi_{AD}(y) = 1 \Longleftrightarrow A_n^2(y) > c_n^{AD}(\alpha) \longrightarrow RC_{AD} = \{\vec{y} \,|\, A_n^2(\vec{y}) > c_n^{AD}(\alpha)\},$$

definiendo así la prueba de Anderson-Darling al nivel de significación α.

Observación 32. *Notar que, bajo H_0 si F_0 es continua, entonces φ_{AD}, es decir, A_n^2 es libre de distribución.*

Observación 33. *El test de Anderson-Darling es óptimo para distribuciones con colas pesadas, pues es especialmente sensible para esta finalidad. Sin embargo, requiere cálculo de logaritmos para cada punto e implica alta complejidad computacional.*

4.2.2. Tests para familias paramétricas

Motivados por el estadístico de Lilliefors [21] —una modificación del estadístico de Kolmogorov-Smirnov para pruebas de bondad de ajuste en hipótesis compuestas, diseñado específicamente para evaluar normalidad a partir de parámetros estimados por máxima verosimilitud—, en este trabajo se trabaja con adaptaciones de los tests expuestos en la sección anterior, incorporando parámetros estimados (a partir de los datos). Quizás lo más simple que viene a la cabeza en este caso es comparar la "mejor" distribución de la clase con la función de distribución empírica para una muestra. Un acercamiento razonable consiste en una estrategia de 2 pasos:

1. Estimar θ por $\hat{\theta} = \hat{\theta}(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ a partir de una muestra aleatoria simple.

2. Aplicar uno de los test vistos para hipótesis simples reemplazando F_0 por $F_{\hat{\theta}}$, que comparan $\hat{F}_n$ con $F_{\hat{\theta}}$, donde $F_{\hat{\theta}} \in (\mathcal{F}_\theta)_{\theta\in\Theta}$.

Sin embargo, el hecho de que $\hat{\theta} = \hat{\theta}(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ dependa de los datos de partida, tiene como consecuencia que el **proceso empírico estimado** $\sqrt{n}\left(\hat{F}_n - F_{\hat{\theta}}\right)(\cdot)$ tendrá, incluso bajo la hipótesis nula, un comportamiento estocástico diferente respecto al de $\sqrt{n}\left(\hat{F}_n - F\right)(\cdot)$, debido a la variabilidad adicional contribuida por $\hat{\theta}$. Además, desafortunadamente, los tests tipo Kolmogorov-Smirnov, Cramér-von Mises y Anderson-Darling basados en $\sqrt{n}\left(\hat{F}_n - F_{\hat{\theta}}\right)(\cdot)$ dejan de ser de libre distribución en general. Sin embargo, sí son de libre distribución o libres de parámetros para cada familia de localización-escala, es decir, la distribución bajo la hipótesis nula sólo depende de la familia de partida $(\mathcal{F}_\theta)_{\theta\in\Theta}$, pero no de θ.

Observación 34. *Notar que introducir la estimación de un parámetro en estos estadísticos afecta a la distribución de estos mismos; de hecho, bajo la hipótesis nula, la distribución será fuertemente influenciada por el tipo de estimador utilizado. Por tanto, ya no es posible utilizar las distribuciones asintóticas dadas en el caso de hipótesis simples. Si no, esto daría lugar a procedimientos inadecuados de testeo, derivando errores graves.*

Por lo tanto, obtenemos los siguientes análogos de los estadísticos de prueba KS, CvM

y AD e introducimos su nueva notación:

$$\bar{D}_n = \sup_t |\hat{F}_n(t) - F_{\hat{\theta}_n}(t)|,$$
$$\bar{W}_n^2 = n \cdot \int (\hat{F}_n - F_{\hat{\theta}_n})^2 \, dF_{\hat{\theta}_n},$$
$$\bar{A}_n^2 = n \cdot \int \frac{(\hat{F}_n - F_{\hat{\theta}_n})^2}{F_{\hat{\theta}_n}(1 - F_{\hat{\theta}_n})} \, dF_{\hat{\theta}_n}.$$

A continuación se presentan algunas propiedades relacionadas con transformaciones medibles, con vista a estadísticos, las cuales serán de importancia fundamental en el desarrollo del presente capítulo.

Definición 4.5. *Sean $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ variables aleatorias independientes e idénticamente distribuidas en los reales, y sea $T : \mathbb{R}^n \longrightarrow \mathbb{R}$ una transformación medible. Entonces llamamos a T:*

- *Equivariante de localización-escala en distribución, si:*

$$\forall b > 0 : \ \forall a \in \mathbb{R} : \ T\left(a + b\mathbf{Y}_1, \ldots, a + b\mathbf{Y}_n\right) \overset{\mathcal{D}}{=} a + bT\left(\mathbf{Y}_1, \cdots, \mathbf{Y}_n\right).$$

- *Invariante de localización y equivariante de escala en distribución, si:*

$$\forall b > 0 : \ \forall a \in \mathbb{R} : \ T\left(a + b\mathbf{Y}_1, \ldots, a + b\mathbf{Y}_n\right) \overset{\mathcal{D}}{=} bT\left(\mathbf{Y}_1, \cdots, \mathbf{Y}_n\right).$$

Siguiendo un hilo invisible pero constante, se prueba en el siguiente resultado el carácter de los estimadores máximo verosímiles, en relación con la definición previa.

Lema 4.4. *[11, Cap. 3] Sea una LoScF generada por la distribución estandarizada $\mathcal{Q}_{0,1}$ arbitraria, $\mathbf{X} \overset{\mathcal{D}}{=} \mu + \sigma\mathbf{Z}$ con $\mu \in \mathbb{R}$ y $\sigma \in (0, \infty)$, con sus correspondientes funciones de distribución $F_{\mu,\sigma}$. Sea $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$ una muestra aleatoria simple de variables i.i.d. de la LoScF. Asumiendo la existencia de los estimadores máximo verosímiles $\hat{\mu} = \hat{\mu}\left(\mathbf{Y}_1, \cdots, \mathbf{Y}_n\right)$ de μ y $\hat{\sigma} = \hat{\sigma}\left(\mathbf{Y}_1, \cdots, \mathbf{Y}_n\right)$ de σ. Entonces $\hat{\mu}$ es equivariante de localización-escala en distribución y $\hat{\sigma}$ es invariante de localización y equivariante de escala en distribución.*

Demostración. Sea $\mathbf{Y}_k \equiv \mathcal{Q}_{\mu,\sigma}$ con $k \in \{1, \ldots, n\}$ arbitrario y unas constantes fijas $a \in \mathbb{R}$ y $b \in (0, \infty)$. Entonces, para $y \in \mathbb{R}$:

$$\mathbb{P}\left(a + b\mathbf{Y}_k \leq y\right) = F_{\mu,\sigma}\left(\frac{y - a}{b}\right) = F_{0,1}\left(\frac{y - a - b\mu}{b\sigma}\right). \tag{4.1}$$

Por otro lado, si $\mathbf{Y}_k \equiv \mathcal{Q}_{a+b\mu,b\sigma}$, observamos que

$$\mathbb{P}(\mathbf{Y}_k \leq y) = F_{0,1}\left(\frac{y - a - b\mu}{b\sigma}\right). \tag{4.2}$$

Obviamente la parte derecha 4.1 y 4.2 coinciden. Notar además, que el índice k es arbitrario. Juntando todo se tiene que

$$\hat{\mu}\left(a+b\mathbf{Y}_1,\ldots,a+b\mathbf{Y}_n\right) \overset{\mathcal{D}}{=} \widehat{a+b\mu}\left(\mathbf{Y}_1,\ldots,\mathbf{Y}_n\right).$$

Ahora por el carácter invariante bajo parametrización del estimador máximo verosímil (cf., eg., Zehna [33] o la sección 2.8 de Barndorf-Nielsen y Cox [3]) tenemos directamente que $\widehat{a+b\mu} = a+b\hat{\mu}$. Por tanto, $\hat{\mu}$ es equivariante de localización-escala en distribución.

