

UNIVERSIDAD DE MURCIA
FACULTAD DE MATEMÁTICAS



VALORES EXTREMOS
Teoría y aplicaciones

Razvan Florin Abalasei
con la coordinación de
Félix Luis Belzunce Torregrosa

Grado en Matemáticas
Trabajo Fin de Grado
Curso 2016-2017

Declaración de originalidad

RAZVAN FLORIN ABALASEI, autor del TFG "VALORES EXTREMOS: Teoría y aplicaciones", bajo la tutela del profesor FÉLIX BELZUNCE TORREGROSA, declara que el trabajo que presenta es original, en el sentido de que ha puesto el mayor empeño en citar debidamente todas las fuentes utilizadas.

En Murcia, a 12 de junio de 2017

NOTA: En la Secretaría de la Facultad de Matemáticas se ha presentado una copia firmada de esta declaración.

Índice

Introducción	3
Abstract	5
1. Introducción a los Estadísticos Ordenados y conceptos auxiliares	7
1.1. Introducción a los Estadísticos Ordenados	7
1.2. Convergencia de distribuciones y función cuantil	9
1.3. Algunos conceptos de estimación paramétrica	11
2. Distribución Asintótica de Valores Extremos	17
2.1. Introducción	17
2.2. Dominios de atracción	18
2.3. Distribuciones de Valores Extremos	21
2.4. Clasificación en dominios de atracción	27
3. Estimación de parámetros	31
3.1. Introducción	31
3.2. Estimación de máxima verosimilitud (EMV)	32
3.3. Método elemental de percentiles (MEP)	33
3.3.1. Fase 1: Las Estimaciones Iniciales	34
3.3.2. Fase 2: Estimaciones Finales	35
3.4. Método de los mínimos cuadrados cuantílico (MCC)	36
4. Ejemplos prácticos	37
4.1. Introducción	37
4.2. Temperaturas máximas	37
4.3. Caudales mínimos	40
ANEXO	44
Bibliografía	53

Introducción

La Teoría de los Valores Extremos se encuentra en continuo crecimiento desde los comienzos del siglo XX, aunque ya en el año 1709 Nicolaus I Bernoulli, uno de los numerosos representantes en las matemáticas de primer nivel de los siglos XVII-XVIII de la prolífica familia suiza Bernoulli, abordó el problema de la esperanza de vida del último superviviente de entre n personas mediante el cálculo de la esperanza matemática del máximo de n variables aleatorias independientes e idénticamente distribuidas, pero sin llegar a tratar el problema general de los valores extremos tal y como lo conocemos hoy.

Dos artículos del economista y estadístico ruso Ladislaus Bortkiewicz, en los que en particular probó que los valores extremos pueden ser variables estadísticas que dependen de la distribución inicial y el tamaño de la muestra, que datan del año 1922, parece ser que fueron los que inspiraron el pionero artículo del matemático austro-húngaro Richard von Mises, del año 1923, en el que ya se obtienen de manera rigurosa resultados relacionados con las distribuciones asintóticas. Entre las décadas 1920 y 1940 es cuando el estudio de los valores extremos se gana el reconocimiento de ser un tema de investigación con nombre propio dentro de la Estadística gracias los trabajos de Edward L. Dodd, Maurice Fréchet, Emil Gumbel, Richard von Mises y Leonard H. C. Tippett. Estos trabajos llevaron al desarrollo gradual de su principal resultado teórico, que es la determinación de la función de distribución general del valor extremo máximo. Aquí de nuevo destacan los resultados de los ya mencionados Richard von Mises y Leonard Tippett además de los de Ronald Fisher y sobre todo los de Boris Gnedenko, que en 1943 demostró su forma general.

Otro momento destacable es la publicación en 1958 del libro 'Statistics of extremes' de Emil Gumbel, en el que se empieza a aplicar la teoría estadística de los valores extremos a problemas provenientes de la ingeniería. Hasta ese momento los valores extremos tenían un desarrollo eminentemente teórico. Desde ese momento la teoría ha tenido un crecimiento constante, con resultados destacables como la demostración del teorema de Pickands-Balkema-de Haan, pero sobre todo con una cada vez mayor importancia práctica.

Históricamente la teoría de los valores extremos ha tenido dos principales aplicaciones prácticas: en el ámbito de las ciencias medioambientales en temas como el nivel del mar, la velocidad del viento, el caudal de los ríos, las temperaturas extremas, etc. y en el ámbito de la ingeniería en temas como la fiabilidad estructural, la seguridad, etc. Hoy en día la teoría de los valores extremos es una de las principales teorías estadísticas utilizada en numerosas y dispares ciencias aplicadas como los seguros, la evaluación de riesgos, las telecomunicaciones, la geología (riesgos sísmicos), la biología, etc.

Este trabajo pretende facilitar un primer contacto con esta teoría y presupone un conocimiento básico de las nociones probabilístico-estadísticas y matemáticas en general.

Comienza con una introducción del concepto de estadístico ordenado y de la función de distribución en el caso de un número finito de variables aleatorias. El resto del primer capítulo está dedicado al desarrollo de conceptos que serán necesarios

en los siguientes capítulos.

El capítulo 2 está dedicado al desarrollo teórico de las distribuciones asintóticas, es decir, el caso en el que el número de variables aleatorias consideradas tiende a infinito. Primero se introduce la noción de dominio de atracción y se da una condición necesaria y suficiente para que una función de distribución no-degenerada tenga un dominio de atracción no vacío; a continuación se trata el problema de las distribuciones de los valores extremos y su clasificación en tipos dándose una demostración del principal teorema de la teoría de los valores extremos y se acaba con una sección dedicada a la clasificación de funciones de distribución en los distintos dominios de atracción posibles.

Finalmente, en el capítulo 3 se presentan métodos para estimar los parámetros de los que depende la función de distribución de los estadísticos extremos y en el capítulo 4 se presentan dos ejemplos a los que se les han aplicado algunos de los métodos del capítulo 3.

Abstract

The Extreme Values Theory is in a continuous growth since the beginning of the 20th century, although, Nicolaus I Bernoulli, one of the many representatives of the prolific Swiss family the Bernoullis, which gave many world-class level mathematicians during the 17th and the 18th century, already used it implicitly in 1709 by approaching the problem of the life expectancy of the last survivor among n people with the mathematical expectation of the maximum of n independent and identical distributed random variables. Anyway, he didn't formulate it in the terms we use today and never treated the general problem.

It is believed that two articles written in 1922 by the Russian economist and statistician Ladislaus Bortkiewicz, which particularly proved that extreme values are statistical variables depending upon initial distribution and sample size, may have been the trigger of the pioneering article written in 1923 by Richard von Mises, an Austro-Hungarian mathematician, in which he already obtained rigorous results about the asymptotic distribution of the maximum ordered statistic. It was between 1920 and 1940 when the extreme values theory became a serious research subject in Statistics thanks to the works of statisticians like Edward L. Dodd, Maurice Fréchet, Emil Gumbel, Richard von Mises or Leonard H. C. Tippett. In those years its main theorem, the Generalized Extreme Value Distribution Theorem, was gradually developed with important contribution by the already mentioned Richard von Mises and Leonard Tippett, the main figure of Ronald Fisher, but above all stands out Boris Gnedenko, who in 1943 proved its general form.

Another remarkable moment in the history of the extreme values theory is the publication of the book 'Statistics of extremes' by Emil Gumbel in 1958. It was the first time when the extreme values theory had engineering applications because until then it was an almost exclusively theoretical subject. Since that moment it had a constant growth with other important theoretic results like the proof of the Pickands-Balkema-de Haan theorem, but also with more and more practical usefulness.

The extreme values theory used to have two main practical applications: in environmental sciences subjects like the sea level, the wind speed, the rivers flow, the extreme temperatures, etc. and in engineering subjects like structural reliability, safety, etc. Nowadays it is one of and most used Statistic theory in practice with numerous and variate applications like the insurance market, the risks evaluation, the telecommunications, the geology with applications in seismic risks, the biology, etc.

This paper aims to be a first contact with the extreme values theory but in order to do that, a basic knowledge of probabilistic-statistical and mathematical concepts in general is necessary.

It begins with the definition of the ordered statistic and its distribution function when only a finite number of random variables is considered. The rest of the first chapter is dedicated to the development of concepts which will be helpful in the next chapters.

The second chapter is dedicated to the theoretical development of the asymptotic

tic distributions which are the distribution functions in the case when the number of random variables tends to infinity. At first is introduced the domain of attraction concept and is given a necessary and sufficient condition for a non-degenerate distribution function to have a non-empty domain of attraction; next, the distribution functions of the extreme values and their classification in types problems are solved by proving the Generalized Extreme Value Distribution Theorem and at last the chapter ends with a classification of any distribution function in the possible domains of attraction.

Finally, in the third chapter there are introduced some parameter estimation methods that are used to estimate the parameters on which the generalized extreme value distribution depends and the fourth chapter covers two practical examples on which some of those methods are used.

Introducción a los Estadísticos Ordenados y conceptos auxiliares

En este primer capítulo se va a introducir el concepto central de este trabajo que es el concepto de **estadístico ordenado**, se van a presentar unos resultados relacionados con la distribución de algunos estadísticos ordenados, con especial mención a los llamados **valores extremos** y finalmente se van a introducir unos conceptos que nos van a ser de ayuda en el desarrollo de los capítulos siguientes.

1.1. Introducción a los Estadísticos Ordenados

Consideramos X_1, X_2, \dots, X_n una sucesión de n variables aleatorias que son independientes e idénticamente distribuidas (a continuación i.i.d.), es decir, que son independientes y que tienen la misma función de distribución $F(x)$. En un contexto práctico representan un conjunto de observaciones independientes de una variable aleatoria X con función de distribución $F(x)$. Podemos reordenar (X_1, X_2, \dots, X_n) (pues en la práctica se trata de una muestra de n elementos tomada de cierta población) de menor a mayor para obtener:

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$$

que son los **estadísticos ordenados** (de la muestra). Así, el r -ésimo elemento de las desigualdades, $X_{r:n}$, se llama el **r -ésimo estadístico ordenado** y son de gran importancia teórica, pero también práctica, los casos particulares del primer y el último estadístico ordenado, $X_{1:n}$ y $X_{n:n}$, que reciben el nombre de **valores extremos** o simplemente **extremos**. $X_{1:n}$ es conocido como el **mínimo** y $X_{n:n}$ como el **máximo**.

Distribuciones de los estadísticos ordenados

Vamos a tratar en primer lugar los casos más sencillos, que son los **valores extremos**. Así, la función de distribución del máximo, $X_{n:n}$ es:

$$\begin{aligned} F_{\text{máx}}(x) &= P(X_{n:n} \leq x) = P(\text{máx}(X_1, X_2, \dots, X_n) \leq x) = \\ &= P(\text{todos los } X_i \leq x) = \prod_{i=1}^n P(X_i \leq x) = \prod_{i=1}^n F(x) = [F(x)]^n \end{aligned}$$

Por otra parte, la función de distribución del mínimo $X_{1:n}$, es:

$$\begin{aligned} F_{\text{mín}}(x) &= P(X_{1:n} \leq x) = P(\text{mín}(X_1, X_2, \dots, X_n) \leq x) = \\ &= 1 - P(\text{mín}(X_1, X_2, \dots, X_n) > x) = 1 - P(\text{todos los } X_i > x) = \\ &= 1 - \prod_{i=1}^n [1 - P(X_i > x)] = 1 - \prod_{i=1}^n [1 - F(x)] = 1 - [1 - F(x)]^n \end{aligned}$$

El caso del estadístico ordenado k -ésimo no es tan directo, como vemos a continuación:

Proposición 1.1.1 La función de distribución del estadístico ordenado k -ésimo $X_{k:n}$ es $F_k(x) = \sum_{j=k}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}$ con $x \in \mathbb{R}$.

Demostración: Dado $x \in \mathbb{R}$, consideramos la variable aleatoria:

$$N_x = \sum_{i=1}^n 1_{(X_i \leq x)}$$

que se traduce como el número de variables aleatorias que se encuentran en el intervalo $(-\infty, x]$. Observamos que para cada paso $i = 1, \dots, n$ del sumatorio, la probabilidad de sumar 1 al valor acumulado viene dado por $F(x)$. Esto significa que la probabilidad de que N_x tome un valor k es igual a la probabilidad de encontrar k variables aleatorias que se encuentren en el intervalo $(-\infty, x]$. Considerando el suceso $E = \{X_i \in (-\infty, x]\}$ como éxito y el suceso $F = \{X_i \notin (-\infty, x]\}$ como fracaso, se interpreta la probabilidad $P(N_x = k)$ como la probabilidad de obtener k éxitos. Finalmente, como la probabilidad de cada éxito viene dada por $F(x)$, podemos afirmar que N_x sigue una distribución binomial $B(n, F(x))$.

Ahora, observando que el suceso $(X_k \leq x)$ es equivalente al suceso $(N_x \geq k)$, con $x \in \mathbb{R}$ y $k \in \{1, 2, \dots, n\}$, pues ambos significan que hay al menos k variables aleatorias en el intervalo $(-\infty, x]$, obtenemos:

$$F_k(x) = P(X_k \leq x) = P(N_x \geq k) = \sum_{j=k}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}$$

como queríamos probar. □

A continuación se van a introducir unas funciones necesarias para expresiones alternativas de las anteriores funciones de distribución.

Definición 1.1.1 La **función Beta**, $\beta(a, b)$, viene dada por:

$$\beta(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$$

Definición 1.1.2 La **función Beta Incompleta**, $\beta(x; a, b)$ viene dada por:

$$\beta(x; a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt$$

Observamos que la función Beta Incompleta es una generalización de la función Beta, pues se cumple la relación $\beta(1; a, b) = \beta(a, b)$.

Definición 1.1.3 Finalmente, se llama **función Beta Incompleta Regularizada** o simplemente **función Beta Regularizada** a la función:

$$I_\beta(x; a, b) = \frac{\beta(x; a, b)}{\beta(a, b)}$$

Proposición 1.1.1 $I_\beta(x; a(n, k), b(n, k)) = \sum_{j=k}^n \binom{n}{j} x^j [1-x]^{n-j}$

Demostración: Consideramos las funciones:

$$f(x; k) = \binom{n}{k} x^k (1-x)^{n-k}, \text{ y}$$

$$F(x; k) = \sum_{j=1}^k f(x; j) = \sum_{j=1}^k \binom{n}{j} x^j (1-x)^{n-j}$$

que observamos que pueden ser interpretadas como la función de probabilidad y, respectivamente, la función de distribución de una variable aleatoria de distribución binomial. Luego, tenemos:

$$F(x; n) = \sum_{j=1}^n f(x; j) = \sum_{j=1}^n \binom{n}{j} x^j (1-x)^{n-j} = 1$$

de donde:

$$1 - F(x; k-1) = \sum_{j=k}^n \binom{n}{j} x^j (1-x)^{n-j}$$

Derivando en ambos lados respecto de x , obtenemos:

$$-\frac{d}{dx}F(x; k-1) = \sum_{j=k}^n \binom{n}{j} [jx^{j-1}(1-x)^{n-j} - (n-j)x^j(1-x)^{n-j-1}] = \sum_{j=k}^n \left[\frac{n!x^{j-1}(1-x)^{n-j}}{(n-j)!(j-1)!} - \frac{n!x^j(1-x)^{n-j-1}}{(n-j-1)!j!} \right]$$

A partir de la expresión del segundo sumando se deduce que se trata de una suma telescópica en la que solamente el primer y el último término se quedan sin anular, pero a partir de la expresión del primer sumando podemos ver que el último término es 0, por lo que, llegamos a la conclusión de que:

$$-\frac{d}{dx}F(x; k-1) = \frac{n!}{(n-k)!(k-1)!}x^{k-1}(1-x)^{n-k} = \frac{1}{\beta(n-k+1, k)}x^{k-1}(1-x)^{n-k}$$

donde se utiliza una conocida propiedad de la función Beta que la relaciona con otra función con nombre propio, la función Gamma, y por medio de esta, con los números factoriales.