Análogamente, como la parte derecha de 4.1 y 4.2 coinciden, tenemos que

$$\hat{\sigma}\left(a+b\mathbf{Y}_1,\ldots,a+b\mathbf{Y}_n\right) \overset{\mathcal{D}}{=} \widehat{b\sigma}\left(\mathbf{Y}_1,\ldots,\mathbf{Y}_n\right).$$

De nuevo, por el carácter invariante bajo parametrización del estimador máximo verosímil concluimos que $\widehat{a+b\sigma} = b\hat{\sigma}$ y por tanto $\hat{\sigma}$ es invariante de localización y equivariante de escala en distribución. □

Lema 4.5. *Sea una LoScF generada por la distribución estandarizada $\mathcal{Q}_{0,1}$ arbitraria, $\mathbf{X} \overset{\mathcal{D}}{=} \mu+\sigma\mathbf{Z}$ con $\mu \in \mathbb{R}$ y $\sigma \in (0,\infty)$, con sus correspondientes funciones de distribución $F_{\mu,\sigma}$. Sea $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$ una muestra aleatoria simple de variables i.i.d. de la LoScF. Sea un estimador genérico equivariante de localización-escala en distribución $\hat{T}_\mu = \hat{T}_\mu\left(\mathbf{Y}_1,\ldots,\mathbf{Y}_n\right)$ de μ y un estimador genérico invariante de localización y equivariante de escala $\hat{T}_\sigma = \hat{T}_\sigma\left(\mathbf{Y}_1,\ldots,\mathbf{Y}_n\right)$ de σ. Entonces, las funciones*

$$\frac{\hat{T}_\mu-\mu}{\sigma}, \qquad\qquad \frac{\hat{T}_\sigma}{\sigma},$$

son funciones pivotales (cantidades pivote), es decir, sus distribuciones no dependen de μ ni σ.

Demostración. Tenemos $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$ una muestra aleatoria simple de variables i.i.d. de la LoScF generada por la distribución estandarizada $\mathcal{Q}_{0,1}$, $\mathbf{X} \overset{\mathcal{D}}{=} \mu+\sigma\mathbf{Z}$ con $\mu \in \mathbb{R}$ y $\sigma \in (0,\infty)$, con sus correspondientes funciones de distribución $F_{\mu,\sigma}$. Estandarizando uno a uno la m.a.s. $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$ v.a.s i.i.d., tenemos $\frac{\mathbf{Y}_1-\mu}{\sigma},\ldots,\frac{\mathbf{Y}_n-\mu}{\sigma}$ una m.a.s. de v.a.s i.i.d. con sus correspondientes funciones de distribución $F_{0,1}$. Además, sabemos que $\hat{T}_\mu$ es equivariante de localización-escala en distribución y $\hat{T}_\sigma$ es invariante de localización y equivariante de escala en distribución. Así conseguimos

$$\hat{T}_\mu\left(\frac{\mathbf{Y}_1-\mu}{\sigma},\ldots,\frac{\mathbf{Y}_n-\mu}{\sigma}\right) \overset{\mathcal{D}}{=} \frac{\hat{T}_\mu\left(\mathbf{Y}_1,\ldots,\mathbf{Y}_n\right)-\mu}{\sigma},$$
$$\hat{T}_\sigma\left(\frac{\mathbf{Y}_1-\mu}{\sigma},\ldots,\frac{\mathbf{Y}_n-\mu}{\sigma}\right) \overset{\mathcal{D}}{=} \frac{\hat{T}_\sigma\left(\mathbf{Y}_1,\ldots,\mathbf{Y}_n\right)}{\sigma}.$$

Concluimos que las funciones $\frac{\hat{T}_\mu - \mu}{\sigma}$ y $\frac{\hat{T}_\sigma}{\sigma}$ son pivotales, pues sus distribuciones no dependen de μ ni σ. □

Corolario 4.6. *Bajo las suposiciones del Lema 4.4, y del Lema 4.5, las funciones*

$$\frac{\hat{\mu} - \mu}{\sigma}, \qquad \frac{\hat{\sigma}}{\sigma},$$

son funciones pivotales (cantidades pivote), es decir, sus distribuciones no dependen de μ ni σ.

Demostración. Caso especial de la demostración del Lema 4.5, pues como se ha visto en el Lema 4.4, tenemos que para los estimadores máximo verosímiles: $\hat{\mu}$ es equivariante de localización-escala en distribución y $\hat{\sigma}$ es invariante de localización y equivariante de escala en distribución. □

Consecuentemente, si somos capaces de escribir test estadísticos únicamente en función de estas cantidades pivote, concluiríamos que la distribución de los test estadísticos bajo la hipótesis nula está completamente determinada por $F_{0,1}$, es decir, son de libre distribución o libres de parámetros para cada familia de localización-escala.

Teorema 4.7. *Bajo las asunciones del Lema 4.5, y recordando adicionalmente que $F_{\mu,\sigma}$ es continua para toda $\mu \in \mathbb{R}$ y $\sigma \in (0, \infty)$. Entonces, la distribución del proceso empírico estimado $\sqrt{n}\left(\hat{F}_n - F_{\hat{T}_\mu, \hat{T}_\sigma}\right)(\cdot)$ no depende de μ ni de σ.*

Demostración. Recordar que

$$\hat{F}_n(y) - F_{\hat{T}_\mu, \hat{T}_\sigma}(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\mathbf{Y}_i \leq y\} - F_{\hat{T}_\mu, \hat{T}_\sigma}(y), \ \forall y \in \mathbb{R}.$$

Sustituyendo ahora $y := F^{-1}_{\hat{T}_\mu, \hat{T}_\sigma}(u), \ 0 \leq u \leq 1$. De esto resulta

$$\hat{F}_n(y) - F_{\hat{T}_\mu, \hat{T}_\sigma}(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\mathbf{Y}_i \leq F^{-1}_{\hat{T}_\mu, \hat{T}_\sigma}(u)\} - u = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{F_{\hat{T}_\mu, \hat{T}_\sigma}(\mathbf{Y}_i) \leq u\} - u,$$

porque $F_{\hat{T}_\mu, \hat{T}_\sigma}$ es continua. Finalmente, notar que para todo $1 \leq i \leq n$, la distribución de

$$\mathbf{Z}_i := F_{\hat{T}_\mu, \hat{T}_\sigma}(\mathbf{Y}_i) = F_{0,1}\left(\frac{\mathbf{Y}_i - \hat{T}_\mu}{\hat{T}_\sigma}\right) = F_{0,1}\left(\frac{\frac{\mathbf{Y}_i - \mu}{\sigma} - \frac{\hat{T}_\mu - \mu}{\sigma}}{\hat{T}_\sigma / \sigma}\right), \tag{4.3}$$

no depende de μ ni σ debido al Lema 4.5. □

Observación 35. *La parte derecha de la Ecuación (4.3) revela que, aún bajo la hipótesis nula, la distribución de* $\mathbf{Z}_i$ *será, como norma general, no uniforme para* $1 \leq i \leq n$*, pues* $\frac{\mathbf{Y}_i - \mu}{\sigma} \equiv F_{0,1}$ *y las funciones pivote* $\frac{\hat{T}_\mu - \mu}{\sigma}$ *y* $\frac{\hat{T}_\sigma}{\sigma}$ *tendrán típicamente distribuciones no degeneradas, por lo menos para valores finitos de* n*.*

Corolario 4.8. *El Teorema 4.7 presenta carácter general. Un caso específico del anterior resultado se da para los estimadores máximo verosímiles* $\hat{\mu}$ *y* $\hat{\sigma}$*.*

Tras este resultado de vital importancia, es posible dilucidar que los estadísticos KS, CvM y AD son de libre distribución para cada familia de localización-escala de partida.

Proposición 4.9. *Bajo las asunciones del Teorema 4.7, se verifica que los estadísticos Kolmogorov-Smirnov, Cramér-von Mises (CvM) y Anderson-Darling (AD) son de libre distribución para cada familia de localización-escala de partida* $F \in \mathcal{F}$ *.*

Demostración. Tenemos $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ una muestra aleatoria simple de variables i.i.d. de la LoScF generada por la distribución estandarizada $\mathcal{Q}_{0,1}$, $\mathbf{X} \overset{\mathcal{D}}{=} \mu + \sigma\mathbf{Z}$ con $\mu \in \mathbb{R}$ y $\sigma \in (0, \infty)$, con sus correspondientes funciones de distribución $F_{\mu,\sigma}$. Estandarizando uno a uno la m.a.s. $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ v.a.s i.i.d., tenemos $\frac{\mathbf{Y}_1 - \mu}{\sigma}, \ldots, \frac{\mathbf{Y}_n - \mu}{\sigma}$ una m.a.s. de v.a.s i.i.d. con sus correspondientes funciones de distribución $F_{0,1}$. Además, sabemos que $\hat{T}_\mu$ es equivariante de localización-escala en distribución y $\hat{T}_\sigma$ es invariante de localización y equivariante de escala en distribución. Dividimos la demostración en tres partes, una para cada estadístico.

La prueba del estadístico de Kolmogorov-Smirnov es directa a partir del Teorema 4.7, basta seguir un razonamiento análogo.

El estadístico CvM con parámetros estimados es:

$$\bar{W}_n^2 = n \int_{-\infty}^{\infty} \left(\hat{F}_n(y) - F_{\hat{T}_\mu, \hat{T}_\sigma}(y) \right)^2 dF_{\hat{T}_\mu, \hat{T}_\sigma}(y).$$

Realizamos el cambio de variable $y = z\hat{T}_\sigma + \hat{T}_\mu$:

$$\begin{aligned}
\bar{W}_n^2 &= \left[dF_{\hat{T}_\mu, \hat{T}_\sigma}(y) = f_{\hat{T}_\mu, \hat{T}_\sigma}(y)dy = \left(\frac{1}{\hat{T}_\sigma} f_{0,1}(z) \right) \hat{T}_\sigma dz = f_{0,1}(z)dz = dF_{0,1}(z) \right] \\
&= n \int_{-\infty}^{\infty} \left(\hat{F}_n(\hat{T}_\sigma z + \hat{T}_\mu) - F_{0,1}(z) \right)^2 dF_{0,1}(z) \\
&= n \int_{-\infty}^{\infty} \left(\frac{1}{n} \sum_{i=1}^{n} I\left(\frac{\mathbf{Y}_i - \hat{T}_\mu}{\hat{T}_\sigma} \leq z \right) - F_{0,1}(z) \right)^2 dF_{0,1}(z) \\
&= n \int_{-\infty}^{\infty} \left(\frac{1}{n} \sum_{i=1}^{n} I(\mathbf{Z}_i \leq z) - F_{0,1}(z) \right)^2 dF_{0,1}(z).
\end{aligned}$$

donde $\mathbf{Z}_i = \frac{\mathbf{Y}_i - \hat{T_\mu}}{\hat{T_\sigma}} = \frac{\frac{\mathbf{Y}_i - \mu}{\sigma} - \frac{\hat{T_\mu} - \mu}{\sigma}}{\hat{T_\sigma}/\sigma}$ no depende de μ ni σ.

El estadístico AD con parámetros estimados es:

$$\bar{A_n}^2 = n \int_{-\infty}^{\infty} \frac{\left(\hat{F}_n(y) - F_{\hat{T_\mu},\hat{T_\sigma}}(y)\right)^2}{F_{\hat{T_\mu},\hat{T_\sigma}}(y)(1 - F_{\hat{T_\mu},\hat{T_\sigma}}(y))} dF_{\hat{T_\mu},\hat{T_\sigma}}(y).$$

Realizando el mismo cambio de variable $y = z\hat{T_\sigma} + \hat{T_\mu}$:

$$\begin{aligned}
\bar{A_n}^2 &= \left[dF_{\hat{T_\mu},\hat{T_\sigma}}(y) = f_{\hat{T_\mu},\hat{T_\sigma}}(y)dy = \left(\frac{1}{\hat{T_\sigma}} f_{0,1}(z)\right) \hat{T_\sigma} dz = f_{0,1}(z)dz = dF_{0,1}(z) \right] \\
&= n \int_{-\infty}^{\infty} \frac{\left(\hat{F}_n(\hat{T_\sigma} z + \hat{T_\mu}) - F_{0,1}(z)\right)^2}{F_{0,1}(z)(1 - F_{0,1}(z))} dF_{0,1}(z) \\
&= n \int_{-\infty}^{\infty} \frac{\left(\frac{1}{n}\sum_{i=1}^{n} I(\mathbf{Z}_i \leq z) - F_{0,1}(z)\right)^2}{F_{0,1}(z)(1 - F_{0,1}(z))} dF_{0,1}(z).
\end{aligned}$$

Nuevamente, por las propiedades de los estimadores, la distribución del estadístico no depende de los parámetros originales. □

La distribución del estadístico bajo H_0 rara vez puede calcularse analíticamente, pero puede aproximarse, con la precisión deseada, aumentando el número de réplicas, mediante simulación de Monte Carlo de la siguiente manera:

(I) Generar una muestra aleatoria $\mathbf{X}_1, \ldots, \mathbf{X}_n$ a partir de $F_{0,1}$.

(II) Calcular el estadístico de prueba para esta muestra aleatoria. Esto significa calcular $T_{\hat{\mu}} = T_{\hat{\mu}}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ y $T_{\hat{\sigma}} = T_{\hat{\sigma}}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, luego calcular $\mathbf{Z}_i = F_{0,1}((\mathbf{X}_i - T_{\hat{\mu}})/T_{\hat{\sigma}})$ y finalmente calcular el estadístico deseado $\bar{D}_n$, $\bar{C}_n$ o $\bar{A}_n$.

(III) Repetir los pasos (I) y (II) un gran número de veces.

4.3. GoF a partir de coeficientes de asimetría y/o curtosis

En años recientes, enfoques alternativos basados en coeficientes de asimetría y curtosis han ganado relevancia por su sensibilidad para detectar desviaciones específicas en la forma de la distribución. Estos coeficientes, al ser invariantes ante transformaciones de localización y escala, ofrecen una herramienta robusta para evaluar el ajuste distribucional, particularmente en distribuciones simétricas o cercanas a la normalidad. En esta sección se presentan algunos de estos tests que han sido explorados recientemente, así como ventajas y posibilidades de construcción de nuevos tests a partir de los propios.

En [15] se prueba que la versión muestral del coeficiente de asimetría de momentos es asintóticamente normal bajo algunas condiciones de regularidad. Por tanto, como se desarrolla en [17, 5] se espera que

$$\sqrt{n}\frac{\hat{\gamma}-\gamma_F}{\sqrt{V(\gamma,F)}} \to N(0,1),$$

donde $\hat{\gamma}$ es la versión muestral del coeficiente de asimetría γ, γ_F es el valor poblacional para el coeficiente de asimetría γ en la distribución F, y $V(\gamma,F)$ denota la varianza asintótica de γ en la distribución F (para la LoScF $\mathcal{F}$ es constante y es posible calcularla explícitamente). Nótese que $\gamma_F = 0$ para todas las distribuciones simétricas; sin embargo, la varianza asintótica $V(\gamma,F)$ depende de la distribución subyacente y varía incluso de una distribución simétrica a otra.

Dado que, como hemos visto, los coeficientes de asimetría son invariantes ante localización y escala, se cumple que $\gamma_F = \gamma_{F'}$ y $V(\gamma,F) = V(\gamma,F')$, para cualquier $F, F' \in \mathcal{F}$. Adicionalmente, se sigue para cualquier coeficiente de asimetría asintóticamente normal que $\sqrt{n}\frac{\hat{\gamma}-\gamma_F}{\sqrt{V(\gamma,F)}} \to N(0,1)$ bajo la hipótesis nula (es decir, para cualquier $F \in \mathcal{F}$), independientemente de los valores de los parámetros de localización y escala. Por lo tanto, basta con calcular γ_F y $V(\gamma,F)$ para definir una región de rechazo basada en el valor muestral de $\hat{\gamma}$. Formalmente, la región de rechazo se define como sigue:

$$RC = \left\{ x \in R^n \middle| |\gamma(x) - \gamma_F| > \frac{\sqrt{V(\gamma,F)}}{\sqrt{n}} z_{1-\alpha/2} \right\},$$

donde $z_{1-\alpha/2}$ es el cuantil de orden $1-\alpha/2$ de una distribución normal.

Observación 36. *Es posible obtener más tests a partir de este, simplemente considerando nuevos coeficientes de asimetría que verifiquen la condición de ser asintóticamente normales bajo ciertas condiciones de regularidad.*

Desafortunadamente, los tests de bondad de ajuste basados en coeficientes de asimetría tienden a exhibir una baja potencia, por lo que es común considerar tests que combinan coeficientes de asimetría y coeficientes de curtosis. Dentro de estos últimos destaca el test de Jarque-Bera [18, 19] inicialmente diseñado específicamente para testear la normalidad. Como se demostró en Moors et al. [22], bajo la asunción de normalidad, se da la siguiente relación

$$\sqrt{n}\begin{pmatrix}\hat{\gamma}\\ \hat{K}\end{pmatrix} \to_{\mathcal{D}} N_2\left(\begin{pmatrix}0\\3\end{pmatrix}, \begin{pmatrix}6 & 0\\ 0 & 24\end{pmatrix}\right),$$

lo que conduce al estadístico de prueba de Jarque-Bera (donde $\hat{K}$ denota la versión muestral del coeficiente de curtosis):

$$T = n\left(\frac{\hat{\gamma}^2}{6} + \frac{(\hat{K}-3)^2}{24}\right) \approx \chi^2_2.$$

Más aún, en la publicación de Brys et al. [6] se aporta una generalización del estadístico.

4.4. PP-plots y QQ-Plots

Los gráficos de probabilidad proporcionan una forma de realizar pruebas de bondad de ajuste de una manera visual. Ofrecen una herramienta rápida para verificar si cierta suposición distribucional es razonable, a pesar de que estas no son pruebas formales. Realmente son métodos gráficos para el diagnóstico de diferencias entre la distribución de probabilidad de una población de la que se ha extraído una muestra aleatoria y una distribución usada para la comparación, siendo habitualmente esta última una distribución teórica.