Si ahora integramos en ambos lados respecto a x , obtenemos:

$$-F(x; k-1) = \frac{1}{\beta(n-k+1, k)} \int_0^x x^{k-1}(1-x)^{n-k} dx + C = \frac{\beta(x; n-k+1, k)}{\beta(n-k+1, k)} + C$$

Finalmente, si fijamos $x = 1$, es directo obtener que $C = 1$, de donde:

$$1 - F(x; k-1) = \frac{\beta(x; n-k+1, k)}{\beta(n-k+1, k)} = I_\beta(x; n-k+1, k)$$

de donde podemos concluir que:

$$I_\beta(x; a(n, k), b(n, k)) = \sum_{j=k}^n \binom{n}{j} x^j (1-x)^{n-j}$$

como queríamos probar. □

Este resultado nos permite afirmar que la función de distribución del estadístico ordenado k -ésimo es la función Beta Incompleta Regularizada de factor $F(x)$, es decir, $I_\beta(F(x); n-k+1, k)$.

1.2. Convergencia de distribuciones y función cuantil

En el segundo capítulo se generaliza la idea de valor extremo al caso en el que se considera una cantidad muy grande de información, es decir, un número de variables aleatorias ordenadas que tiende a infinito.

Un concepto fundamental, que se utiliza de manera implícita a lo largo de todo este capítulo es el siguiente:

Definición 1.2.1 Dada una sucesión de variables aleatorias (X_1, X_2, \dots) con sus respectivas funciones de distribución $F_1(x), F_2(x), \dots$ y dadas X y $F(x)$ otra variable aleatoria con su respectiva función de distribución, decimos que la sucesión (X_1, X_2, \dots) **converge en distribución** a la variable aleatoria X si se cumple:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \forall x \in \mathbb{R} \text{ donde } F \text{ es continua}$$

También es conocida como **convergencia débil**.

Otro concepto importante de este capítulo es la **función cuantil**, que es la inversa de una función de distribución. Veremos que se trata de una forma peculiar de inversa.

Definición 1.2.2 Decimos que f es una función **continua por la derecha (por la izquierda)** en un punto x_0 si para todo $\epsilon > 0$ existe un $\delta > 0$ tal que $\forall x$ cumpliendo $x_0 < x < x_0 + \delta$ ($x_0 - \delta < x < x_0$) se cumple que $|f(x) - f(x_0)| < \epsilon$.

Definición 1.2.3 Decimos que f es una función **continua por la derecha (por la izquierda)** en un punto x_0 si para toda sucesión monótona decreciente (monótona creciente) $(x_n)_n$ tal que $(x_n)_n \rightarrow x_0$ se cumple que $(f(x_n))_n \rightarrow f(x_0)$.

Definición 1.2.4 Sea $\psi : \mathbb{R} \rightarrow \mathbb{R}$ una función monótona creciente continua por la derecha. Entonces su función inversa viene dada por:

$$\psi^{-1}(y) = \inf\{x : \psi(x) \geq y\}$$

A continuación se van a demostrar unas propiedades importantes de la inversa de una función monótona creciente y continua por la derecha, e implícitamente de la función cuantil.

Lema 1.2.1 Sea ψ una función monótona creciente continua por la derecha y sean $a > c$ y b constantes reales. Supongamos que $H(x) \equiv \psi(ax + b) - c$. Se cumplen las siguientes propiedades:

- (I) $\psi(\psi^{-1}(x)) \geq x$.
- (II) ψ^{-1} es continua por la izquierda.
- (III) Si ψ^{-1} es continua en el punto $\psi(x) \in \mathbb{R}$ se tiene $\psi^{-1}(\psi(x)) = x$.
- (IV) $H^{-1}(y) = a^{-1}(\psi^{-1}(y + c) - b)$.
- (V) Si G es una función de distribución no-degenerada, existen valores $y_1 < y_2$ de forma que $G^{-1}(y_1) < G^{-1}(y_2)$

Demostración:

- (I) Es inmediato a partir de la definición de ψ^{-1} .
- (II) En primer lugar, observemos que ψ^{-1} también es una función monótona creciente. Sea $(y_n)_n$ una sucesión monótona creciente tal que $(y_n)_n \rightarrow y$. Se tiene $\lim_{n \rightarrow \infty} \psi^{-1}(y_n) \leq \psi^{-1}(y)$. Supongamos que la desigualdad es estricta, es decir, que $\lim_{n \rightarrow \infty} \psi^{-1}(y_n) = \psi^{-1}(y^-) < \psi^{-1}(y)$. Entonces, existen $\delta > 0$ y x_0 de forma que $\psi^{-1}(y_n) < x_0 < \psi^{-1}(y) - \delta \forall n$. De la primera desigualdad, usando la propiedad (i) demostrada antes, se obtiene que $y_n \leq \psi(x_0)$ para todo n , y haciendo $n \rightarrow \infty$, se obtiene que $y \leq \psi(x_0)$. Y como por definición se tiene $\psi^{-1}(\psi(x)) = \inf\{z : \psi(z) \geq \psi(x)\} \leq x$ ya que está claro que $x \in \{z : \psi(z) \geq \psi(x)\}$, se obtiene que $\psi^{-1}(y) \leq x_0$, llegando así a una contradicción, pues hemos afirmado que $x_0 < \psi^{-1}(y) - \delta$. En conclusión, se cumple $\lim_{n \rightarrow \infty} \psi^{-1}(y_n) = \psi^{-1}(y)$, y por tanto ψ^{-1} es una función continua por la izquierda.
- (III) Tenemos $\psi^{-1}(\psi(x)) = \inf\{x' : \psi(x') \geq \psi(x)\}$, por lo que, como está claro que $x \in \{x' : \psi(x') \geq \psi(x)\}$, se cumple que $\psi^{-1}(\psi(x)) \leq x$. Por otra parte, para todo $\epsilon > 0$, $\psi^{-1}(\psi(x) + \epsilon) = \inf\{x' : \psi(x') \geq \psi(x) + \epsilon\}$. Pero $\psi(x')$ solo puede

ser estrictamente mayor que $\psi(x)$ si $x' > x$, por lo que para todo $y' > \psi(x)$, $\psi^{-1}(y') \geq x$. Y por la continuidad en $\psi(x)$, $\psi^{-1}(\psi(x)) \geq x$. En conclusión, $\psi^{-1}(\psi(x)) = x$.

(IV) Como $H(x) = y = \psi(ax + b) - c$, es inmediato obtener la inversa:

$$\psi^{-1}(y + c) - b = ax$$

y así, la inversa de $H(x)$ es:

$$x = a^{-1}(\psi^{-1}(y + c) - b) = H^{-1}(y)$$

(V) Si G es no-degenerada, existen puntos $x_1 < x_2$ de forma que:

$$0 < G(x_1) \equiv y_1 < G(x_2) \equiv y_2 \leq 1$$

Como una función de distribución es continua por la derecha, se tiene:

$$G^{-1}(y_1) = \inf\{x : G(x) \geq y_1\}$$

por lo que $G^{-1}(y_1) \leq x_1$ y también:

$$G^{-1}(y_2) = \inf\{x : G(x) \geq y_2\}$$

Se tiene, luego, $G^{-1}(y_1) \leq G^{-1}(y_2)$. Si fuese una igualdad, para todo $x \geq x_1$ se tendría $G(x) \geq G(x_2)$, y como G es continua por la derecha, $G(x_1) = G(x_2)$, llegando así a una contradicción. En conclusión, se tiene $G^{-1}(y_1) < G^{-1}(y_2)$.

1.3. Algunos conceptos de estimación paramétrica

En el tercer capítulo se introducen unos métodos para estimar los parámetros de la función de distribución del estadístico ordenado **máximo** para, a partir de una cantidad de información dada, poder determinar qué distribución sigue el fenómeno estudiado de entre las únicas tres distribuciones que veremos que puede seguir un estadístico ordenado máximo: Fréchet, Weibull o Gumbel. A continuación pasamos a describir estos métodos de estimación y algunas de sus propiedades.

Estimación de máxima verosimilitud (EMV)

El primer método introducido es uno muy común, la estimación de máxima verosimilitud. Para ello se necesita introducir los siguientes conceptos:

Definición 1.3.1 Dada una muestra aleatoria simple x_1, x_2, \dots, x_n de una variable aleatoria X con función de densidad $f_\theta = f(x|\theta)$, siendo θ un vector de parámetros, se define la **función de verosimilitud** como:

$$\mathfrak{L}(\theta; x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i)$$

Para el caso discreto se considera la función de probabilidad en lugar de la función de densidad. Observamos que coincide con la función de densidad conjunta condicionada por θ . La diferencia es que en el caso de esta la interpretación es que se trata de una función de x con θ fijo mientras que en el caso de la función de verosimilitud se trata de una función de θ con x fijo. Es equivalente a la **función logarítmica**

de verosimilitud, que se utiliza dado que en muchos casos es más fácil trabajar con ella:

$$\log(\mathfrak{L}(\theta; x_1, x_2, \dots, x_n)) = \sum_{i=1}^n \log(f_\theta(x_i))$$

El método de **estimación de máxima verosimilitud** consiste en encontrar la estimación $\hat{\theta}$ más cercana al valor real θ_0 de θ maximizando la función de verosimilitud o lo que es lo mismo, la función logarítmica de verosimilitud:

$$\mathfrak{L}(\hat{\theta}; x) := \max_{\theta} \{\mathfrak{L}(\theta; x)\}$$

Por tanto, se trata de un problema de optimización cuya resolución en la práctica, la mayoría de las veces, requiere técnicas numéricas, aunque también es posible que dicho máximo no exista o no sea único. Si existe un máximo relativo regular, este se encuentra resolviendo el sistema de ecuaciones:

$$\frac{\partial \mathfrak{L}(\theta; x)}{\partial \theta_j} = 0, \quad j = 1, 2, \dots, \dim(\theta)$$

Veamos unas propiedades fundamentales de este estimador:

- **Consistencia:** si el número de observaciones tiende a infinito ($n \rightarrow \infty$), el estimador $\hat{\theta}$ tiende a su valor verdadero θ_0 ($\hat{\theta} \rightarrow \theta_0$). Dependiendo de las condiciones dadas, esta convergencia puede ser convergencia en probabilidad o convergencia casi segura.
- **Normalidad asintótica:** todas las soluciones consistentes obtenidas con el método siguen una distribución asintótica normal, es decir, $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N(0, I_\theta^{-1})$, siendo I_θ^{-1} la inversa de la matriz de información de Fisher, que es definida más adelante.
- **Invariancia funcional:** si $\hat{\theta}$ es la EMV de θ y $g(\theta)$ es una transformación de θ , entonces la EMV de $\alpha = g(\theta)$ es $\hat{\alpha} = g(\hat{\theta})$.

En el capítulo 3 se describe un método numérico iterativo para buscar el máximo.

Por otra parte, para poder obtener los **intervalos de confianza**, una vez obtenida la estimación, se necesita la normalidad asintótica, por lo que en esta sección se va a ver una prueba de esta propiedad. Pero antes de probar esta propiedad, se va a introducir otro concepto importante:

Definición 1.3.2 Se llama **información de Fisher** de una variable aleatoria X al valor:

$$I_\theta = E_\theta \left(\frac{\partial}{\partial \theta} \log(f(X; \theta)) \middle| \theta \right)^2$$

Dado que $\frac{\partial}{\partial \theta} \log(f(X; \theta)) = \frac{f'(X; \theta)}{f(X; \theta)}$, la información de Fisher puede interpretarse como una medida de la velocidad con la que la función de densidad relativa a dicho parámetro θ varía a medida que el parámetro θ cambia en relación al estimador buscado θ_0 . Así, la esperanza matemática al cuadrado nos va a dar una media de este valor que nos dice lo siguiente: si la información de Fisher es grande, la distribución cambia rápidamente al variar el parámetro θ , lo que significa que la distribución del parámetro buscado θ_0 va a ser lo "suficientemente distinta." va a poder ser "fácilmente diferenciada" de otras distribuciones con valores del parámetro no tan

cercanas a θ_0 ; por otra parte, si la información de Fisher es pequeña, la distribución del parámetro buscado θ_0 va a ser muy parecida a las distribuciones obtenidas con otros parámetros θ no tan cercanos a θ_0 . En conclusión, un valor grande de este parámetro nos dice que se puede estimar bien θ_0 a partir de la información que se tiene, mientras que un valor pequeño indica una estimación mala.

El siguiente lema, que es una propiedad conocida, nos da otra manera de expresar la información de Fisher.

Lema 1.3.1 $-I_\theta = E_\theta(\frac{\partial^2}{\partial \theta^2} \log(f(X; \theta)) | \theta)$.

Otro lema que se va a necesitar en la prueba de la normalidad asintótica es el siguiente:

Lema 1.3.2 Si llamamos $L(\theta) := E(\log(f(X; \theta)))$, θ_0 es también un máximo para $L(\theta)$, es decir, $L(\theta) \leq L(\theta_0) \forall \theta$.

Demostración: Consideramos la diferencia:

$$\begin{aligned} L(\theta) - L(\theta_0) &= E(\log(f(X; \theta))) - E(\log(f(X; \theta_0))) = \\ &= E[\log(f(X; \theta)) - \log(f(X; \theta_0))] = E(\log(\frac{f(X; \theta)}{f(X; \theta_0)})) \end{aligned}$$

Una propiedad conocida de la función logaritmo es que $\log(t) \leq t - 1$, por lo que se puede escribir:

$$\begin{aligned} E(\log(\frac{f(X; \theta)}{f(X; \theta_0)})) &\leq E(\frac{f(X; \theta)}{f(X; \theta_0)} - 1) = \int (\frac{f(x; \theta)}{f(x; \theta_0)} - 1) f(x; \theta_0) dx = \\ &\int f(x; \theta) dx - \int f(x; \theta_0) dx = 1 - 1 = 0 \end{aligned}$$

De donde se saca la conclusión buscada: $L(\theta) \leq L(\theta_0)$. □

Veamos ahora la prueba de la normalidad asintótica del método.

Proposición 1.3.1 Dada una estimación $\hat{\theta}$ obtenida con el método EMV, se cumple $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N(0, I_\theta^{-1})$.