Sea una LoScF generada por la distribución estandarizada $\mathcal{Q}_{0,1}$ arbitraria, $\mathbf{X} \overset{\mathcal{D}}{=} \mu + \sigma\mathbf{Z}$ con $\mu \in \mathbb{R}$ y $\sigma \in (0,\infty)$, con sus correspondientes funciones de distribución $F_{\mu,\sigma}$, denotamos por $\mathcal{F}$ a $\mathcal{F} = \{F_{\mu,\sigma} : \mu \in \mathbb{R}, \sigma > 0\}$. Notar que $\mathbf{X} \sim F_{0,1}$, por tanto $F_{\mu,\sigma}(x) = F_{0,1}\left(\frac{x-\mu}{\sigma}\right)$. Para asegurar la comprensión del concepto de estos gráficos, se ejemplifica la situación.

Ejemplo 4.1. *Consideramos una variable aleatoria $\mathcal{N}(\mu,\sigma^2)$, que se obtiene a partir de una variable aleatoria normal estandarizada Z mediante la transformación lineal $\mathbf{Z} \mapsto \mu + \sigma\mathbf{Z}$. Sean $\mathbf{X}_1,\ldots,\mathbf{X}_n$ datos arbitrarios de alguna distribución. Así, tenemos que $F^{-1}_{\mu,\sigma}(\hat{F}_n(\mathbf{X}_{(i)})) = F^{-1}_{\mu,\sigma}\left(\frac{i}{n}\right)$, recordando la relación $\hat{F}_n(\mathbf{X}_{(i)}) = i/n$. Ahora bien, si los datos provienen de una distribución en $\mathcal{F}$, entonces $\hat{F}_n(\mathbf{X}_{(i)}) \approx F_{\mu,\sigma}(\mathbf{X}_{(i)})$, y así, $\mathbf{X}_{(i)} \approx F^{-1}_{\mu,\sigma}\left(\frac{i}{n}\right) = \mu + \sigma F^{-1}_{0,1}\left(\frac{i}{n}\right)$. Dicho de otra manera, si los datos provienen de una distribución en $\mathcal{F}$, esperamos que los puntos $\left(\mathbf{X}_{(i)}, F^{-1}_{0,1}\left(\frac{i}{n+1}\right)\right)$ se encuentren aproximadamente en una línea recta. Notar que se ha reemplazado $\frac{i}{n}$ por $\frac{i}{n+1}$ para asegurar el no evaluar $F^{-1}_{0,1}(1)$, que puede ser infinito. El gráfico de estos puntos se conoce como* ***gráfico cuantil-cuantil (QQ-Plot)****.*

Para construir el gráfico de probabilidades (PP-Plot), comparamos las probabilidades teóricas con las empíricas. Si los datos provienen de una distribución en $\mathcal{F}$, entonces esperamos que $\hat{F}_n(X_{(i)}) \approx F_{\mu,\sigma}(\mathbf{X}_{(i)})$, y por tanto, los puntos $\left(F_{\mu,\sigma}(\mathbf{X}_{(i)}), \frac{i}{n+1}\right)$ deberían estar cercanos a la diagonal $y = x$. De nuevo, el uso de $\frac{i}{n+1}$ en lugar de $\frac{i}{n}$ se debe a

una corrección común para evitar comparar directamente con 1, especialmente cuando se usan funciones de distribución con soporte no acotado. El gráfico de estos puntos se conoce como ***gráfico de probabilidad-probabilidad (PP-Plot)****. Si los datos provienen de la distribución teórica, se espera que los puntos sigan aproximadamente una línea recta creciente con pendiente 1 y ordenada al origen 0.*

El uso de gráficos de probabilidad requiere cierta práctica, pero son muy comunes y útiles. Si se desea hacer la comprobación para una distribución F_θ, el razonamiento anterior implica que los puntos para el QQ-Plot $\left(F_\theta^{-1}\left(\frac{i}{n+1}\right), \mathbf{X}_{(i)}\right)$ ó para el PP-Plot $\left(F_\theta(\mathbf{X}_{(i)}), \frac{i}{n+1}\right)$ deberían estar en una línea recta si θ es conocido. Si θ es desconocido, se pueden utilizar parámetros estimados $\hat{\theta}$.

En la Figura 4.1, es posible observar un ejemplo de PP-Plot y QQ-Plot para datos generados a partir de una distribución normal y parámetros estimados por máxima verosimilitud.

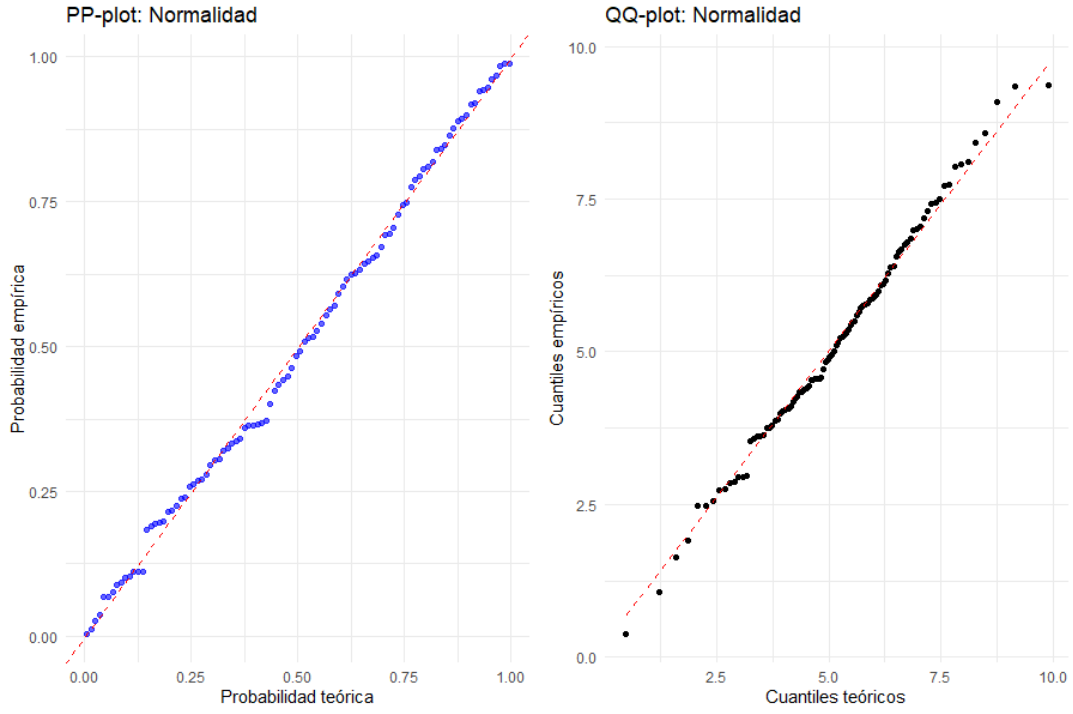

Figura 4.1: PP-Plot y QQ-Plot a datos provenientes de una LoScF normal con parámetros esimados por máxima-verosimilitud.

CAPÍTULO 5

Experimentación

En este Capítulo se expone la metodología seguida para la experimentación, así como un análisis de los resultados obtenidos para los distintos tests seleccionados, considerando dos posibilidades para los estimadores de los parámetros de localización y escala (por un lado, estimación por máxima verosimilitud (MLE), dependiente de la familia de localización-escala; y por otro lado, el uso de la media y desviación típica muestrales). Se proporcionan tablas y mapas de calor para facilitar la comprensión de los resultados obtenidos.

5.1. Metodología

Se estudiará el comportamiento de aquellos tests basados en la función de distribución empírica (ECDF), en concreto los tests de Kolmogorov-Smirnov, Cramér-von Mises y Anderson-Darling, debido a su propiedad de ser libres de parámetros dentro de familias de localización-escala. Se evaluará el comportamiento de los distintos tests en términos de potencia estadística bajo distintos enfoques, considerando numerosos factores: diferentes LoScF (normal, logística, exponencial, uniforme y Laplace), diversos tamaños muestrales ($n \in \{30, 100\}$) y dos métodos de estimación de parámetros por un lado, el método de máxima verosimilitud (MLE) y por otro, la estimación mediante momentos (media y desviación típica).

Para lograr el objetivo de estudio, se usará en repetidas ocasiones el método de simulación por Monte Carlo con 10^4 réplicas para cada casuística. Se trabajará con el nivel de significación fijado $\alpha = 0.05$. En primera instancia, se fija una semilla (1) con el fin de poder reproducir los resultados.

Al realizar el estudio de potencias, antes se tabulará la distribución de cada estadístico

con los parámetros de estimación de localización-escala seleccionados, utilizando el método de Monte Carlo para este fin a partir de muestras de la distribución seleccionada (H_0) con parámetro de localización 0 y escala 1. Como se ha probado la libertad de parámetros dentro de la familia, esta elección es arbitraria.

Una vez tabulada la distribución del estadístico para la distribución seleccionada, de nuevo mediante método de simulación por Monte Carlo se calculará la potencia, repitiendo el procedimiento: generar una muestra aleatoria con parámetro de localización 0 y escala 1 a partir de la LoScF seleccionada para el contraste (de la que provienen los datos), calcular el valor del estadístico para dicha muestra, obtener el p-valor del test comparando este con la tabla. Finalmente se realiza la media de estos "rechazos" para conseguir el valor de la potencia deseada.

Esto se ha programado para su repetición automática en lo que respecta a todas las combinaciones de casos: familias LoScF, tamaños muestrales y métodos de estimación de parámetros. Ver código.

5.2. Estudio de potencias: Kolmogorov-Smirnov

5.2.1. Parámetros estimados por máxima verosimilitud

Este estudio evalúa la potencia del test de Kolmogorov-Smirnov (KS), utilizando estimadores de máxima verosimilitud (MLE). A la vista de las Tablas 5.1, 5.2 y las Figuras 5.1, 5.2 se revelan patrones significativos en el comportamiento del test para tamaños muestrales pequeños ($n = 30$) y grandes ($n = 100$).

Para muestras pequeñas ($n = 30$), el test muestra un adecuado control del error Tipo I, manteniendo tasas de rechazo bajo H_0 cercanas al nivel de significación $\alpha = 0.05$. Sin embargo, se observan comportamientos diferenciales según la distribución analizada.

La distribución exponencial presenta elevada potencia para detectar desviaciones de la familia, tanto al partir de datos provenientes de dicha distribución como al realizar contrastes siendo el supuesto bajo H_0 que pertenece a la LoScF exponencial. Las distribuciones normal y logística muestran una notable dificultad para ser discriminadas entre sí, reflejando su cercanía. Aparte, los contrastes entre distribuciones normales/logísticas frente a Laplace exhiben potencias bajas.

Al incrementar el tamaño muestral a $n = 100$, se mantiene el control del error Tipo I mientras se observa una mejora generalizada en la potencia: las tasas de rechazo correcto

aumentan significativamente para la mayoría de los casos; persiste la dificultad para distinguir entre las LoScF normal y logística, y mejora limitadamente la potencia en contrastes LoScF normal/logística contra la LoScF de Laplace.

Tabla 5.1: Potencias Kolmogorov-Smirnov con $n = 30$, estimadores MLE: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Distribución usada \ H_0	Laplace	Uniforme	Exponencial	Logística	Normal
Laplace	0.0512	0.7203	0.9854	0.1265	0.2816
Uniforme	0.3843	0.0500	0.6370	0.2658	0.1664
Exponencial	0.7243	0.9790	0.0462	0.7538	0.7793
Logística	0.0810	0.5196	0.9653	0.0505	0.0910
Normal	0.1258	0.3074	0.9465	0.0599	0.0475

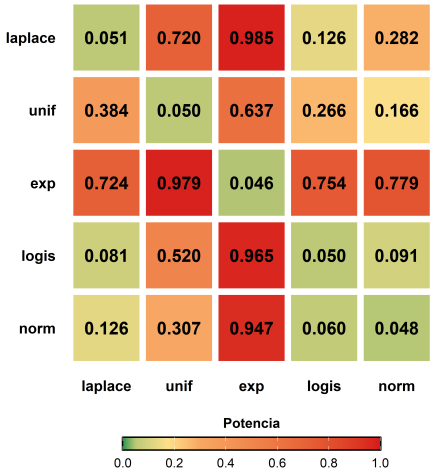

Figura 5.1: Potencias Kolmogorov-Smirnov con $n = 30$, estimadores MLE: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Tabla 5.2: Potencias Kolmogorov-Smirnov con $n = 100$, estimadores MLE: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Distribución usada \ H_0	Laplace	Uniforme	Exponencial	Logística	Normal
Laplace	0.0566	1.0000	1.0000	0.3244	0.6851
Uniforme	0.9534	0.0480	0.9991	0.8342	0.5929
Exponencial	1.0000	1.0000	0.0570	1.0000	0.9999
Logística	0.2148	0.9959	1.0000	0.0534	0.1366
Normal	0.3999	0.9693	1.0000	0.0894	0.0449

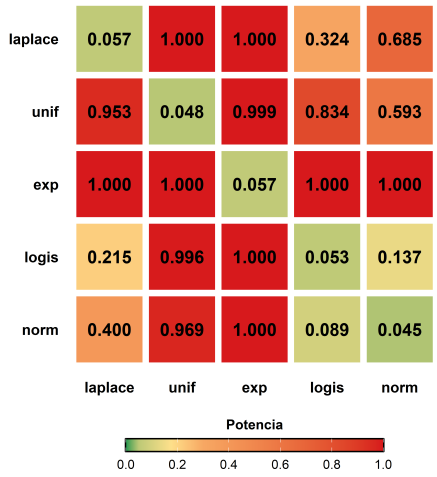

Figura 5.2: Potencias Kolmogorov-Smirnov con $n = 100$, estimadores MLE: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

5.2.2. Parámetros estimados: media desviación típica

El estudio del test de Kolmogorov-Smirnov utilizando estimadores de momentos (media y desviación típica) revela un comportamiento diferencial marcado según el tamaño muestral considerado. A la vista de las Tablas 5.3, 5.4 y las Figuras 5.3, 5.4 se demuestra que, si bien el test mantiene adecuadamente controlado el error de tipo I en torno al nivel de significación para ambos tamaños muestrales, su potencia presenta variaciones significativas para otras casuísticas.

Fijando la atención en las muestras pequeñas ($n = 30$), los resultados obtenidos son desalentadores. Exceptuando algunos casos específicos, se observa una potencia general notablemente baja y la capacidad del test para detectar desviaciones resulta insuficiente (exceptuando el caso de la LoScF exponencial). Igualmente preocupante es el pobre desempeño en los contrastes entre las LoScF Normal/Logística frente a Laplace, donde las tasas de rechazo correcto alcanzan valores mínimos. Estos hallazgos sugieren que, en contextos con tamaños muestrales reducidos, el uso de estimadores por momentos podría conducir a conclusiones erróneas con relativa frecuencia, al menos en comparación con el uso de estimadores de máxima verosimilitud.

La situación mejora sustancialmente al aumentar el tamaño muestral a $n = 100$. En este escenario, el test muestra un rendimiento destacable para la distribución exponencial, alcanzando tasas de rechazo correcto del 100 %. Las distribuciones uniforme y Laplace también presentan resultados satisfactorios. Sin embargo, persisten ciertas limitaciones: la

discriminación entre las LoScF normal y logística sigue siendo problemática, y los contrastes de estas frente a Laplace continúan mostrando potencias reducidas, aunque con cierta mejoría respecto a los resultados para $n = 30$.

Tabla 5.3: Potencias Kolmogorov-Smirnov con $n = 30$, estimadores media y desviación típica: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Distribución usada \ H_0	Laplace	Uniforme	Exponencial	Logística	Normal
Laplace	0.0469	0.6925	0.4239	0.1536	0.2983
Uniforme	0.1882	0.0478	0.4662	0.1880	0.1458
Exponencial	0.6854	0.9160	0.0541	0.7335	0.7829
Logística	0.0289	0.3878	0.2581	0.0471	0.0967
Normal	0.0418	0.1932	0.2279	0.0421	0.0476

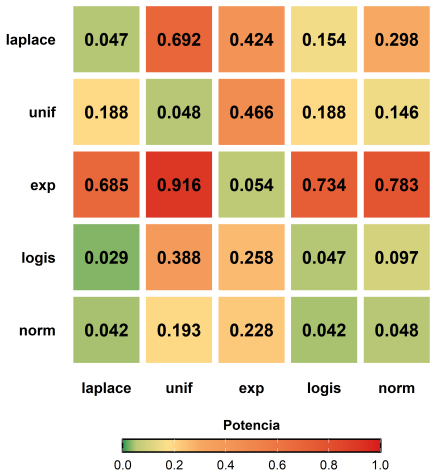

Figura 5.3: Potencias Kolmogorov-Smirnov con $n = 30$, estimadores media y desviación típica: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Tabla 5.4: Potencias Kolmogorov-Smirnov con $n = 100$, estimadores media y desviación típica: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Distribución usada \ H_0	Laplace	Uniforme	Exponencial	Logística	Normal
Laplace	0.0510	0.9983	0.9106	0.3499	0.7007
Uniforme	0.9542	0.0557	0.9960	0.8468	0.5775
Exponencial	0.9999	1.0000	0.0506	0.9998	0.9999
Logística	0.1006	0.9255	0.7454	0.0518	0.1456
Normal	0.2358	0.6582	0.7940	0.0686	0.0466

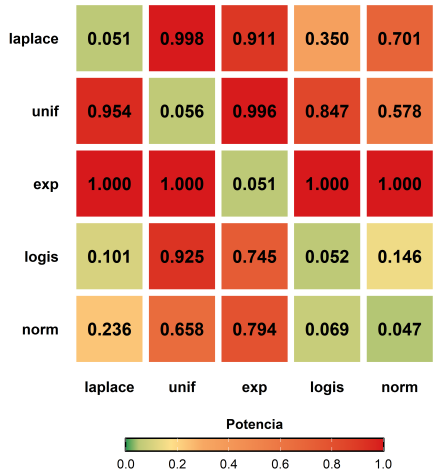

Figura 5.4: Potencias Kolmogorov-Smirnov con $n = 100$, estimadores media y desviación típica: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

5.3. Estudio de potencias: Cramér-Von-Mises

5.3.1. Parámetros estimados por máxima verosimilitud

El estudio del test de Cramér-von Mises (CvM) utilizando estimadores de máxima verosimilitud (MLE) revela un comportamiento diferencial marcado según el tamaño muestral considerado. A la vista de las Tablas 5.5, 5.6 y las Figuras 5.5, 5.6 se demuestra un comportamiento robusto en términos de control del error Tipo I para ambos tamaños muestrales, con tasas que oscilan entre 0.0463 y 0.0496, consistentemente cercanas al nivel de significación. Esta precisión en el control del error Tipo I valida la fiabilidad del test bajo la hipótesis nula.