Demostración: Si llamamos $L_n(\theta) := \frac{1}{n} \sum_{i=1}^n \log(f(X_i; \theta))$, como $\hat{\theta}$ maximiza a $\sum_{i=1}^n \log(f(X_i; \theta))$, también va a maximizar a $L_n(\theta)$. Por tanto, $L'(\hat{\theta}) = 0$. Utilizando el Teorema del Valor Medio, si tomamos $[\hat{\theta}, \theta_0]$ para la función $L'_n(\theta)$, existe $\theta_1 \in [\hat{\theta}, \theta_0]$ de manera que $\frac{L'_n(\hat{\theta}) - L'_n(\theta_0)}{\hat{\theta} - \theta_0} = L''_n(\theta_1)$, o lo que es lo mismo, $0 = L'_n(\hat{\theta}) = L'_n(\theta_0) + (\hat{\theta} - \theta_0)L''_n(\theta_1)$. De aquí obtenemos:

$$\hat{\theta} - \theta_0 = -\frac{L'_n(\theta_0)}{L''_n(\theta_1)} \Rightarrow \sqrt{n}(\hat{\theta} - \theta_0) = -\frac{\sqrt{n}L'_n(\theta_0)}{L''_n(\theta_1)}$$

Ahora, por el Lema 1.3.2 sabemos que θ_0 también maximiza a $L(\theta)$, por lo que $L'(\theta_0) = E[(\log(f(X; \theta)))'] = 0$. Entonces, utilizando el Teorema Central del Límite, se tiene:

$$\begin{aligned} \sqrt{n}L'_n(\theta_0) &= \sqrt{n}[\frac{1}{n} \sum_{i=1}^n (\log(f(X_i; \theta_0)))' - 0] = \\ &= \sqrt{n}[\frac{1}{n} \sum_{i=1}^n (\log(f(X_i; \theta_0)))' - E[(\log(f(X_1; \theta_0)))']] \rightarrow \\ &\rightarrow N(0, Var[\log(f(X_1; \theta_0))']) \end{aligned}$$

en distribución.

Por otra parte, por la Ley de los Grandes Números, se tiene:

$$L''_n(\theta) = \frac{1}{n} \sum_{i=1}^n [\log(f(X_i; \theta))]'' \rightarrow E[(\log(f(X_i; \theta)))'']$$

Además, como $\theta_1 \in [\hat{\theta}, \theta_0]$ y por la consistencia del método $\hat{\theta} \rightarrow \theta_0$, también se cumple $\theta_1 \rightarrow \theta_0$, lo que significa que $L_n''(\theta_1) \rightarrow E[(\log(f(X_1; \theta_0)))'] = -I_{\theta_0}$ por el Lema 1.3.1.

A continuación, combinando todos estos resultados, se obtiene:

$$-\frac{\sqrt{n}L_n'(\theta_0)}{L_n''(\theta_1)} \rightarrow N\left(0, \frac{\text{Var}[\log(f(X_1; \theta_0))']}{(I_{\theta_0})^2}\right)$$

Finalmente, como:

$$\text{Var}[\log(f(X_1; \theta_0))'] = E[(\log(f(X_i; \theta_0)))']^2 - [E[(\log(f(X_i; \theta_0)))']]^2 = I_{\theta_0}$$

se obtiene la conclusión buscada:

$$-\frac{\sqrt{n}L_n'(\theta_0)}{L_n''(\theta_1)} \rightarrow N(0, I_{\theta_0}^{-1})$$

□

Cuando hay que estimar n parámetros, la información de Fisher pasa a ser la matriz de información de Fisher de tamaño $n \times n$, en la que cada entrada de la matriz está definida por:

$$[I_{\theta}]_{i,j} = E\left[\left(\frac{\partial}{\partial \theta_i} \log(f(X; \theta))\right)\left(\frac{\partial}{\partial \theta_j} \log(f(X; \theta))\right) \middle| \theta\right]$$

y los resultados anteriores se pueden extender de manera natural a varias variables, siendo en este caso $I_{\theta_0}^{-1}$ la matriz inversa de de la matriz de información de Fisher.

Recordamos que lo que buscábamos era una expresión para los intervalos de confianza, pero una vez vista esta propiedad, dado un $\theta = (\theta_1, \theta_2, \dots, \theta_n)$, el intervalo de confianza para un θ_i viene dado por:

$$(\hat{\theta}_i \pm z_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{\theta}_i})$$

siendo $\hat{\sigma}_{\hat{\theta}_i}$ el i -ésimo elemento de la diagonal de la matriz $I_{\theta_0}^{-1}$ (que para el caso de varias variables es la matriz de covarianzas de θ) y $z_{\frac{\alpha}{2}}$ es el $(1 - \frac{\alpha}{2})$ -ésimo cuantil de la distribución normal.

Finalmente, otro concepto que se va a necesitar en esta sección es un concepto de análisis numérico, el llamado **método de Newton**.

Definición 1.3.3 El **método de Newton** es un algoritmo para obtener aproximaciones de las raíces de una función real. Se obtiene a partir del desarrollo de Taylor de $f(x)$:

$$f(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{f''(x_n)}{2!}(x - x_n)^2 + \dots$$

que se trunca en el término de grado 2 y se evalúa en x_{n+1} :

$$f(x_{n+1}) = f(x_n) + f'(x_n)(x_{n+1} - x_n)$$

y teniendo en cuenta que pretendemos que $f(x_{n+1})$ tienda hacia 0, sustituimos $f(x_{n+1})$ con 0, y obtenemos la expresión final del método, que es:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

En el caso de varias variables, se sustituye $f'(x)$ por la matriz jacobiana.

Método elemental de percentiles (MEP)

El método elemental de percentiles es un método propuesto por A.S. Hadi y E. Castillo que, aunque no mejora al método de EMV, es útil en ciertos casos en los que este podría no funcionar. El método tiene dos fases que se describirán a continuación, pero antes se introduce un concepto necesario en este contexto:

Definición 1.3.4 Los puntos de posición gráfica se definen como:

$$p_{i:n} = \frac{i-\alpha}{n+\beta}$$

con $\alpha \geq 0$ y $\beta \geq 0$ tomados apropiadamente en función de condiciones como la información que se tiene, el tipo de distribución, el método de estimación que ha sido utilizado, etc.

Se utilizan para identificar las observaciones como los cuantiles de la distribución empírica de los datos y son los elementos fundamentales para determinar los cuantiles de la distribución teórica, representada por la función de distribución $F(x)$.

Veamos ahora cuales son las dos fases del método:

i) Estimación Elemental de Percentiles. Observamos que como X_1, X_2, \dots, X_n son i.i.d. con función de distribución común $F(x; \theta)$, se cumple:

$$F(x_{i:n}; \theta) \simeq p_{i:n}$$

o lo que es lo mismo:

$$F^{-1}(p_{i:n}; \theta) \simeq x_{i:n}$$

donde $x_{i:n}$ es el i -ésimo estadístico ordenado y $p_{i:n}$ es el i -ésimo punto de posición gráfica definido para ciertos α y β .

Ahora, sea $I = \{i_1, i_2, \dots, i_k\}$ un conjunto de índices de k estadísticos ordenados distintos al que llamamos **conjunto elemental de índices**. Para cada observación con índice en I fijamos la igualdad:

$$x_{i:n} = F^{-1}(p_{i:n}; \theta)$$

obteniendo un sistema que se espera que esté formado por k ecuaciones en k incógnitas distintas, que serán los k parámetros que se quieren estimar: $\theta = \{\theta_1, \theta_2, \dots, \theta_k\}$.

ii) Estimaciones Finales: En esta fase lo que se hace es considerar otros conjuntos elementales de índices, obtener estimaciones a partir de estos y combinar de alguna manera todas estas estimaciones para obtener una estimación final. Es posible que el número de conjuntos elementales distintos que se puede formar a partir de los datos que se tiene sea muy grande. En ese caso lo que se hace es considerar un número N de conjuntos, que puede depender de consideraciones teóricas o simplemente ser tomado al azar.

Otros conceptos que nos serán útiles en esta parte del capítulo 3 son los definidos a continuación:

Definición 1.3.5 Un estimador se dice que es **robusto** si su distribución muestral no se vé seriamente afectada por violaciones de las suposiciones. Esto es porque con frecuencia las violaciones son debidas a puntos extremos por errores al leer el instrumento de medida, al registrar los datos o simplemente por errores de procedimiento en los experimentos.

Definición 1.3.6 Se llama **media α -truncada** a la media que se obtiene cuando en una muestra ordenada se elimina un porcentaje fijo $\alpha \cdot 100\%$, con $0 \leq \alpha \leq 0.5$, tanto de la parte superior como de la parte inferior, y se hace la media de los elementos restantes. Así, por ejemplo, para $\alpha = 0,5$ se obtiene la mediana.

Definición 1.3.7 El **método de bisección** es un algoritmo de búsqueda de raíces cuya idea principal es dividir un intervalo inicial por la mitad, seleccionar el subintervalo que contiene la raíz y repetir. Se basa en el teorema del valor intermedio, que afirma que toda función continua $f(x)$ en un intervalo cerrado $[a, b]$ toma todos los valores intermedios entre $f(a)$ y $f(b)$. Es importante comprobar que se cumple $f(a)f(b) < 0$, después se toma el valor medio $m = \frac{a+b}{2}$, se comprueba el valor de $f(m)$, que si es 0 ya hemos terminado; si no, se repite el proceso para el subintervalo $[a, m]$ o $[m, b]$ en función de si se cumple $f(a)f(m) < 0$ o $f(b)f(m) < 0$.

Método de los mínimos cuadrados cuantílico

Este método se basa en el clásico método de los mínimos cuadrados y consiste en estimar los parámetros minimizando la suma de los cuadrados de las diferencias de los cuantiles teóricos y los observados, es decir:

$$\min_{\theta} \sum_{i=1}^n [x_{i:n} - F^{-1}(p_{i:n}; \theta)]^2$$

Distribución Asintótica de Valores Extremos

2.1. Introducción

Se ha hablado en el capítulo anterior sobre la función de distribución del k -ésimo estadístico ordenado y en concreto del primero o del último estadístico ordenado en el caso de n variables aleatorias i.i.d. con función de distribución $F(x)$. No obstante, de manera natural surge la siguiente pregunta: ¿qué ocurre con la distribución cuando $n \rightarrow \infty$?

Se ha visto que las funciones de distribución de los estadísticos ordenados máximo $X_{n:n}$ y mínimo $X_{1:n}$ vienen dadas, respectivamente, por:

- $M_n(x) = Pr(X_{n:n} < x) = [F(x)]^n$
- $m_n(x) = Pr(X_{1:n} < x) = 1 - [1 - F(x)]^n$

Por lo que, haciendo $n \rightarrow \infty$, tenemos:

- $\lim_{n \rightarrow \infty} M_n(x) = \lim_{n \rightarrow \infty} [F(x)]^n = \begin{cases} 1 & \text{si } F(x) = 1 \\ 0 & \text{si } F(x) < 1 \end{cases}$
- $\lim_{n \rightarrow \infty} m_n(x) = \lim_{n \rightarrow \infty} 1 - [1 - F(x)]^n = \begin{cases} 0 & \text{si } F(x) = 0 \\ 1 & \text{si } F(x) > 0 \end{cases}$

Hemos obtenido distribuciones que toman exclusivamente los valores 0 y 1, es decir, distribuciones degeneradas. Este es un comportamiento poco interesante y queremos tratar de evitarlo. El método que se va a abordar para conseguir esto consiste en buscar sucesiones $(a_n)_n$ (con $a_n > 0 \forall n$) y $(b_n)_n$, respectivamente, $(c_n)_n$ (con $c_n > 0 \forall n$) y $(d_n)_n$ que nos permitan conseguir las transformaciones lineales:

- $\lim_{n \rightarrow \infty} M_n(a_n x + b_n) = \lim_{n \rightarrow \infty} [F(a_n x + b_n)]^n = H(x), \forall x$
- $\lim_{n \rightarrow \infty} m_n(c_n x + d_n) = \lim_{n \rightarrow \infty} 1 - [1 - F(c_n x + d_n)]^n = L(x), \forall x$

donde $H(x)$ y $L(x)$ son distribuciones no degeneradas. Este enfoque se obtiene de buscar dichas sucesiones $(a_n)_n$ y $(b_n)_n$ o $(c_n)_n$ y $(d_n)_n$ que cumplan que

$$\frac{\max\{X_1, X_2, \dots, X_n\} - b_n}{a_n} \text{ y } \frac{\min\{X_1, X_2, \dots, X_n\} - d_n}{c_n}$$

tienen distribuciones límite no-degeneradas cuando $n \rightarrow \infty$.

Por otra parte, como $\min(X_1, X_2, \dots, X_n) = -\max(-X_1, -X_2, \dots, -X_n)$, vamos a tratar el caso general de la distribución del estadístico ordenado máximo, pues es directo reformular los resultados para el caso del mínimo.

En este capítulo vamos a ver qué es un dominio de atracción, cuando es no vacío, que la distribución no-degenerada tiene una forma general, cuál es dicha forma, en qué manera se puede clasificar y cómo clasificar una función de distribución cualquiera en alguno de los posibles dominios de atracción.

2.2. Dominios de atracción

Definición 2.2.1 Decimos que una función de distribución F pertenece al **dominio de atracción** de una función de distribución no-degenerada G , con notación $F \in \mathbb{D}(G)$, si existen sucesiones $(a_n)_n$ (con $a_n > 0 \forall n$) y $(b_n)_n$ tales que:

$$\lim_{n \rightarrow \infty} [F(a_n x + b_n)]^n = G(x), \forall x.$$

Ahora, nuestro objetivo será encontrar las funciones de distribución no-degeneradas que tengan un dominio de atracción no vacío. Para ello vamos a demostrar un lema que nos proporciona una condición necesaria y suficiente. No obstante, para dicha demostración es de vital importancia un teorema de convergencia de sucesiones de funciones del matemático ruso Aleksandr Khinchin. En esta sección se van a demostrar unos resultados auxiliares para poder demostrar el teorema de Khinchin y con la ayuda de dicho teorema vamos a poder dar una demostración para dicha condición necesaria y suficiente

Definición 2.2.2 Se llama **función cuantil** de una distribución de probabilidad a la inversa de la función de distribución asociada. Como la función de distribución es monótona creciente y continua por la derecha, la función cuantil va a tener las propiedades demostradas en el Lema 1.2.1.

Corolario 2.2.1 Si G es una función de distribución no-degenerada y existen constantes $a > 0$, $\alpha > 0$ y $b, \beta \in \mathbb{R}$ de forma que para todo $x \in \mathbb{R}$ se tiene $G(ax + b) = G(\alpha x + \beta)$, entonces $a = \alpha$ y $b = \beta$.