Para muestras pequeñas ($n = 30$), el test exhibe una potencia elevada cuando los datos provienen de la LoScF exponencial, detectando eficazmente desviaciones de esta familia. Sin embargo, se observan limitaciones significativas al discriminar entre las LoScF normal y logística, donde las tasas de rechazo son cercanas al nivel de significación. Esta dificultad persiste incluso al contrastar datos provenientes de las LoScF normal/logística con la LoScF de Laplace, mostrando potencias reducidas.

Al considerar tamaños muestrales mayores ($n = 100$), el rendimiento del test mejora sustancialmente. Las potencias alcanzan valores altos para las familias exponencial, uniforme y Laplace, destacándose la precisión casi perfecta en el caso exponencial. No obstante,

la discriminación entre distribuciones normales y logísticas sigue siendo problemática, cometiendo errores elevados de tipo II. También se pone de manifiesto una baja potencia al contrastar datos logísticos bajo la hipótesis nula de Laplace (difíciles de distinguir incluso con muestras moderadamente grandes).

Tabla 5.5: Potencias Cramér-Von Mises con $n = 30$, estimadores MLE: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Distribución usada \ $\mathbf{H_0}$	Laplace	Uniforme	Exponencial	Logística	Normal
Laplace	0.0481	0.7085	0.9898	0.1432	0.3313
Uniforme	0.4364	0.0483	0.7752	0.3797	0.2576
Exponencial	0.6313	0.9867	0.0476	0.7690	0.8963
Logística	0.0721	0.4831	0.9799	0.0463	0.0991
Normal	0.1060	0.2725	0.9705	0.0621	0.0496

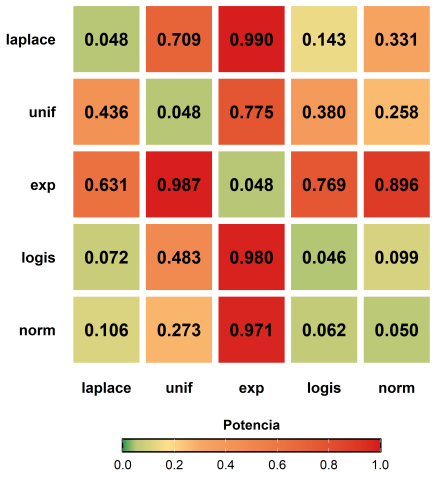

Figura 5.5: Potencias Cramér-Von Mises con $n = 30$, estimadores MLE: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Tabla 5.6: Potencias Cramér-Von Mises con $n = 100$, estimadores MLE: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Distribución usada \ H_0	Laplace	Uniforme	Exponencial	Logística	Normal
Laplace	0.0504	1.0000	1.0000	0.3983	0.8101
Uniforme	0.9939	0.0475	1.0000	0.9592	0.8572
Exponencial	0.9990	1.0000	0.0549	0.9997	1.0000
Logística	0.1909	0.9978	1.0000	0.0499	0.1943
Normal	0.4062	0.9794	1.0000	0.1006	0.0498

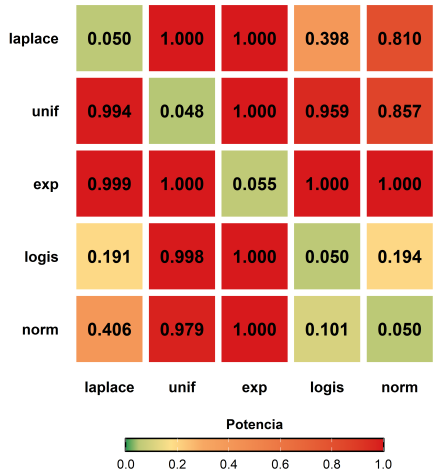

Figura 5.6: Potencias Cramér-Von Mises con $n = 100$, estimadores MLE: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

5.3.2. Parámetros estimados: media desviación típica

El estudio del test de Cramér-von Mises (CvM) utilizando estimadores de momentos (media y desviación típica), como se puede observar en las Tablas 5.7, 5.8 y las Figuras 5.7, 5.8 revela lo siguiente.

Para muestras pequeñas ($n = 30$), el test muestra un rendimiento limitado, con potencias aceptables únicamente cuando los datos proceden de la LoScF exponencial, mientras que en los demás casos presenta dificultades significativas para detectar desviaciones de la hipótesis nula. Esta situación mejora notablemente al aumentar el tamaño muestral a $n = 100$, donde el test demuestra una capacidad adecuada para rechazar correctamente la hipótesis nula en la mayoría de los escenarios considerados.

Sin embargo, persisten ciertas limitaciones que merecen especial atención. En particular, el test continúa mostrando dificultades para discriminar entre las LoScF normal

y logística, independientemente del tamaño muestral utilizado, siendo más fácil detectar diferencias cuando los datos provienen de una logística, la relación no es simétrica. Asimismo, se observa un comportamiento subóptimo cuando se contrastan datos procedentes de la LoScF logística frente a la hipótesis nula de la LoScF de Laplace, evidenciando la similaridad entre las familias.

Tabla 5.7: Potencias Cramér-Von Mises con $n = 30$, estimadores media y desviación típica: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Distribución usada \ H_0	Laplace	Uniforme	Exponencial	Logística	Normal
Laplace	0.0484	0.7681	0.5864	0.1734	0.3515
Uniforme	0.4141	0.0444	0.6799	0.3253	0.2237
Exponencial	0.8028	0.9221	0.0503	0.8644	0.8946
Logística	0.0311	0.4425	0.4830	0.0479	0.1074
Normal	0.0577	0.2147	0.4730	0.0430	0.0494

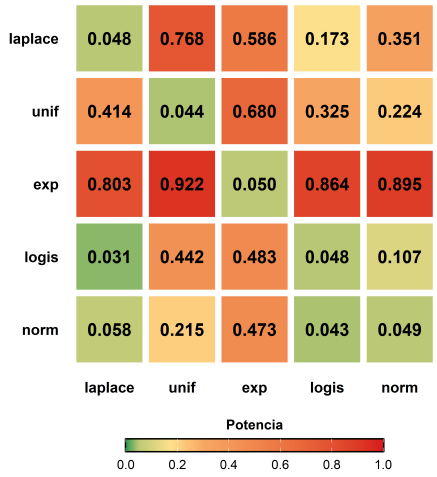

Figura 5.7: Potencias Cramér-Von Mises con $n = 30$, estimadores media y desviación típica: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Tabla 5.8: Potencias Cramér-Von Mises con $n = 100$, estimadores media y desviación típica: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Distribución usada \ H_0	Laplace	Uniforme	Exponencial	Logística	Normal
Laplace	0.0474	0.9996	0.9912	0.4083	0.8228
Uniforme	0.9977	0.0492	0.9994	0.9757	0.8419
Exponencial	1.0000	1.0000	0.0494	1.0000	1.0000
Logística	0.1369	0.9713	0.9804	0.0492	0.2080
Normal	0.4145	0.7922	0.9876	0.0877	0.0502

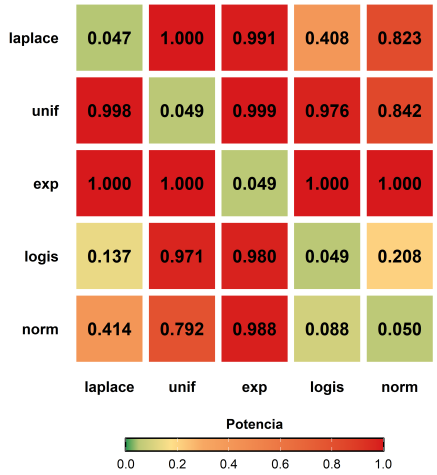

Figura 5.8: Potencias Cramér-Von Mises con $n = 100$, estimadores media y desviación típica: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

5.4. Estudio de potencias: Anderson-Darling

5.4.1. Parámetros estimados por máxima verosimilitud

El estudio del test de Anderson-Darling (AD) con estimación por máxima verosimilitud (MLE), como se puede comprobar en las Tablas 5.9, 5.10 y las Figuras 5.9, 5.10 demuestra un excelente comportamiento en términos de control del error Tipo I, manteniendo tasas entre 4.85 % y 5.02 % para ambos tamaños muestrales ($n = 30$ y $n = 100$).

Para muestras pequeñas ($n = 30$), el test muestra una potencia moderada-alta en la mayoría de los escenarios evaluados, es particularmente efectivo para detectar desviaciones de las LoScF exponencial y en menor medida, la uniforme. Los valores de potencia alcanzados son consistentemente elevados, con la excepción característica del caso norma-

l/logística, donde persiste la conocida dificultad para discriminar entre estas dos familias. Esta limitación, común a varios tests de bondad de ajuste, parece ser una propiedad intrínseca de la similitud entre estas distribuciones más que un defecto específico del test.

Al incrementar el tamaño muestral a $n = 100$, los resultados muestran estabilidad, con mejoras considerables en la mayoría de LoScF, alcanzando grandes niveles de rendimiento (rechazos correctos). Esta consistencia en el comportamiento del test para diferentes tamaños muestrales refuerza su utilidad práctica, particularmente en situaciones donde solo se dispone de muestras de tamaño moderado.

Tabla 5.9: Potencias Anderson-Darling con $n = 30$, estimadores MLE: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Distribución usada \ H_0	Laplace	Uniforme	Exponencial	Logística	Normal
Laplace	0.0505	0.5982	0.9751	0.1560	0.3432
Uniforme	0.4473	0.0517	0.6323	0.4184	0.3283
Exponencial	0.7742	0.9839	0.0485	0.8981	0.9335
Logística	0.0645	0.3798	0.9567	0.0486	0.1079
Normal	0.0944	0.1906	0.9351	0.0541	0.0517

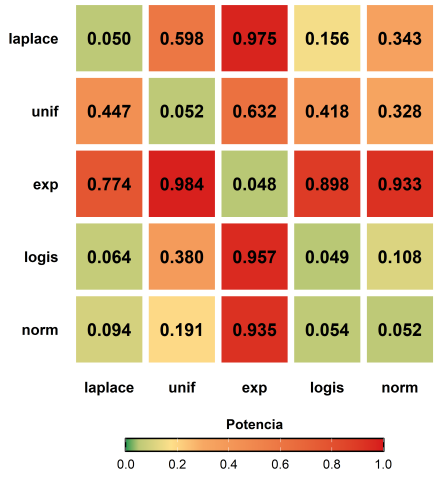

Figura 5.9: Potencias Anderson-Darling con $n = 30$, estimadores MLE: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Tabla 5.10: Potencias Anderson-Darling con $n = 100$, estimadores MLE: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Distribución usada \ H_0	Laplace	Uniforme	Exponencial	Logística	Normal
Laplace	0.0497	1.0000	1.0000	0.3734	0.8134
Uniforme	0.9980	0.0492	1.0000	0.9917	0.9543
Exponencial	1.0000	1.0000	0.0529	1.0000	1.0000
Logística	0.1565	0.9983	1.0000	0.0499	0.2227
Normal	0.3617	0.9843	1.0000	0.0924	0.0496

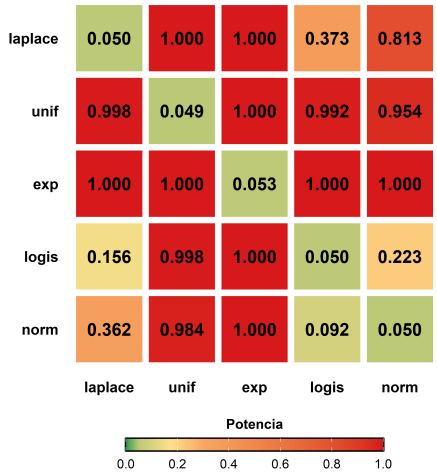

Figura 5.10: Potencias Anderson-Darling con $n = 100$, estimadores MLE: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

5.4.2. Parámetros estimados: media desviación típica

El estudio del test de Anderson-Darling (AD) utilizando estimadores de momentos (media y desviación típica), siguiendo los resultados en las Tablas 5.11, 5.12 y las Figuras 5.11, 5.12 se infieren las siguientes conclusiones.

Se demuestra un comportamiento diferenciado según el tamaño muestral. Para muestras pequeñas ($n = 30$), aunque controla adecuadamente el error Tipo I (4.74 %-5.03 %), presenta potencias desiguales: moderada-alta para datos de LooScF exponencial, moderada para LoScF uniforme, y muy baja para la LoScF de Laplace, logística y normal. En definitiva, para este tamaño muestral los resultados no son los deseables.

La situación cambia radicalmente con $n = 100$, donde las potencias mejoran muy significativamente, equiparándose a las obtenidas con máxima verosimilitud, aunque en general son ligeramente inferiores. Este cambio abrupto sugiere un umbral muestral a partir del

cual la estimación por momentos alcanza plena eficacia. No obstante, persiste la dificultad característica para discriminar entre distribuciones normales y logísticas, limitación común a varios tests de bondad de ajuste.

Estos resultados indican que, mientras para muestras pequeñas se recomienda preferentemente la estimación por máxima verosimilitud, en muestras grandes ambos métodos de estimación producen resultados suficientemente cercanos, casi equivalentes. Por tanto, cabe plantear un debate teniendo en cuenta la complejidad computacional que acompaña a la estimación máximo-verosimil cuando las fórmulas no son cerradas.

Tabla 5.11: Potencias Anderson-Darling con $n = 30$, estimadores media y desviación típica: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Distribución usada \ H_0	Laplace	Uniforme	Exponencial	Logística	Normal
Laplace	0.0478	0.1395	0.0816	0.1681	0.3588
Uniforme	0.6179	0.0474	0.6940	0.4459	0.2895
Exponencial	0.8692	0.3068	0.0495	0.9023	0.9315
Logística	0.0428	0.0931	0.1626	0.0486	0.1160
Normal	0.0898	0.0795	0.2610	0.0469	0.0503

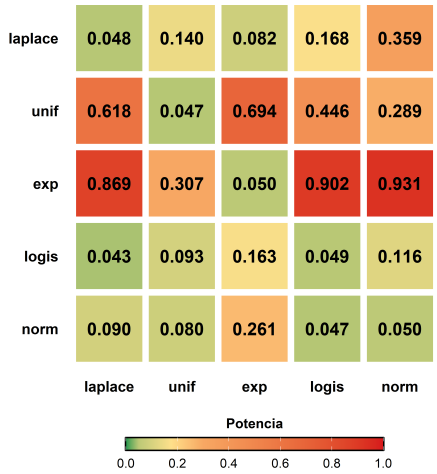

Figura 5.11: Potencias Anderson-Darling con $n = 30$, estimadores media y desviación típica: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Tabla 5.12: Potencias Anderson-Darling con $n = 100$, estimadores media y desviación típica: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

Distribución usada \ H_0	Laplace	Uniforme	Exponencial	Logística	Normal
Laplace	0.0492	0.8190	0.4726	0.3929	0.8243
Uniforme	0.9999	0.0479	0.9998	0.9967	0.9473
Exponencial	1.0000	0.9490	0.0513	1.0000	1.0000
Logística	0.1825	0.6246	0.7696	0.0491	0.2364
Normal	0.5425	0.5269	0.9429	0.1022	0.0499

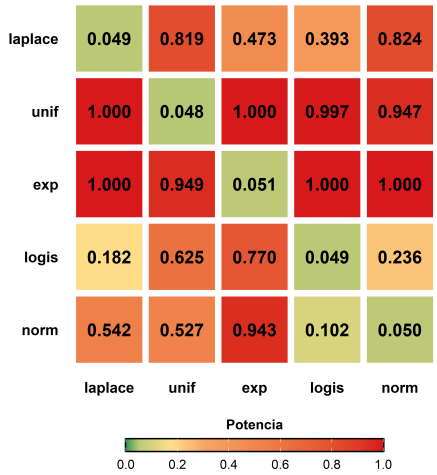

Figura 5.12: Potencias Anderson-Darling con $n = 100$, estimadores media y desviación típica: distribución muestra de partida (filas) y distribución bajo H_0 (columnas).

5.5. Síntesis

Los resultados obtenidos revelan patrones que permiten establecer recomendaciones prácticas. En primer lugar, se observa que la combinación de tests Cramér-Von Mises y Anderson-Darling con estimación por máxima verosimilitud ofrece el mejor rendimiento global, siendo similares los resultados obtenidos para sendos tamaños muestrales, ligeramente más marcados para muestras pequeñas.