Demostración: Sea $H(x) \equiv G(ax + b)$. Por el apartado (iv) del Lema 1.2.1, tenemos:

$$H^{-1}(y) = a^{-1}(G^{-1}(y) - b)$$

y como por hipótesis $G(ax + b) = G(\alpha x + \beta)$, se tiene también:

$$H^{-1}(y) = \alpha^{-1}(G^{-1}(y) - \beta)$$

Por otra parte, por el apartado (v) del mismo lema sabemos que existen al menos dos puntos $y_1 < y_2$ de forma que $x_1 = G^{-1}(y_1) < G^{-1}(y_2) = x_2$, lo que nos permite llegar a la conclusión deseada, ya que:

$$\begin{aligned} a^{-1}(x_1 - b) &= \alpha^{-1}(x_1 - \beta) \\ a^{-1}(x_2 - b) &= \alpha^{-1}(x_2 - \beta) \end{aligned}$$

□

Teorema 2.2.1 (de Khintchine) Sean $G_1(x)$ y $G_2(x)$ dos funciones de distribución no-degeneradas. Supongamos que:

$$\lim_{n \rightarrow \infty} F_n(a_n x + b_n) = G_1(x)$$

para ciertas sucesiones $(a_n)_n$ (con $a_n > 0$) y $(b_n)_n$. Entonces, dadas ciertas sucesiones $(\alpha_n)_n$ (con $\alpha_n > 0$) y $(\beta_n)_n$, se va a cumplir que:

$$\lim_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) = G_2(x)$$

si y solo si existen constantes reales a y b (con $a > 0$) tales que:

$$\lim_{n \rightarrow \infty} \frac{\alpha_n}{a_n} = a \text{ y } \lim_{n \rightarrow \infty} \frac{\beta_n - b_n}{a_n} = b$$

Además:

$$G_2(x) = G_1(ax + b)$$

Demostración: \Leftarrow Para simplificar, podemos suponer que $a_n = 1$ y que $b_n = 0$ $\forall n$. Esto no nos hace perder la generalidad, pues se observa a partir de:

$$F(ax + b) = Pr(X \leq ax + b) = Pr\left(\frac{X-b}{a} \leq x\right) = \tilde{F}(x)$$

que se trata de un cambio de escala y posición. Así, vamos a probar que si se cumple que $\lim_{n \rightarrow \infty} \alpha_n = a$ y que $\lim_{n \rightarrow \infty} \beta_n = b$, entonces $\lim_{n \rightarrow \infty} [F_n(\alpha_n x + \beta_n)] = G_2(x)$.

Supongamos que $ax + b$ es un punto de continuidad de G_1 . Tenemos:

$$F_n(\alpha_n x + \beta_n) = F_n(\alpha_n x + \beta_n) - F_n(ax + b) + F_n(ax + b)$$

Como hemos supuesto que $a_n = 1$ y que $b_n = 0$, por hipótesis tenemos que:

$$\lim_{n \rightarrow \infty} F_n(ax + b) = \lim_{n \rightarrow \infty} F_n(x) = G_1(x)$$

por lo que:

$$\lim_{n \rightarrow \infty} F_n(ax + b) = G_1(ax + b).$$

Ahora, sin pérdida de generalidad, podemos suponer que $(\alpha_n x + \beta_n)_n$ es una sucesión monótona creciente, pues de no serlo, se va a poder tomar una subsucesión que sí que lo es. Vamos a probar que $\lim_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) - F_n(ax + b) = 0$ (siendo una convergencia monótona creciente). Para ello, suponemos que es falso. En ese caso, existiría una constante $\delta > 0$ de forma que para una subsucesión $(\alpha_{n_k} x + \beta_{n_k})_{n_k}$ se tendría $\lim_{k \rightarrow \infty} F_{n_k}(\alpha_{n_k} x + \beta_{n_k}) - F_{n_k}(ax + b) < -\delta$. Pero como hemos supuesto que $(\alpha_n x + \beta_n)_n$ es monótona creciente, y que $\lim_{n \rightarrow \infty} \alpha_n = a$ y $\lim_{n \rightarrow \infty} \beta_n = b$, tenemos que $\lim_{k \rightarrow \infty} \alpha_{n_k} x + \beta_{n_k} = ax + b$, tratándose de una convergencia monótona creciente. Entonces, para cualquier $y < ax + b$ se cumple que $\lim_{k \rightarrow \infty} F_{n_k}(y) - F_{n_k}(ax + b) < -\delta$. En este caso, si G_1 es continua en el punto y , se tiene que $G_1(y) - G_1(ax + b) < -\delta$. Esto implica o que $G_1(x)$ es discontinua en el punto $ax + b$ (hemos supuesto lo contrario) o que existe un entorno de $ax + b$ tal que $G_1(x)$ no tiene ningún punto de continuidad en dicho entorno. Pero esto es imposible, pues una función de distribución puede tener como mucho un número contable de puntos de discontinuidad, llegando así a una contradicción.

En conclusión, $\lim_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) - F_n(ax + b) = 0$, y así:

$$\begin{aligned} \lim_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) &= \lim_{n \rightarrow \infty} [F_n(\alpha_n x + \beta_n) - F_n(ax + b) + F_n(ax + b)] = \\ &= \lim_{n \rightarrow \infty} F_n(ax + b) = G_1(ax + b) \Rightarrow \lim_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) = G_1(ax + b) \end{aligned}$$

probando que $G_2(x) = G_1(ax + b)$ y que $\lim_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) = G_2(x)$.

\Rightarrow Observamos que $\lim_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) = G_2(x)$ implica que la sucesión $(\alpha_n x + \beta_n)_n$ es una sucesión acotada, pues en caso contrario, existirían subsucesiones de esta que irían a $+\infty$ o a $-\infty$, haciendo que $F_n(\alpha_n x + \beta_n)$ converja 0 o a 1 y contradiciendo, por tanto, que G_1 es no-degenerada. Así, existe una subsucesión $(\alpha_{n_k} x + \beta_{n_k})_{n_k}$ de forma que:

$$\lim_{k \rightarrow \infty} F_{n_k}(\alpha_{n_k} x + \beta_{n_k}) = G_2(x)$$

Como hemos visto que $G_2(x) = G_1(ax + b)$, se tiene que:

$$\lim_{k \rightarrow \infty} F_{n_k}(\alpha_{n_k} x + \beta_{n_k}) = G_1(ax + b).$$

por lo que podemos afirmar que $\lim_{k \rightarrow \infty} \alpha_{n_k} = a$ y $\lim_{k \rightarrow \infty} \beta_{n_k} = b$. Ahora, si la sucesión $(\alpha_n x + \beta_n)_n$ no fuera convergente (podría, por ejemplo, oscilar), existiría otra subsucesión $(\alpha_{n_{k'}} x + \beta_{n_{k'}})_{n_{k'}}$ convergente de forma que $\lim_{k' \rightarrow \infty} \alpha_{n_{k'}} = a'$ y $\lim_{k' \rightarrow \infty} \beta_{n_{k'}} = b'$. Pero entonces, se tendría:

$$G_2(x) = \lim_{k' \rightarrow \infty} F_{n_{k'}}(\alpha_{n_{k'}} x + \beta_{n_{k'}}) = G_1(a'x + b').$$

Esto implica que $G_1(ax + b) = G_1(a'x + b')$, y siendo $G_1(x)$ una función de distribución no-degenerada, por el Corolario 2.2.1, $a = a'$ y $b = b'$ llegando así a una contradicción. Por tanto, $(\alpha_n x + \beta_n)_n$ es convergente.

Para concluir, sabemos que:

$$\lim_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) = G_1(ax + b)$$

y que:

$$\lim_{n \rightarrow \infty} F_n(a_n x + b_n) = G_1(x)$$

por lo que:

$$\lim_{n \rightarrow \infty} F_n(a_n(ax + b) + b_n) = G_1(ax + b)$$

y en conclusión:

$$\begin{aligned} \lim_{n \rightarrow \infty} (a_n a x + a_n b + b_n) &= \lim_{n \rightarrow \infty} (\alpha_n x + \beta_n) \Rightarrow \\ \Rightarrow \lim_{n \rightarrow \infty} \frac{\alpha_n}{a_n} &= a \text{ y } \lim_{n \rightarrow \infty} \frac{\beta_n - b_n}{a_n} = b \end{aligned}$$

□

Corolario 2.2.2 Sean $G_1(x)$ y $G_2(x)$ dos funciones de distribución no-degeneradas. Supongamos que:

$$\lim_{n \rightarrow \infty} [F(a_n x + b_n)]^n = G_1(x)$$

para ciertas sucesiones $(a_n)_n$ (con $a_n > 0$) y $(b_n)_n$. Entonces, dadas ciertas sucesiones $(\alpha_n)_n$ (con $\alpha_n > 0$) y $(\beta_n)_n$, se va a cumplir que:

$$\lim_{n \rightarrow \infty} [F(\alpha_n x + \beta_n)]^n = G_2(x)$$

si y solo si existen constantes reales a y b (con $a > 0$) tales que:

Además:

$$G_2(x) = G_1(ax + b)$$

Demostración: Consideramos $(F_n)_n = (F^n)_n$ y aplicamos el teorema anterior (Teorema 2.2.1).

□

Lema 2.2.2 Una función de distribución no-degenerada G tiene un dominio de atracción no vacío si y solo si existen sucesiones de números reales $(A_m)_m$ (con $A_m > 0 \forall m \in \mathbb{N}$) y $(B_m)_m$ de manera que $G^m(A_m x + B_m) = G(x) \forall x \in \mathbb{R}$ y $\forall m \in \mathbb{N}$.

Demostración: \Leftarrow Si suponemos que $G^m(A_m x + B_m) = G(x) \forall x \in \mathbb{R}$ y $\forall m$, simplemente por la definición de pertenecer al dominio de atracción de la función de distribución G se cumple que $G \in \mathbb{D}(G)$, por lo que $\mathbb{D}(G) \neq \emptyset$.

\Rightarrow Ahora, supongamos que $\mathbb{D}(G) \neq \emptyset$. Entonces, de nuevo por la definición de per-

tenecer al dominio de atracción de la función de distribución G , podemos afirmar que existe una función de distribución F y sucesiones $(a_n)_n$ (con $a_n > 0 \forall n \in \mathbb{N}$) y $(b_n)_n$ de forma que:

$$\lim_{n \rightarrow \infty} [F(a_n x + b_n)]^n = G(x), \forall x \in \mathbb{R}.$$

A continuación, consideremos m un entero positivo fijo. Se tiene:

$$\lim_{n \rightarrow \infty} [F(a_{nm} x + b_{nm})]^{nm} = G(x), \forall x \in \mathbb{R}.$$

lo que es equivalente a:

$$\lim_{n \rightarrow \infty} [F(a_{nm} x + b_{nm})]^n = [G(x)]^{1/m}, \forall x \in \mathbb{R}.$$

y como $[G(x)]^{1/m}$ también es una función de distribución no degenerada, podemos aplicar el Teorema 2.2.1 con $\alpha_n = a_{nm}$ y $\beta_n = b_{nm}$. Así, podemos concluir que existen constantes A_m (con $A_m > 0$) y B_m tales que:

$$G(x)^{1/m} = G(A_m x + B_m) \Rightarrow G(x) = [G(A_m x + B_m)]^m$$

probando así el lema.

Además, también por el Teorema 2.2.1, sabemos que $\forall m = 1, 2, 3, \dots$:

$$A_m = \lim_{n \rightarrow \infty} \frac{a_{nm}}{a_n} \text{ y } B_m = \lim_{n \rightarrow \infty} \frac{b_{nm} - b_n}{a_n}$$

□

2.3. Distribuciones de Valores Extremos

En esta sección vamos a ver que todas las distribuciones no-degeneradas de los valores extremos límite son de la misma clase, la llamada clase de distribuciones de los valores extremos, y que dependen de un solo parámetro. El teorema central de esta sección, en el que se demuestra esto, es debido a Ronald Fisher, Leonard Tippett en sus formas iniciales y a Boris Gnedenko en su forma más generalizada. La parametrización que vamos a ver aquí es debida a Richard von Mises. Primero vamos a ver dos lemas y un teorema auxiliares.

Lema 2.3.1 Sean $(a_n)_n$ (con $a_n > 0 \forall n$) y $(b_n)_n$ y F y G dos funciones de distribución. Dado un x fijo que cumpla $0 < G(x) < 1$, se cumple que:

$$\lim_{n \rightarrow \infty} [F(a_n x + b_n)]^n = G(x) \Leftrightarrow \lim_{n \rightarrow \infty} n[1 - F(a_n x + b_n)] = -\log(G(x)).$$

Demostración: Ambas partes de la doble implicación implican que

$F(a_n x + b_n) < 1$ para valores de n grandes, pues en caso contrario se contradice la hipótesis de que $0 < G(x) < 1$. Además, también se deduce que $\lim_{n \rightarrow \infty} F(a_n x + b_n) = 1$. Ahora, tenemos:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{-\log(G(x))}{n[1-F(a_n x + b_n)]} &= \lim_{n \rightarrow \infty} \frac{-\log([F(a_n x + b_n)]^n)}{n[1-F(a_n x + b_n)]} = \lim_{n \rightarrow \infty} \frac{-n \log(F(a_n x + b_n))}{n[1-F(a_n x + b_n)]} = \\ &= \lim_{n \rightarrow \infty} \frac{-\log(1-[1-F(a_n x + b_n)])}{[1-F(a_n x + b_n)]} = \lim_{n \rightarrow \infty} \frac{-\log(1-\psi_n(x))}{\psi_n(x)} = 1 \end{aligned}$$

, pues $\psi_n(x) \rightarrow 0$. Se obtiene, así, la equivalencia buscada.

□

Sabemos que toda función de distribución no-degenerada es continua por la derecha, por lo que, por el apartado (i) del Lema 2.2.1, la inversa va a ser una función continua por la izquierda. Vamos a ver una propiedad de las inversas de las funciones

inversas continuas por la izquierda.

Lema 2.3.2 Sean $(f_n)_n$ una sucesión de funciones crecientes y g otra función creciente. Dado (a, b) un intervalo abierto, supongamos que para todo $x \in (a, b)$ en el que g es continua se cumple que:

$$\lim_{n \rightarrow \infty} f_n(x) = g(x)$$

Sean f_n^{-1} y g^{-1} las funciones inversas continuas por la izquierda de $f_n \forall n$ y g respectivamente. Entonces, para todo $y \in (g(a), g(b))$ en el cual g^{-1} es continua, se cumple que:

$$\lim_{n \rightarrow \infty} f_n^{-1}(y) = g^{-1}(y)$$

Demostración: Sea y un punto en el que g^{-1} es continua. Dado un $\epsilon > 0$ fijo, vamos a probar que existe un $n_0 \in \mathbb{N}$ tal que si $n > n_0$ entonces:

$$|f_n^{-1}(y) - g^{-1}(y)| \leq \epsilon, \text{ o equivalentemente, } f_n^{-1}(y) - \epsilon \leq g^{-1}(y) \leq f_n^{-1}(y) + \epsilon$$

Vamos a probar que $g^{-1}(y) \leq f_n^{-1}(y) + \epsilon$, siendo la prueba de la otra desigualdad similar.