De forma general, si bien para muestras grandes las diferencias entre los distintos tests y métodos de estimación se atenúan (sobre todo en el caso de parámetros estimados mediante la media y desviación típica) en el rango de tamaño muestral pequeño las ventajas de las elecciones mencionadas previamente son particularmente notorias frente a las demás.

Un hallazgo relevante es la presumible dificultad que presentan estos tests para dis-

criminar entre las LoScF normal y logística, siendo más fácil para todos ellos detectar desviaciones de la LoScF de contraste cuando los datos generados provienen de la logística. No es una relación simétrica, pues no se detecta el mismo patrón al generar datos de la LoScF normal y realizar el contraste contra la logística.

Merecedora de mención es la similitud implícita entre las LoScF de Laplace y logística, lo que provoca que las potencias de contraste entre ambas sean reducidas. De igual manera notar que esta relación no es simétrica, pues se detecta con más facilidad las desviaciones respecto a la familia de contraste cuando los datos generados provienen de la LoScF de Laplace que cuando provienen de la logística.

Estas dos últimas casuísticas particulares mencionadas anteriormente sugieren la necesidad de recurrir a pruebas más especializadas. Por el contrario, es honorífico el caso de la LoScF exponencial, para la que todos los tests sin excepción demostraron una precisión envidiable.

Por último, concluir, en vista de los resultados, que para tamaños muestrales pequeños se recomienda la elección de estimadores máximo-verosímiles, y elección de test Anderson-Darling o Cramér-von Mises, en función de la complejidad computacional que esté dispuesto a soportar el usuario. Si bien los resultados del test Kolmogorov-Smirnov con parámetros estimados no son desalentadores, existen alternativas superiores y solo se recomienda su uso práctico para fines recreativos.

CAPÍTULO 6

Conclusiones

Este estudio explora y desarrolla un marco teórico para diversos tests de bondad de ajuste compuestos aplicados a familias de localización-escala. Tras el desarrollo teórico de estos tests, se seleccionaron aquellos basados en la función de distribución empírica (ECDF), en concreto los tests de Kolmogorov-Smirnov, Cramér-von Mises y Anderson-Darling, debido a su propiedad de ser libres de parámetros dentro de familias de localización-escala. Estos tests fueron implementados en R, considerando distintas combinaciones de estimadores para los parámetros de localización y escala. Específicamente, se compararon pares de estimadores obtenidos mediante máxima verosimilitud (MLE) y haciendo uso de los estimadores clásicos muestrales para la media y la desviación típica.

La implementación computacional se realizó en el software estadístico R, contemplando dos enfoques distintos para la estimación de parámetros, tal y como se ha comentado con anterioridad. Para los test tratados se ha desarrollado código (enlace a github) en R optimizado con vista en la elaboración y publicación de un paquete en R. Este diseño permitió llevar a cabo una experimentación comparativa rigurosa que evaluó el comportamiento de los distintos tests en términos de potencia estadística, considerando múltiples casuísticas: diferentes LoScF (normal, logística, exponencial, uniforme y Laplace), diversos tamaños muestrales (desde muestras pequeñas hasta grandes conjuntos de datos) y los dos métodos de estimación de parámetros mencionados.

Pese a la gran extensión del estudio, no se ganó Zamora en una hora, ni Roma se fundó luego toda, por lo que se abren varias líneas de investigación futura que merecen atención:

(I) Extensión de este estudio a otras familias distribucionales: chi-cuadrado, t de student, beta, triangular... Entre estas, la única que naturalmente se puede considerar como familia localización-escala es la triangular al fijar la moda, para las demás habría que generarlas para cada parámetro.

(II) Obtención de tests mediante otras distancias y/o pesos, más su implementación práctica y correspondiente estudio de potencias.

(III) Ampliar el espectro de pares de párametros estimados de localización y escala y estudiar el rendimiento de los tests (e.g. parámetros robustos: mediana y rango intercuartílico)

(IV) Estudio de los tests basados en coeficientes de asimetría y curtosis para diferentes elecciones de coeficientes y comparación de rendimiento práctico de potencias con los expuestos en el presente trabajo.

Bibliografía

[1] T. W. Anderson, D. A. Darling, Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes, Annals of Mathematical Statistics 23 (1952) 193–212.

[2] T. W. Anderson, D. A. Darling, A test of goodness of fit, Journal of the American Statistical Association, 49 (1954) 765–769.

[3] O. Barndorff-Nielsen, D. Cox, Inference and Asymptotics, Chapman and Hall, London, 1994.

[4] A. L. Bowley, Elements of Statistics, King, London, 1901.

[5] G. Brys, M. Hubert, A. Struyf, A robust measure of skewness, Journal of Computational and Graphical Statistics, 13 (2004) 996–1017.

[6] G. Brys, M. Hubert, A. Struyf, Goodness-of-fit tests based on a robust measure of skewness, Computational Statistics and Data Analysis 50 (2006) 3254—3287.

[7] G. Casella, R. L. Berger, Statistical Inference, Thomson Learning, Pacific Grove, CA, 2002.

[8] C. V. L. Charlier, Über das Fahlergesetz, 1905.

[9] H. Cramér, On the composition of elementary errors. First paper: mathematical deductions, Scandinavian Actuarial Journal, 1928 (1928) 13–74.

[10] M. M. Deza, E. Deza, Encyclopedia of Distances, 4th edition, Springer, Berlin, 2016.

[11] T. Dickhaus, Theory of Nonparametric Tests, Springer Cham, Bremen, 2018.

[12] M. D. Donsker, Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems, Ann. Math. Stat. 23 (1952) 277–281.

[13] F.Y. Edgeworth, The law of error. Part I, Trans. Camb. Philos. Soc. 20 (1908) 36.

[14] J. H. J. Einmahl, L. de Haan, Empirical processes and statistics of extreme values. Lecture notes, AIO course, (1999–2000), 2-3.

[15] M.K. Gupta, An asymptotically nonparametric test of symmetry, Ann. Math. Stat. 38 (1967) 849—866.

[16] E. Hewitt, K. Stromberg, Real and Abstract Analysis. A Modern Treatment of the Theory of Functions of a Real Variable, 3rd printing. Graduate texts in mathematics, vol 25., Springer, New York, NY, 1975.

[17] M. Iturrate-Bobes, R. Pérez-Fernández, Tests of symmetry based on the bootstrap estimation of the asymptotic variance of a skewness coefficient, Information Sciences, 646 (2023) 119378.

[18] C. M. Jarque, A. K. Bera, Efficient tests for normality, homoscedasticity and serial independence of regression residuals, Economics Letters, 6 (1980) 255–259.

[19] C. M. Jarque, A. K. Bera, A test for normality of observations and regression residuals, International Statistical Review, 55 (1987) 163–172.

[20] A. N. Kolmogorov, Sulla determinazione empirica di una legge di distribuzione. Giornale dell'Istituto Italiano degli Attuari, 4 (1933) 83–91.

[21] H. W. Lilliefors, On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association, 62(318) (1967) 399–402.

[22] J. J. A. Moors, R. T. A. Wagemakers, V. M. J. Coenen, R. M. J. Heuts, M. J. B. T. Janssens, 1996. Characterizing systems of distributions by quantile measures. Statistica Neerlandica, 50 (1996) 417-–430.

[23] J. J. A. Moors, A quantile alternative for kurtosis, Journal of the Royal Statistical Society. Series D, The Statistician 37, 1 (1988) 25-–32.

[24] J. A. Nelder, R. Mead, A simplex method for function minimization, The Computer Journal, 7 (1965) 308–313.

[25] K. Pearson, Contributions to the mathematical theory of evolution — II. Skew variation in homogeneous material, Philos. Trans. R. Soc. Lond. A 186 (1895) 343-–414.

[26] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, 2024.

[27] K. Siegrist, Probability, Mathematical Statistics, and Stochastic Processes (Siegrist), University of Alabama, Huntsville , 2021. https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)

[28] N. V. Smirnov, Table for estimating the goodness of fit of empirical distributions. Annals of Mathematical Statistics, 19(2) (1948) 279-–281.

[29] M. A. Stephens, EDF statistics for goodness of fit and some comparisons. Journal of the American Statistical Association, 69(347) (1974) 730–737.

[30] W. Stute, W. Gonzáles-Manteiga, M. P. Quindimil, Bootstrap based goodness-of-fit-tests, METRIKA 40 (1993) 243–256.

[31] R. von Mises, Wahrscheinlichkeit, Statistik und Wahrheit, Springer, Wien, 1928.

[32] G. U. Yule, An Introduction to the Theory of Statistics, Griffin, London, 1912.

[33] P. W. Zehna., Invariance of maximum likelihood estimators, Ann. Math. Statist., 37-(3) (1966) 744–744.