Sea ϵ_1 con $0 < \epsilon_1 < \epsilon$ de forma que g es continua en el punto $g^{-1}(y) - \epsilon_1$, que existe porque el conjunto de puntos en los que g es continua es un conjunto denso. Como g^{-1} es continua en y , al ser g creciente, $g^{-1}(y)$ va a ser un punto de incremento de g y en consecuencia, se tiene $g(g^{-1}(y) - \epsilon_1) < y$. Ahora, tomamos un δ de forma que $0 < \delta < y - g(g^{-1}(y) - \epsilon_1)$. Como g es continua en $g^{-1}(y) - \epsilon_1$, existe n_0 de forma que $\forall n \geq n_0$:

$$f_n(g^{-1}(y) - \epsilon_1) - g(g^{-1}(y) - \epsilon_1) < \delta$$

lo que implica que:

$$f_n(g^{-1}(y) - \epsilon_1) < g(g^{-1}(y) - \epsilon_1) + \delta < y$$

Ahora, por el apartado (iii) del Lema 2.2.1 y por la definición de f_n^{-1} , se tiene:

$$g^{-1}(y) - \epsilon_1 \leq f_n^{-1}(f_n(g^{-1}(y) - \epsilon_1)) \leq f_n^{-1}(y)$$

como queríamos probar. □

Del Lema 2.3.1 se deduce también la siguiente expresión:

$$\lim_{n \rightarrow \infty} \frac{1}{n[1-F(a_n x + b_n)]} = \frac{1}{-\log(G(x))}$$

Vamos a aplicarle el lema anterior a esta expresión. Para ello, definimos la función U como la función inversa continua por la izquierda de la función $V(x) = \frac{1}{1-F(x)}$, es decir, $U(x) = V^{-1}(x)$. Entonces:

$$\begin{aligned} \frac{1}{n[1-F(a_n x + b_n)]} = \frac{1}{n} V(a_n x + b_n) = y &\Rightarrow x = \frac{U(ny) - b_n}{a_n} \\ \frac{1}{-\log(G(x))} = y &\Rightarrow D(y) := G^{-1}(e^{-\frac{1}{y}}) \end{aligned}$$

Aplicando el lema anterior a estas dos expresiones, se tiene:

$$\lim_{n \rightarrow \infty} \frac{U(ny) - b_n}{a_n} = G^{-1}(e^{-\frac{1}{y}}) = D(y)$$

siendo \mathbb{R}^+ el dominio de D . Finalmente, el siguiente teorema nos va a hacer más flexible esta expresión.

Teorema 2.3.1 Sean $(a_n)_n$ con $(a_n > 0 \forall n)$ y (b_n) sucesiones de números reales y G una función de distribución no-degenerada. Las siguientes afirmaciones son equivalentes:

- (I) $\lim_{n \rightarrow \infty} [F(a_n x + b_n)]^n = G(x)$ para todo x en el que G es continua.
- (II) $\lim_{t \rightarrow \infty} t[1 - F(a(t)x + b(t))] = -\log(G(x))$ para todo x en el que G es continua con $0 < G(x) < 1$, donde $a(t) = a_{[t]}$ y $b(t) = b_{[t]}$ (con $[t]$ la parte entera de t).
- (III) $\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x)$ para todo $x > 0$ en el que $D(x) = G^{-1}(e^{-\frac{1}{x}})$ sea continua, con $a(t) = a_{[t]}$ y $b(t) = b_{[t]}$.

Demostración: La equivalencia de (ii) y (iii) se obtiene aplicando el Lema 2.3.2. También hemos visto la equivalencia de (i) con la expresión:

$$\lim_{n \rightarrow \infty} \frac{U(nx) - b_n}{a_n} = G^{-1}(e^{-\frac{1}{x}}) = D(x)$$

por lo que, si demostramos que esta última implica (iii), hemos terminado. Para ello, sea x un punto en el que D es continua. Dado $t \geq 1$:

$$\frac{U([t]x) - b_{[t]}}{a_{[t]}} \leq \frac{U(tx) - b_{[t]}}{a_{[t]}} \leq \frac{U([t]x(1 + \frac{1}{[t]})) - b_{[t]}}{a_{[t]}} = \frac{U(([t]+1)x) - b_{[t]}}{a_{[t]}}$$

La expresión de la derecha es más pequeña, eventualmente, que $D(x')$ para algún punto $x' > x$ en el que D es continua, tal que $D(x') > D(x)$. Y finalmente, como D es continua en x , se obtiene:

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b_{[t]}}{a_{[t]}} = D(x)$$

como queríamos probar. □

Ahora estamos en las condiciones de probar el teorema principal de esta sección, que nos va a ayudar a clasificar las funciones de distribuciones no-degeneradas asintóticas que pueden tener los valores extremos.

Teorema 2.3.2 (Fisher, Tippett, Gnedenko) La clase de distribuciones no-degeneradas del valor extremo máximo viene dada por $G_\gamma(ax + b)$ para $a > 0$ y b constantes reales, donde:

$$G_\gamma(x) = \exp(-(1 + \gamma x)^{-\frac{1}{\gamma}}) \text{ para } 1 + \gamma x > 0$$

γ es un número real y el caso $\gamma = 0$ es interpretado como $G_0(x) = \exp(-e^{-x})$.

Demostración: Sea D la clase de las funciones límite de $\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} =: D(x)$. Para la demostración, vamos a suponer primero que D es continua en el punto 1. Para todo punto $x > 0$ en el que tengamos continuidad, definimos:

$$E(x) := D(x) - D(1) = \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)}$$

que observamos que mantiene la continuidad por la izquierda de D . También consideramos un punto $y > 0$ y escribimos:

$$\frac{U(txy) - U(t)}{a(t)} = \frac{U(txy) - U(ty)}{a(ty)} \frac{a(ty)}{a(t)} + \frac{U(ty) - U(t)}{a(t)}$$

Vamos a probar que los límites $\lim_{t \rightarrow \infty} \frac{U(ty) - U(t)}{a(t)}$ y $\lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)}$ existen. Para ello, supongamos que no existen. Sean A_1, A_2, B_1 y B_2 con $A_1 \neq A_2$ o $B_1 \neq B_2$, con

B_i puntos límite de $\frac{U(ty)-U(t)}{a(t)}$ y A_i puntos límite de $\frac{a(ty)}{a(t)}$, $i = 1, 2$ cuando $t \rightarrow \infty$. Entonces, por la definición de $E(x)$, tenemos:

$$E(xy) = \lim_{t \rightarrow \infty} \frac{U(txy)-U(t)}{a(t)} = \lim_{t \rightarrow \infty} \left(\frac{U(txy)-U(ty)}{a(ty)} \frac{a(ty)}{a(t)} + \frac{U(ty)-U(t)}{a(t)} \right) = E(x)A_i + B_i$$

con $i = 1, 2$ para todo punto x en el que $E(\cdot)$ y $E(\cdot y)$ con continuas. Ahora, para un punto cualquiera x , consideramos una sucesión creciente de puntos $(x_n)_n$ en los que $E(\cdot)$ y $E(\cdot y)$ son continuas de forma que $\lim_{n \rightarrow \infty} x_n = x$. Como E es continua por la izquierda, se va a cumplir que:

$$\lim_{n \rightarrow \infty} E(x_n y) = E(xy) \text{ y } \lim_{n \rightarrow \infty} E(x_n) = E(x)$$

por lo que, como el conjunto de puntos en los que E es continua es denso en su dominio, se va a cumplir que para todo x e y positivos, se tiene $E(xy) = E(x)A_i + B_i$. Ahora, hacemos la diferencia entre las dos expresiones que se obtienen para $i = 1$ o $i = 2$, y obtenemos:

$$0 = E(x)(A_1 - A_2) + (B_1 - B_2) \Rightarrow B_2 - B_1 = E(x)(A_1 - A_2) \quad \forall x > 0$$

Pero esto implicaría que $E(x) = \frac{B_2 - B_1}{A_1 - A_2} \equiv Cte$, lo que a su vez haría que $D(x)$ fuese constante y, en fin, implicaría que la función $G(x) := \lim_{n \rightarrow \infty} [F(a_n x + b_n)]^n$ es una función constante, y por tanto, degenerada, lo que es una contradicción a las hipótesis iniciales. Por lo tanto, se debe tener $A_1 = A_2$ y también $B_1 = B_2$. En conclusión,

$$\lim_{t \rightarrow \infty} \frac{U(ty)-U(t)}{a(t)} = E(y) \text{ y } \lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)} =: A(y)$$

para todo $y > 0$, y por tanto, para todo $x > 0$ e $y > 0$ se cumple que:

$$E(xy) = E(x)A(y) + E(y)$$

Ahora consideramos $s := \log(x)$ y $t := \log(y)$ ($x, y \neq 1$) y definimos $H(x) := E(e^x)$. Entonces, a partir de la expresión anterior, se tiene:

$$H(t + s) = H(s)A(e^t) + H(t)$$

Además, como $H(0) = E(e^0) = E(1) = D(1) - D(1) = 0$, se tiene:

$$\frac{H(t+s)-H(t)}{s} = \frac{H(s)-H(0)}{s} A(e^t)$$

H es monótona, por lo que va a existir al menos un punto t_0 en el que sea diferenciable. Así, en ese punto t_0 se tiene:

$$H'(t_0) = \lim_{s \rightarrow 0} \frac{H(t_0+s)-H(t_0)}{s} = \lim_{s \rightarrow 0} \frac{H(s)-H(0)}{s} A(e^{t_0}) = k$$

por lo que vemos que el que sea diferenciable no depende del punto t en el que nos encontramos, lo que nos permite afirmar que H es diferenciable en todo su dominio. Así, tenemos:

$$\lim_{s \rightarrow 0} \frac{H(t+s)-H(t)}{s} = \lim_{s \rightarrow 0} \frac{H(s)-H(0)}{s} A(e^t) \Rightarrow H'(t) = H'(0)A(e^t)$$

Definimos ahora $Q(t) := \frac{H(t)}{H'(0)}$, que está bien definida porque $H'(0) \neq 0$, ya que en caso contrario se tendría $H'(t) = 0 \quad \forall t$, lo que a su vez haría que H fuese constante y llegaríamos de nuevo a la conclusión contradictoria de que la función de distribución límite G debe ser degenerada. Observamos que $Q(0) = 0$, que $Q'(0) = 1$ y que $Q(t+s) - Q(t) = Q(s)A(e^t)$. Por otra parte, de $H'(t) = H'(0)A(e^t)$ obtenemos que $A(e^t) = \frac{H'(t)}{H'(0)} = Q'(t)$, por lo que $Q(t+s) - Q(t) = Q(s)Q'(t)$. Ahora si hacemos

la diferencia entre las dos expresiones que se obtienen al intercambiar las variables s y t en la expresión anterior, obtenemos:

$$Q(t) \frac{Q'(s)-1}{s} = \frac{Q(s)}{s} (Q'(t) - 1) \Rightarrow Q(t) \frac{Q'(s)-Q'(0)}{s} = \frac{Q(s)}{s} (Q'(t) - 1)$$

donde, al hacer $s \rightarrow 0$, se obtiene:

$$Q(t)Q''(0) = Q'(t) - 1$$

De aquí vemos que Q es dos veces diferenciable, por lo que derivando en ambos lados de la expresión, se obtiene:

$$Q'(t)Q''(0) = Q''(t)$$

En consecuencia:

$$Q''(0) = \frac{Q''(t)}{Q'(t)} = [\log(Q'(t))]' =: \gamma \in \mathbb{R} \quad \forall t$$

Resolviendo la ecuación diferencial (teniendo en cuenta que $Q'(0) = 1$, se obtiene que $Q'(t) = e^{\gamma t}$ y como $Q(0) = 0$, tenemos:

$$Q(t) = \int_0^t e^{\gamma s} ds$$

Ahora de $Q(t)Q''(0) = Q'(t) - 1$, $Q(t) = \frac{H(t)}{H'(0)}$ y $E(x) := D(x) - D(1)$ obtenemos:

$$D(t) = D(1) + H'(0) \frac{e^{\gamma t} - 1}{\gamma}$$

Como $t = \log(y)$, despejando y en la expresión anterior, se obtiene la expresión de la función inversa continua por la derecha D^{-1} de D como:

$$D^{-1}(x) = (1 + \gamma \frac{x-D(1)}{H'(0)})^{\frac{1}{\gamma}}$$

Anteriormente hemos llegado también a que $D(y) = G^{-1}(e^{-\frac{1}{x}})$, de donde:

$$D^{-1}(x) = \frac{1}{-\log(G(x))}$$

Uniando ambas expresiones y haciendo el cambio de variable $z = \frac{x-D(1)}{H'(0)}$, obtenemos la función de distribución buscada:

$$G_\gamma(z) = \exp(-(1 + \gamma z)^{-\frac{1}{\gamma}})$$

como queríamos probar.

Si D no es continua en el punto 1, es suficiente con repetir la misma demostración para un punto x_0 en el que sea continua sustituyendo la función $U(t)$ con $U(tx_0)$.

Es análogo reformular el teorema para obtener la clase de distribuciones no-degeneradas para el valor extremo mínimo: $L_\gamma(x) = 1 - \exp(-(1 - \gamma x)^{-\frac{1}{\gamma}})$ o en el caso $\gamma = 0$, $L_0(x) = 1 - \exp(-e^x)$. □

La importancia del teorema reside en que demuestra que las funciones de distribución límite de los valores extremos forman una familia de funciones que depende de un solo parámetro, exceptuando los valores $D(1) =: \lambda$ y $H'(0) =: \delta$, que pueden ser interpretados como parámetro de posición y parámetro de escala, respectivamente. Podemos separar la clase en tres distintas subclases, siendo estas las tres únicas distribuciones asintóticas que puede tener un estadístico ordenado máximo:

- $\boxed{\gamma > 0}$: En este caso se tiene $G_\gamma(x) < 1 \forall x$ y se cumple que $G_\gamma(x) = 0$ para $x \leq \frac{-1}{\gamma}$. Una distribución destacable se obtiene considerando el cambio de variable $z = \frac{x-1}{\gamma}$:

$$\Phi_\alpha(z) := \begin{cases} 0 & \text{si } z \leq 0 \\ \exp(-z^{-\alpha}) & \text{si } z > 0 \end{cases}$$

donde $\alpha = \frac{1}{\gamma}$. Recibe el nombre de clase de **funciones de distribución de Fréchet (maximal)** en honor al matemático francés Maurice Fréchet.

- $\boxed{\gamma < 0}$: En este caso se cumple que $G_\gamma(x) = 1$ para todo $x \geq \frac{-1}{\gamma}$. Una distribución destacable, en este caso, se obtiene haciendo ahora el cambio de variable $z = -\frac{1+x}{\gamma}$:

$$\Psi_\alpha(z) := \begin{cases} \exp(-(-z)^\alpha) & \text{si } z < 0 \\ 1 & \text{si } z \geq 0 \end{cases}$$

donde $\alpha = -\frac{1}{\gamma}$. Recibe el nombre de clase de **funciones de distribución de Weibull (maximal)** en honor al matemático sueco Waloddi Weibull.

- $\boxed{\gamma = 0}$: Por la definición hemos visto que en este caso se tiene:

$$G_0(x) = \exp(-e^{-x})$$

Esta distribución recibe el nombre de **distribución de Gumbel (maximal)** en honor al matemático alemán Emil Gumbel.

Estas mismas distribuciones, para el caso del estadístico ordenado mínimo, tienen la siguiente forma:

- $\boxed{\gamma > 0}$:

$$\Phi_\alpha(z) := \begin{cases} 1 - \exp(-(-z)^{-\alpha}) & \text{si } z < 0 \\ 1 & \text{si } z \geq 0 \end{cases}$$

donde $\alpha = \frac{1}{\gamma}$. Recibe el nombre de clase de **funciones de distribución de Fréchet (minimal)**.

- $\boxed{\gamma < 0}$:

$$\Psi_\alpha(z) := \begin{cases} 0 & \text{si } z \leq 0 \\ 1 - \exp(-z^\alpha) & \text{si } z > 0 \end{cases}$$

donde $\alpha = -\frac{1}{\gamma}$. Recibe el nombre de clase de **funciones de distribución de Weibull (minimal)**.

- $\boxed{\gamma = 0}$:

$$L_0(x) = 1 - \exp(-e^{-x})$$

Esta distribución recibe el nombre de **distribución de Gumbel (minimal)**.

Ahora estamos en las condiciones de reformular el Teorema 2.3.1:

Teorema 2.3.3 Para un $\gamma \in \mathbb{R}$ las siguientes afirmaciones son equivalentes:

(I) Existen sucesiones $(a_n)_n$ (con $a_n > 0 \forall n$) y $(b_n)_n$ de forma que:

$$\lim_{n \rightarrow \infty} [F(a_n x + b_n)]^n = G_\gamma(x) = \exp(-(1 + \gamma x)^{-\frac{1}{\gamma}}) \quad \forall x \text{ con } 1 + \gamma x > 0$$

(II) Existe una función positiva a de forma que para todo $x > 0$:

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D_\gamma(x) = \frac{x^\gamma - 1}{\gamma} \quad (\text{si } \gamma = 0, D_0(x) = \log(x))$$

(III) Existe una función positiva a de forma que:

$$\lim_{t \rightarrow \infty} t(1 - F(a(t)x + U(t))) = (1 + \gamma x)^{-\frac{1}{\gamma}}$$

2.4. Clasificación en dominios de atracción

En esta sección vamos a ver criterios para que cierta función de distribución pertenezca al dominio de atracción de G_γ . Esto significa que si la función de distribución de ciertas variables aleatorias i.i.d. es F y $F \in \mathbb{D}(G_\gamma)$, la función de distribución asintótica de los estadísticos extremos de dichas variables aleatorias será G_γ . En primer lugar, vamos a ver una condición suficiente, la llamada condición de von Mises.

Definición 2.4.1 Se llama **extremo derecho** de una función de distribución F al punto $x^* = \sup\{x : F(x) < 1\}$.

Teorema 2.4.1 (Condición de von Mises): Sea F una función de distribución y sea x^* su extremo derecho. Supongamos que $F''(x)$ existe y que $F'(x)$ es positiva para todo x perteneciente a un entorno (por la izquierda) de x^* . Si se cumple:

$$\lim_{t \rightarrow x^*} \left(\frac{1-F(t)}{F'(t)} \right)' = \gamma, \text{ equivalente a } \lim_{t \rightarrow x^*} \frac{(1-F(t))F''(t)}{[F''(t)]^2} = -\gamma - 1$$

entonces $F \in \mathbb{D}(G_\gamma)$.

Demostración: Para la prueba de este teorema, al igual que en la sección anterior, nos será más útil utilizar la función U , que recordamos que era la inversa de $\frac{1}{1-F}$.

Ahora, si derivamos en la expresión $t = \frac{1}{1-F(U(t))}$, obtenemos:

$$1 = \frac{F'(U(t))U'(t)}{[1-F(U(t))]^2} \Rightarrow U'(t) = \frac{[1-F(U(t))]^2}{F'(U(t))}$$

Derivando de nuevo en esta expresión, se obtiene:

$$U''(t) = U'(t) \left[-2(1 - F(U(t))) - \frac{F''(U(t))[1-F(U(t))]^2}{[F'(U(t))]^2} \right] \Rightarrow$$

y utilizando que $\frac{1}{t} = 1 - F(U(t))$:

$$\frac{tU''(t)}{U'(t)} = -2 - \frac{F''(U(t))[1-F(U(t))]}{[F'(U(t))]^2}$$

Por otra parte, por los Teoremas 2.3.1 y 2.3.3, sabemos que lo que queremos probar es equivalente a probar que si $\lim_{t \rightarrow \infty} \frac{tU''(t)}{U'(t)} = -2 + \gamma + 1 = \gamma - 1$, entonces

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}.$$

Dado un $1 < x_0 < x$, $\int_{x_0}^x \frac{U''(s)}{U'(s)} ds = \log(U'(x)) - \log(U'(x_0)) = \log\left(\frac{U'(x)}{U'(x_0)}\right)$, por lo que dado $x > 0$ y $tx > 1$, se tiene:

$$\log\left(\frac{U'(tx)}{U'(t)}\right) = \int_1^x \frac{A(ts)}{s} ds \text{ donde } A(t) := \frac{tU''(t)}{U'(t)} \rightarrow \gamma - 1 \text{ si } t \rightarrow \infty$$

De aquí se obtiene, para $0 < a < b < \infty$:

$$\lim_{t \rightarrow \infty} \sup_{a \leq x \leq b} |\log\left(\frac{U'(tx)}{U'(t)}\right) - \log(x^{\gamma-1})| = 0$$

Y como en un entorno compacto se cumple que $|e^s - e^t| \leq c|s - t|$ para cierta constante c , tenemos:

$$\begin{aligned} & \lim_{t \rightarrow \infty} \sup_{a \leq x \leq b} \left| \frac{U'(tx)}{U'(t)} - x^{\gamma-1} \right| \leq \\ & \leq c \lim_{t \rightarrow \infty} \sup_{a \leq x \leq b} |\log\left(\frac{U'(tx)}{U'(t)}\right) - \log(x^{\gamma-1})| = 0 \end{aligned}$$

En consecuencia, por el teorema de la convergencia dominada, obtenemos:

$$\lim_{t \rightarrow \infty} \int_1^x \left(\frac{U'(ts)}{U'(t)} - s^{\gamma-1} \right) ds = \lim_{t \rightarrow \infty} \left[\frac{U(tx) - U(t)}{tU'(t)} - \frac{x^{\gamma-1}}{\gamma} \right] = 0$$

Lo que, como hemos dicho, nos permite afirmar que $F \in \mathbb{D}(G_\gamma)$. □

De manera equivalente, para la función U , la condición de von Mises es:

Corolario 2.4.1 Sea F una función de distribución, sea x^* su extremo derecho y sea U la función inversa de la función $\frac{1}{1-F}$. Supongamos que $F''(x)$ existe y que $F'(x)$ es positiva para todo x perteneciente a un entorno (por la izquierda) de x^* . Si se cumple:

$$\lim_{t \rightarrow \infty} \frac{tU''(t)}{U'(t)} = \gamma - 1$$

entonces $F \in \mathbb{D}(G_\gamma)$.

Es posible encontrar condiciones más sencillas en el caso en el que $\gamma \neq 0$. En primer lugar, supongamos que $\gamma > 0$.

Teorema 2.4.2 Dada una función de distribución F , supongamos que $x^* = \infty$ y que existe F' . Si $\lim_{t \rightarrow \infty} \frac{tF'(t)}{1-F(t)} = \frac{1}{\gamma}$, entonces $F \in \mathbb{D}(G_\gamma)$.

Demostración: Al igual que en la demostración del teorema anterior, a partir de $t = \frac{1}{1-F(U(t))}$, derivando obtenemos $U'(t) = \frac{1-F(U(t))}{tF'(U(t))}$, de donde:

$$\lim_{t \rightarrow \infty} \frac{tU'(t)}{U(t)} = \lim_{t \rightarrow \infty} \frac{1-F(U(t))}{U(t)F'(U(t))} = \gamma$$

Por otra parte, se tiene:

$$\log(U(tx)) - \log(U(t)) = \log\left(\frac{U(tx)}{U(t)}\right) = \int_1^x \frac{tsU'(ts)}{U(ts)} ds$$

de donde, procediendo como en la demostración del Teorema 2.4.1, $\forall x > 0$, se obtiene:

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma \Rightarrow \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{\gamma U(t)} = \frac{x^\gamma - 1}{\gamma}$$

lo que, finalmente, por el Teorema 2.3.3 significa que $F \in \mathbb{D}(G_\gamma)$, como queríamos probar. □

Corolario 2.4.2 Dada una función de distribución F , supongamos que $x^* = \infty$ y que existe F' . Si $\lim_{t \rightarrow \infty} \frac{tU'(t)}{U(t)} = \gamma$, entonces $F \in \mathbb{D}(G_\gamma)$.

Supongamos finalmente que $\gamma < 0$.

Teorema 2.4.3 Dada una función de distribución F , supongamos que $x^* < \infty$ y que existe F' para $x < x^*$. Si $\lim_{t \rightarrow x^*} \frac{(x^*-t)F'(t)}{1-F(t)} = -\frac{1}{\gamma}$, entonces $F \in \mathbb{D}(G_\gamma)$.

Demostración: Al igual que antes, a partir de $t = \frac{1}{1-F(U(t))}$, derivando obtenemos $tU'(t) = \frac{1-F(U(t))}{F'(U(t))}$, de donde:

$$\lim_{t \rightarrow \infty} \frac{tU'(t)}{U(\infty)-U(t)} = \lim_{t \rightarrow \infty} \frac{1-F(U(t))}{(x^*-U(t))F'(U(t))} = -\gamma$$

Por otra parte, se tiene:

$$\log[U(\infty) - U(tx)] - \log[U(\infty) - U(t)] = \log\left(\frac{U(\infty)-U(tx)}{U(\infty)-U(t)}\right) = -\int_1^x \frac{tsU'(ts)}{U(\infty)-U(ts)} \frac{ds}{s}$$

de donde, procediendo como en la demostración del Teorema 2.4.1, $\forall x > 0$, se obtiene:

$$\lim_{t \rightarrow \infty} \frac{U(\infty)-U(tx)}{U(\infty)-U(t)} = x^\gamma \Rightarrow \lim_{t \rightarrow \infty} \frac{U(tx)-U(t)}{-\gamma[U(\infty)-U(t)]} = \frac{x^\gamma-1}{\gamma}$$

lo que, finalmente, por el Teorema 2.3.3 significa que $F \in \mathbb{D}(G_\gamma)$, como queríamos probar. □

Corolario 2.4.3 Dada una función de distribución F , supongamos que $x^* < \infty$ y que existe F' para $x < x^*$. Si $\lim_{t \rightarrow \infty} \frac{tU'(t)}{U(\infty)-U(t)} = -\gamma$, entonces $F \in \mathbb{D}(G_\gamma)$.

Finalmente, para acabar el capítulo, veamos el enunciado de un teorema que nos da una condición necesaria y suficiente para que una función de distribución no-degenerada se encuentre en el dominio de atracción de G_γ y un corolario que nos dice que cualquier función de distribución que se encuentra en el dominio de atracción de G_γ va a tener como distribución límite la función de distribución de Fréchet, la de Weibull o la de Gumbel. Las respectivas demostraciones de este teorema y este corolario precisan de la introducción y la demostración de conceptos que exceden el objetivo de este trabajo, por lo que solamente se van a ver enunciados. Para un desarrollo completo de sus demostraciones junto a los conceptos necesarios, véase [11].

Teorema 2.4.4 Dada una función de distribución F , se va a cumplir que $F \in \mathbb{D}(G_\gamma)$ si y solo si:

- (I) Para $\gamma > 0$, $x^* = \infty$ y $\lim_{t \rightarrow \infty} \frac{1-F(tx)}{1-F(t)} = x^{-\frac{1}{\gamma}} \forall x > 0$.
- (II) Para $\gamma < 0$, $x^* < \infty$ y $\lim_{t \rightarrow 0^+} \frac{1-F(x^*-tx)}{1-F(x^*-t)} = x^{-\frac{1}{\gamma}} \forall x > 0$.
- (III) Para $\gamma = 0$, x^* es finito o infinito y $\lim_{t \rightarrow x^*} \frac{1-F(t+xf(t))}{1-F(t)} = e^{-x} \forall x \in \mathbb{R}$, donde f es una función positiva adecuada. Además, si dicha función f existe, se cumple $\int_1^{x^*} [1-F(s)]ds < \infty$ para $t < x^*$, y el límite también converge al mismo valor e^{-x} para la función $f(t) := \frac{\int_1^{x^*} [1-F(s)]ds}{1-F(t)}$.

Corolario 2.4.4 Si $F \in \mathbb{D}(G_\gamma)$, se tiene:

- (i) Para $\gamma > 0$, $\lim_{n \rightarrow \infty} [F(a_n x)]^n = \exp(-x^{-\frac{1}{\gamma}}) \forall x > 0$, donde $a_n := U(n)$.
- (ii) Para $\gamma < 0$, $\lim_{n \rightarrow \infty} [F(a_n x + x^*)]^n = \exp(-(-x)^{-\frac{1}{\gamma}}) \forall x < 0$, donde $a_n := x * -U(n)$.
- (iii) Para $\gamma = 0$, $\lim_{n \rightarrow \infty} [F(a_n x + b_n)]^n = \exp(e^{-x}) \forall x \in \mathbb{R}$, donde $a_n := f(U(n))$, $b_n = U(n)$ y f es una función creciente como en el enunciado del teorema anterior.

Estimación de parámetros

3.1. Introducción

En este capítulo se van a presentar unos métodos de estimación aplicados a la estimación de los parámetros de la función de distribución asintótica de los valores extremos (véase [3]).

Recordamos del capítulo anterior que la función de distribución de los valores extremos (para el caso máximo) viene dada por

$$G_\gamma(x) = \exp(-(1 + \gamma x)^{-\frac{1}{\gamma}})$$

o, teniendo también en cuenta los parámetros de posición y escala:

$$G_\gamma(x) = \exp(-(1 + \gamma \frac{x-\lambda}{\delta})^{-\frac{1}{\gamma}}) \text{ o}$$

$$G_0(x) = \exp(-\exp(\frac{\lambda-x}{\delta})) \text{ si } \gamma = 0$$

Así, su función de densidad se obtiene derivando $G_\gamma(x)$ (o $G_0(x)$) respecto de x :

$$g_\gamma(x) = \exp(-(1 + \gamma \frac{x-\lambda}{\delta})^{-\frac{1}{\gamma}}) (1 + \gamma \frac{x-\lambda}{\delta})^{-\frac{\gamma+1}{\gamma}} \frac{1}{\delta}$$

$$g_0(x) = \exp(-\exp(\frac{\lambda-x}{\delta})) \exp(\frac{\lambda-x}{\delta}) \frac{1}{\delta}$$

Sus dominios son $x \geq \lambda + \frac{\delta}{\gamma}$ si $\gamma < 0$ o $x \leq \lambda + \frac{\delta}{\gamma}$ si $\gamma > 0$ y $-\infty < x < \infty$ respectivamente.

Al igual que ha pasado en el capítulo anterior, las estimaciones se van a hacer para el caso del valor extremo máximo. Se utiliza el siguiente teorema para poder reformular los resultados para estimaciones del valor extremo mínimo:

Teorema 3.1.1 Si una variable aleatoria X tiene función de distribución $G_\gamma(x)$ con parámetros λ y δ , entonces la variable aleatoria $Y = -X$ tiene función de distribución $L_\gamma(x)$ con parámetros $-\lambda$ y δ , y viceversa.

Demostración: Si $Y = -X$, se tiene:

$$F_Y(y) = P(Y \leq y) = P(-X \leq y) = P(X \geq -y) = 1 - P(X < -y) = 1 - G_\gamma(-y) =$$

$$1 - \exp(-(1 + \gamma \frac{(-y)-\lambda}{\delta})^{-\frac{1}{\gamma}}) = 1 - \exp(-(1 - \gamma \frac{y-(-\lambda)}{\delta})^{-\frac{1}{\gamma}}) = L_\gamma(y).$$

Observamos que ahora el parámetro de posición tiene signo negativo. La prueba para el caso inverso es similar. □

Los pasos que se siguen para la estimación de parámetros en el caso del mínimo son:

1. Se cambia el signo de los valores de la información: $x_i \rightarrow -x_i$ para $i = 1, 2, \dots, n$.
2. Se hace la estimación de los parámetros $\hat{\delta}$, $\hat{\lambda}$ y $\hat{\gamma}$ para el máximo.
3. Ahora $\delta = \hat{\delta}$, $\lambda = -\hat{\lambda}$ y $\gamma = \hat{\gamma}$ son los parámetros estimados para el mínimo.

3.2. Estimación de máxima verosimilitud (EMV)

Vamos a diferenciar los casos $\gamma \neq 0$ y $\gamma = 0$. Si $\mathbf{x} = (x_1, x_2, \dots, x_n)$, en el caso $\gamma \neq 0$, la función logarítmica de verosimilitud viene dada, para $\theta = \{\lambda, \delta, \gamma\}$, por:

$$\begin{aligned} \log(\mathfrak{L}(\theta; x)) &= \sum_{i=1}^n \log[\exp(-(1 + \gamma \frac{x_i - \lambda}{\delta})^{-\frac{1}{\gamma}})(1 + \gamma \frac{x_i - \lambda}{\delta})^{-\frac{\gamma+1}{\gamma}} \frac{1}{\delta}] = \\ &= -\sum_{i=1}^n \log[\exp(1 + \gamma \frac{x_i - \lambda}{\delta})^{-\frac{1}{\gamma}}] - \sum_{i=1}^n \log(1 + \gamma \frac{x_i - \lambda}{\delta})^{-\frac{\gamma+1}{\gamma}} - \sum_{i=1}^n \log[\delta] = \\ &= -\sum_{i=1}^n \exp[-\frac{1}{\gamma} \log(1 + \gamma \frac{x_i - \lambda}{\delta})] - (1 + \gamma) \sum_{i=1}^n \log[1 + \gamma \frac{x_i - \lambda}{\delta}] - n \log(\delta) = \\ &= -n \log(\delta) - (1 + \gamma) \sum_{i=1}^n z_i - \sum_{i=1}^n \exp(-z_i) \text{ donde } z_i = \frac{1}{\gamma} \log[1 + \gamma \frac{x_i - \lambda}{\delta}] \end{aligned}$$

De manera similar, para $\theta = \{\lambda, \delta\}$ y $\gamma = 0$, se tiene:

$$\log(\mathfrak{L}(\theta; x)) = -n \log(\delta) - \sum_{i=1}^n \exp[-(\frac{x_i - \lambda}{\delta})] - \sum_{i=1}^n (\frac{x_i - \lambda}{\delta})$$

El problema ahora es maximizar estas funciones. Lo más rápido es utilizar métodos numéricos y paquetes de optimización. Vamos a ver a continuación un método numérico iterativo. De nuevo vamos a separar los casos $\gamma \neq 0$ y $\gamma = 0$.

Caso $\gamma \neq 0$

Se considera la fórmula iterativa:

$$\theta_{j+1} = \theta_j + \mathbf{I}_\theta^{-1} \nabla_\theta l \text{ para } j = 0, 1, \dots$$

El método es obtenido de aplicar el método de Newton descrito en el capítulo 1 a la función gradiente de la función logarítmica de verosimilitud y sustituir la matriz de las derivadas parciales de esta (hessiana de la función logarítmica) por la matriz de información de Fisher.

Tanto \mathbf{I}_θ^{-1} como $\nabla_\theta l$ están evaluados en θ_j . Aquí l es el vector que en cada entrada tiene la función logarítmica de verosimilitud con un parámetro como variable y los otros dos como constante; $\nabla_\theta l$ es su gradiente e \mathbf{I}_θ es la llamada matriz de información de Fisher, dada en este caso por:

$$\mathbf{I}_\theta = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$$

donde:

- $m_{11} = E(\frac{\partial^2 l}{\partial \lambda^2}) = \frac{n}{\delta^2} p$
- $m_{22} = E(\frac{\partial^2 l}{\partial \delta^2}) = \frac{n}{\delta^2 \gamma^2} (1 - 2\Gamma(2 + \gamma) + p)$
- $m_{33} = E(\frac{\partial^2 l}{\partial \gamma^2}) = \frac{n}{\gamma^2} (\frac{\pi^2}{6} + (1 - \hat{\gamma} + \frac{1}{\gamma})^2 - \frac{2q}{\gamma} + \frac{p}{\gamma^2})$
- $m_{12} = m_{21} = E(\frac{\partial^2 l}{\partial \lambda \partial \delta}) = -\frac{n}{\delta^2 \gamma} (p - \Gamma(2 + \gamma))$
- $m_{13} = m_{31} = E(\frac{\partial^2 l}{\partial \lambda \partial \gamma}) = \frac{n}{\delta \gamma} (q - \frac{p}{\gamma})$
- $m_{23} = m_{32} = E(\frac{\partial^2 l}{\partial \delta \partial \gamma}) = \frac{n}{\delta \gamma^2} (1 - \hat{\gamma} + \frac{1 - \Gamma(2 + \gamma)}{\gamma} - q + \frac{p}{\gamma})$

Γ es la función Gamma con $\Gamma(u) = \int_0^\infty y^{u-1} e^{-y} dy$; $\hat{\gamma}$ es la constante de Euler-Mascheroni con valor $\hat{\gamma} = \lim_{n \rightarrow \infty} (-\ln(n) + \sum_{k=1}^n \frac{1}{k}) = \int_0^\infty (\frac{1}{[x]} - \frac{1}{x}) dx = 0.57721566\dots$;

$p = (1 + \gamma)^2 \Gamma(1 + 2\gamma)$; $\psi(u) = \frac{d \log(\Gamma(u))}{du}$ y $q = \Gamma(2 + \gamma) [\psi(1 + \gamma) + \frac{1+\gamma}{\gamma}]$.

Observamos que para obtener una estimación de los parámetros utilizando este método es necesario tener un $\theta_0 = \{\lambda_0, \delta_0, \gamma_0\}$, es decir, unas estimaciones iniciales. Una manera para proceder es utilizar alguno de los métodos que se van a ver en las siguientes secciones.

Por otra parte, para los **intervalos de confianza**, se va a utilizar lo visto en el capítulo 1. Así, una vez maximizada la función de verosimilitud y obtenida la EMV, a partir de la inversa de la matriz de información de Fischer \mathbf{I}_θ , evaluada en el punto $(\hat{\lambda}, \hat{\delta}, \hat{\gamma})$, obtenemos la matriz de covarianza estimada de $(\hat{\lambda}, \hat{\delta}, \hat{\gamma})$. Las raíces cuadradas de los elementos de la diagonal de esta matriz, $(\hat{\sigma}_\lambda, \hat{\sigma}_\delta, \hat{\sigma}_\gamma)$, son las respectivas desviaciones estándar de los respectivos valores estimados $(\hat{\lambda}, \hat{\delta}, \hat{\gamma})$. Por tanto, los $(1 - \alpha)100\%$ intervalos de confianza de los parámetros son:

$$\begin{aligned}\lambda &\in (\hat{\lambda} \pm z_{\frac{\alpha}{2}} \hat{\sigma}_\lambda) \\ \delta &\in (\hat{\delta} \pm z_{\frac{\alpha}{2}} \hat{\sigma}_\delta) \\ \gamma &\in (\hat{\gamma} \pm z_{\frac{\alpha}{2}} \hat{\sigma}_\gamma)\end{aligned}$$

Caso $\gamma = 0$

Es análogo al caso $\gamma \neq 0$, se utiliza la fórmula iterativa:

$$\theta_{j+1} = \theta_j + \mathbf{I}_\theta^{-1} \nabla_\theta l \text{ para } j = 0, 1, \dots$$

y sus términos significan lo mismo que antes, con la diferencia que ahora l es un vector formado por solo dos entradas, lo que también cambia el tamaño de la matriz de información de Fischer, que ahora es:

$$\mathbf{I}_\theta = \begin{bmatrix} \frac{n}{\delta^2} & \frac{n(\hat{\gamma}-1)}{\delta^2} \\ \frac{n(\hat{\gamma}-1)}{\delta^2} & \frac{n(6-12\hat{\gamma}+6\hat{\gamma}^2+n^2)}{6\delta^2} \end{bmatrix}$$

Procediendo de una manera similar al caso $\gamma \neq 0$, con la respectiva información para $\theta = (\lambda, \delta)$, se obtienen los **intervalos de confianza**:

$$\begin{aligned}\lambda &\in (\hat{\lambda} \pm z_{\frac{\alpha}{2}} \hat{\sigma}_\lambda) \\ \delta &\in (\hat{\delta} \pm z_{\frac{\alpha}{2}} \hat{\sigma}_\delta)\end{aligned}$$

3.3. Método elemental de percentiles (MEP)

En el caso de la estimación de parámetros para el valor extremo máximo el método EMV puede presentar problemas como no poder encontrar un máximo local, no mantener las propiedades asintóticas, necesitar soluciones numéricas, etc. En esta sección vamos a ver como aplicar el **método elemental de percentiles** y se va a describir un algoritmo para resolver la ecuación de donde se obtiene la estimación.

Para este método vamos a necesitar la función cuantil correspondiente a la función de distribución del valor extremo máximo (asintótica). Para ello, se invierte

la función $G_\gamma(x)$ o $G_0(x)$, obteniendo:

$$\begin{aligned}x_p &= \lambda + \frac{\delta}{\gamma}[-1 + (-\log(p))^{-\gamma}] \text{ o} \\x_p &= \lambda - \delta \log(-\log(p))\end{aligned}$$

respectivamente.

Se va a separar la descripción del método en sus dos fases y para la primera fase se va a diferenciar el caso $\gamma \neq 0$ del caso $\gamma = 0$, que al final de esta fase ya tendrá una solución definitiva.

3.3.1. Fase 1: Las Estimaciones Iniciales

Caso $\gamma \neq 0$

En este caso $\theta = (\lambda, \delta, \gamma)$, es decir, hay tres parámetros, por lo que cada conjunto elemental está formado por tres estadísticos ordenados distintos. Sea $I = (i, j, k) : i < j < k \in \{1, 2, \dots, n\}$ un conjunto de índices de tres estadísticos ordenados distintos. Igualando el cuantil de la muestra con el cuantil teórico dado por la función cuantil, se obtiene:

$$\begin{aligned}x_{i:n} &= \lambda + \frac{\delta}{\gamma}[-1 + (-\log(p_{i:n}))^{-\gamma}] \\x_{j:n} &= \lambda + \frac{\delta}{\gamma}[-1 + (-\log(p_{j:n}))^{-\gamma}] \\x_{k:n} &= \lambda + \frac{\delta}{\gamma}[-1 + (-\log(p_{k:n}))^{-\gamma}]\end{aligned}$$

donde $p_{r:n}$ es un punto de posición gráfica que en este caso es considerado con el valor $p_{r:n} = \frac{r-0.35}{n}$. Restándole el tercero al segundo y después al primero y tomando la razón entre estas diferencias se obtiene la siguiente ecuación:

$$\frac{x_{j:n} - x_{k:n}}{x_{i:n} - x_{k:n}} = \frac{C_j^{-\gamma} - C_k^{-\gamma}}{C_i^{-\gamma} - C_k^{-\gamma}} = \frac{1 - A_{jk}^{-\gamma}}{1 - A_{ik}^{-\gamma}}$$

donde $C_r = -\log(p_{r:n})$ y $A_{rl} = \frac{C_r}{C_l}$. Se ha obtenido así una ecuación donde la única incógnita es γ . Resolviéndola se obtiene la estimación $\hat{\gamma}_{ijk}$, indicando que es una estimación en función de los estadísticos i -ésimo, j -ésimo y k -ésimo. Sustituyendo $\hat{\gamma}_{ijk}$ en las ecuaciones obtenidas al igualar el cuantil teórico con el de la muestra y resolviendo el sistema, se obtienen las estimaciones:

$$\begin{aligned}\hat{\delta}_{ijk} &= \frac{\hat{\gamma}_{ijk}(x_{j:n} - x_{i:n})}{C_j^{-\hat{\gamma}_{ijk}} - C_i^{-\hat{\gamma}_{ijk}}} \text{ y} \\ \hat{\lambda}_{ijk} &= x_{i:n} - \frac{\hat{\delta}_{ijk}(C_i^{-\hat{\gamma}_{ijk}} - 1)}{\hat{\gamma}_{ijk}}\end{aligned}$$

Para mantener la coherencia de las predicciones, deben cumplir las condiciones impuestas por los dominios de las distintas funciones de distribución. Así, para $\hat{\gamma} < 0$ debe cumplirse $x_{n:n} \leq \hat{\lambda} - \frac{\hat{\delta}}{\hat{\gamma}}$ y para $\hat{\gamma} > 0$ debe cumplirse $x_{1:n} \geq \hat{\lambda} - \frac{\hat{\delta}}{\hat{\gamma}}$.

Ahora veamos un algoritmo útil para resolver la ecuación que nos proporciona una estimación inicial de $\hat{\gamma}$. Llamamos $D_{ijk} := \frac{x_{j:n} - x_{k:n}}{x_{i:n} - x_{k:n}}$ y consideramos la ecuación

$D_{ijk} = \frac{1 - A_{jk}^{-\gamma}}{1 - A_{ik}^{-\gamma}}$, que al resolverla para γ , obtenemos una estimación $\hat{\gamma}_{ijk}$. Para ello:

1. Si $D_{ijk} < \frac{\log(A_{jk})}{\log(A_{ik})}$, se usa el método de bisección para obtener el 0 de la función:

$$h(\gamma) = \frac{1-A_{jk}^{-\gamma}}{1-A_{ik}^{-\gamma}} - D_{ijk}$$

en el intervalo $(\frac{\log(D_{ijk})}{\log(A_{ji})}, 0)$.

2. En caso contrario, se usa el método de bisección para encontrar el 0 en el intervalo $(0, \frac{\log(1-D_{ijk})}{\log(A_{jk})})$.

Caso $\gamma = 0$

En este caso $\theta = (\lambda, \delta)$, por lo que los conjuntos elementales solo están formados por dos estadísticos ordenados distintos, es decir, $I = (i, j)$ con $i < j \in \{1, 2, \dots, n\}$. De nuevo, igualando el cuantil de la muestra con el cuantil teórico dado por la función cuantil, se obtiene:

$$\begin{aligned} x_{i:n} &= \lambda - \delta \log(-\log(p_{i:n})) \\ x_{j:n} &= \lambda - \delta \log(-\log(p_{j:n})) \end{aligned}$$

En este caso es inmediato resolver el sistema y obtener las estimaciones:

$$\begin{aligned} \hat{\delta}_{ij} &= \frac{x_{i:n} - x_{j:n}}{\log(C_j) - \log(C_i)} \text{ y} \\ \hat{\lambda}_{ij} &= x_{i:n} + \hat{\delta}_{ij} \log(C_i) \end{aligned}$$

siendo $C_k = -\log(p_{k:n})$.

3.3.2. Fase 2: Estimaciones Finales

Las estimaciones obtenidas en el paso anterior solo están basadas en tres estadísticos ordenados distintos. Ahora, para obtener unas estimaciones más robustas y estadísticamente eficientes, vamos a usar otros estadísticos ordenados de la siguiente manera. Se considera un número N de conjuntos elementales de 3 estadísticos ordenados, que puede ser un número prefijado, un número al azar o incluso pueden ser todas las combinaciones distintas posibles. Para cada uno de estos conjuntos se da una estimación como la de la fase anterior, obteniendo un conjunto $\{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n\}$, con $\hat{\theta}_i = \{\hat{\lambda}_i, \hat{\delta}_i, \hat{\gamma}_i\}$, de estimaciones elementales. Las que son incoherentes según el criterio del dominio se descartan. Finalmente se utiliza alguna función, preferiblemente robusta, para combinar estas estimaciones elementales y obtener una estimación final. Ejemplos de este tipo de funciones robustas son la mediana o la media α -truncada. Así, pueden ser estimaciones finales de $\theta = \{\lambda, \delta, \gamma\}$:

$$\begin{aligned} \hat{\gamma}_{MED} &= \text{Mediana}\{\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_n\} \\ \hat{\delta}_{MED} &= \text{Mediana}\{\hat{\delta}_1, \hat{\delta}_2, \dots, \hat{\delta}_n\} \\ \hat{\lambda}_{MED} &= \text{Mediana}\{\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_n\} \end{aligned}$$

o

$$\begin{aligned} \hat{\gamma}_{MT_\alpha} &= MT_\alpha\{\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_n\} \\ \hat{\delta}_{MT_\alpha} &= MT_\alpha\{\hat{\delta}_1, \hat{\delta}_2, \dots, \hat{\delta}_n\} \\ \hat{\lambda}_{MT_\alpha} &= MT_\alpha\{\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_n\} \end{aligned}$$

Para evitar incoherencias se puede estimar $\lambda - \frac{\delta}{\gamma}$ en vez de λ , es decir, estimar:

$$(\hat{\lambda} - \frac{\hat{\delta}}{\hat{\gamma}})_{MED} = \text{Mediana}\{(\hat{\lambda} - \frac{\hat{\delta}}{\hat{\gamma}})_{1,2,n}, (\hat{\lambda} - \frac{\hat{\delta}}{\hat{\gamma}})_{1,3,n}, \dots, (\hat{\lambda} - \frac{\hat{\delta}}{\hat{\gamma}})_{1,n-1,n}\}$$

y después obtener $\hat{\lambda}$ de:

$$\hat{\lambda}_{MED} = (\hat{\lambda} - \frac{\hat{\delta}}{\hat{\gamma}})_{MED} + \frac{\hat{\delta}_{MED}}{\hat{\gamma}_{MED}}$$

De manera similar se hace lo mismo para la media α -truncada.

Se procede de manera análoga en el caso $\gamma = 0$.

3.4. Método de los mínimos cuadrados cuantílico (MCC)

Tal como su nombre indica, este método consiste en aplicar el método de los mínimos cuadrados a los cuantiles. Así, se va a tratar de minimizar la suma de los cuadrados de las diferencias entre los cuantiles teóricos y los observados. Se trata de un problema de optimización, que en el caso $\gamma \neq 0$ es:

$$\min_{\lambda, \delta, \gamma} \sum_{i=1}^n [x_{i:n} - \lambda - \frac{\delta}{\gamma} [-1 + (-\log(p_{i:n}))^{-\gamma}]^2$$

y para el caso $\gamma = 0$, se tiene:

$$\min_{\lambda, \delta, \gamma} \sum_{i=1}^n [x_{i:n} - \lambda + \delta \log(-\log(p_{i:n}))]^2$$

Estos dos métodos, MEP y MCC, pueden ser utilizados para obtener una estimación inicial para EMV.

Ejemplos prácticos

4.1. Introducción

En este último capítulo se van a analizar dos casos prácticos. Se van a utilizar dos métodos distintos para obtener las maximizaciones de la función logarítmica de máxima verosimilitud, y por tanto, las estimaciones para λ , δ y γ : un método de optimización ya implementado en R en el paquete “evd” y el método iterativo descrito en la sección 3.2 de este trabajo con una estimación inicial obtenida por el método descrito en la sección 3.3. El código de estos métodos del segundo caso está descrito en el anexo del final.

Mi opción ha sido escoger dos fenómenos locales cuya naturaleza es conocida: las altas temperaturas en un verano en Murcia y el bajo caudal del río Segura en su paso por la región de Murcia (por el municipio Abarán). El objetivo no es llegar a alguna conclusión nueva respecto de estos dos fenómenos sino que, partiendo de unos fenómenos conocidos, se pretende obtener una confirmación de que la modelización de valores extremos es efectiva mediante las distribuciones y los métodos teóricos presentados en los capítulos anteriores.

Un concepto que es frecuentemente encontrado en las aplicaciones prácticas de la modelización estadística es el siguiente:

Definición 4.1.1 Si $F(x)$ es la función de distribución obtenida al modelizar el estadístico ordenado máximo en cierto período de tiempo de una variable aleatoria X , se denomina **período de retorno**, notado con τ_x , del suceso $\{X > x\}$ al valor $\tau_x := \frac{1}{1-F(x)}$, medido en la misma longitud de tiempo que el período cubierto por dicho máximo. El valor es el promedio de periodos en los que se mide el fenómeno correspondiente, por ejemplo en años si lo que medimos es el máximo anual, que tardaremos en observar un valor máximo de magnitud mayor que x . De manera análoga, para el estadístico ordenado mínimo, el **período de retorno** del suceso $\{X < x\}$ es $\tau_x := \frac{1}{F(x)}$.

La interpretación que se le puede dar a este concepto es que proporciona una idea de como de raro puede ser considerado cierto suceso. Así, si suponemos que cierto proyecto de ingeniería falla si y solo si se cumple el suceso A, la vida media de dicho proyecto es el período de retorno de A. También se utiliza en el sentido de que, a priori, un proyecto de ingeniería suele ser condicionado por criterios de diseño que dependen de períodos de retorno, como por ejemplo pedir que una presa tenga una vida útil de al menos 50 años.

4.2. Temperaturas máximas

Los primeros datos analizados han sido tomados de la Agencia Estatal de Meteorología y representan la temperatura máxima en grados centígrados (°C) que se ha dado en Murcia en cada día del verano de 2016, considerando verano el conjunto

de los meses Junio, Julio y Agosto. Dichos datos se encuentran en la Figura 1.

30.8	31.5	31.8	32.7	31.0	32.5	31.5	35.6
37.2	37.2	33.5	34.6	40.3	35.1	35.7	29.7
30.4	30.4	29.8	30.1	30.7	31.1	33.1	37.0
33.2	30.0	30.2	31.5	34.3	36.8	35.5	37.7
32.2	31.1	37.3	34.6	31.4	37.5	34.9	37.9
38.1	38.3	32.2	31.0	30.3	30.4	32.2	31.7
34.2	36.9	37.4	35.2	32.1	32.9	34.6	36.4
36.1	34.9	36.1	40.1	40.2	32.6	34.3	35.9
40.0	34.0	32.7	32.6	34.5	34.1	30.5	31.6
30.7	31.5	32.0	33.0	32.5	38.4	36.8	38.4
39.6	33.1	33.1	32.7	33.3	33.3	32.6	35.0
37.7	34.5	32.5	32.2				

Figura 1: Máximas en °C, Verano 2016, Murcia

Un análisis inicial nos da la siguiente información:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
29.70	31.67	33.25	33.98	36.10	40.30

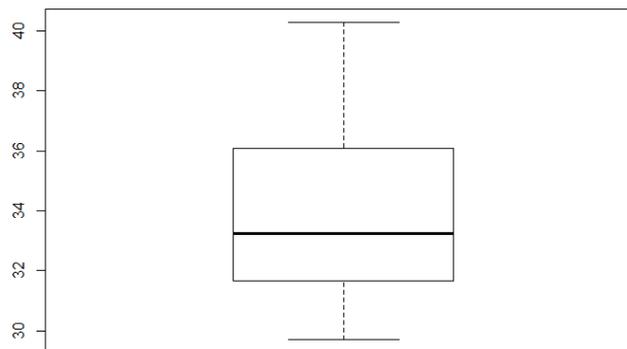


Figura 2: Gráfico de cajas de las temperaturas

Esto nos permite hacernos una primera idea del comportamiento de estas temperaturas máximas.

Ahora veamos qué distribución de valor extremo máximo se le asocia mejor a estos datos. Para ello, se va a tratar de estimar los tres parámetros de los que depende dicha función de distribución. Utilizando en primer lugar el paquete de R “evd”, con la función “fgev”, obtenemos los siguientes valores:

$$\begin{aligned}\lambda &= 32.68802 \\ \delta &= 2.30714 \\ \gamma &= -0.02814\end{aligned}$$

Esto nos indica un valor del parámetro γ muy cercano a 0. Además, con la función “confint” se obtienen los respectivos intervalos de confianza para un $\alpha = 0.05$, que son:

$$\begin{aligned}\lambda &\in [32.11441, 33.26164] \\ \delta &\in [1.86737, 2.74691] \\ \gamma &\in [-0.25996, 0.20367]\end{aligned}$$

lo que nos indica la posibilidad de que $\gamma = 0$. Y es que aplicando, por otra parte, el método iterativo, en caso de no suponer que $\gamma = 0$, a partir de cierto paso la matriz de información de Fisher pasará a tomar valores muy cercanos a 0, lo que hará que su determinante sea prácticamente 0 y en consecuencia, se obtendrá un error que indica que dicha matriz no tiene inversa. Utilizando, por tanto, el método iterativo para la estimación de solo dos parámetros, con 500 conjuntos de índices elementales distintos considerados a la hora de obtener una estimación inicial, se obtiene la siguiente estimación para λ y δ :

$$\begin{aligned}\lambda &= 32.65688 \\ \delta &= 2.28858\end{aligned}$$

Observamos que la diferencia es del orden de las centésimas, por lo que podemos confiar en que el método iterativo ha funcionado correctamente en este caso y que se puede considerar $\gamma = 0$. Ahora, si finalmente se considera como estimación definitiva para λ y δ la media de estas dos estimaciones:

$$\lambda = 32.67245 \text{ y } \delta = 2.29836$$

y $\gamma = 0$, podemos modelar la función de distribución de las temperaturas máximas del verano de 2016 de Murcia según la función de distribución:

$$G_0(x) = \exp\left(-e^{-\frac{x-32.67245}{2.29836}}\right)$$

con dominio todo \mathbb{R} y función de densidad de probabilidad:

$$g_0(x) = \frac{1}{2.29836} \exp\left(-e^{-\frac{x-32.67245}{2.29836}}\right) e^{-\frac{x-32.67245}{2.29836}}$$

también definida en todo \mathbb{R} . Pueden ser vistas sus respectivas representaciones en las figuras 3 y 4.

Esto nos permite sacar conclusiones que, por la experiencia, sabemos que se corresponden con bastante precisión a la realidad: que la temperatura máxima más probable en un día cualquiera del verano de 2016 en Murcia ha sido de 32.67°C, que es muy poco probable (probabilidad de 0.0568) de que la temperatura máxima en un día cualquiera del verano de 2016 en Murcia haya sido menor de 30°C o que es casi imposible (probabilidad de 7.15×10^{-12}) que la temperatura máxima en un día cualquiera del verano de 2016 en Murcia haya sido menor de 25°C.

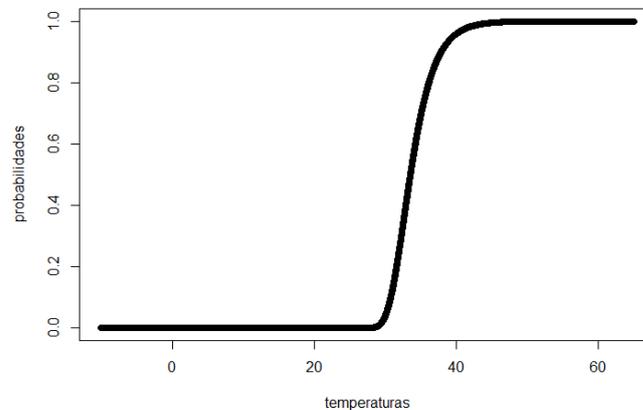


Figura 3: Función de distribución $G_0(x)$

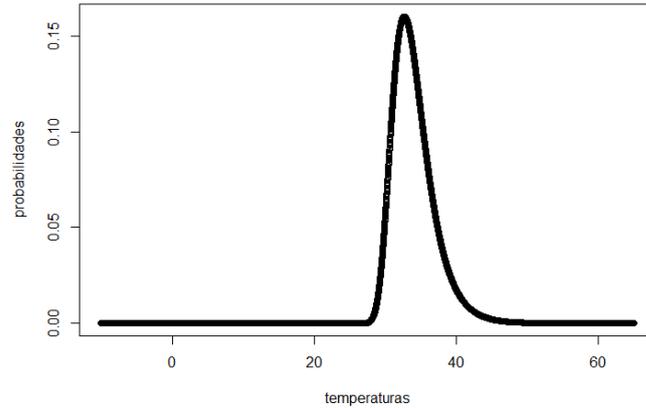


Figura 4: Función de densidad de probabilidad $g_0(x)$

Así, el período de retorno en este caso nos da información como que una temperatura máxima de 30°C se dará, en media, una vez cada 1.0425 días o una temperatura máxima de 40°C se dará, en media, una vez cada 24.7473 días.

Está claro que desde un punto de vista estadístico las temperaturas máximas recogidas en el verano de un solo año no es una información muy representativa para un verano cualquiera, y por tanto, para tratar de predecir, por ejemplo, si hará o no mucho calor el próximo verano, pero es un ejemplo que nos permite hacernos una idea sobre la modelización de máximos y un buen punto de partida.

4.3. Caudales mínimos

En el segundo caso analizado los datos han sido tomados de la página web del Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente y corresponden al caudal medio (medido en m^3/s) de cada mes entre los años 1994 y 2004 del río Segura en su paso por el municipio murciano de Abarán. El número de datos no es 120, como cabía esperar, sino que es de solo 96 porque falta información en algunos meses. La información puede ser vista en la figura 5.

19.94	11.05	11.26	17.02	19.70	15.33	17.18	23.70
9.50	13.23	17.49	12.78	11.67	15.75	22.69	11.24
15.49	14.40	15.76	12.92	14.87	23.48	20.26	11.08
5.98	20.05	24.91	11.04	16.04	23.87	17.69	13.95
11.13	15.61	23.22	18.06	20.79	18.59	26.22	17.90
12.95	29.31	33.87	30.59	24.17	25.12	29.35	25.43
32.17	19.70	31.69	24.05	28.49	25.49	38.71	28.79
28.43	11.09	16.89	34.54	30.12	30.71	41.77	29.52
23.20	22.78	35.28	31.57	24.84	30.72	34.88	48.93
17.50	31.75	52.40	19.72	47.10	24.59	43.07	25.21
29.20	35.69	37.05	28.01	48.56	30.99	39.49	26.21
23.00	31.64	29.95	29.24	38.48	27.60	34.24	22.65

Figura 5: Caudales medios del río Segura entre 1994 y 2004

Procediendo como en el caso anterior, nos hacemos una idea inicial del comportamiento de los datos que tenemos mediante un primer análisis:

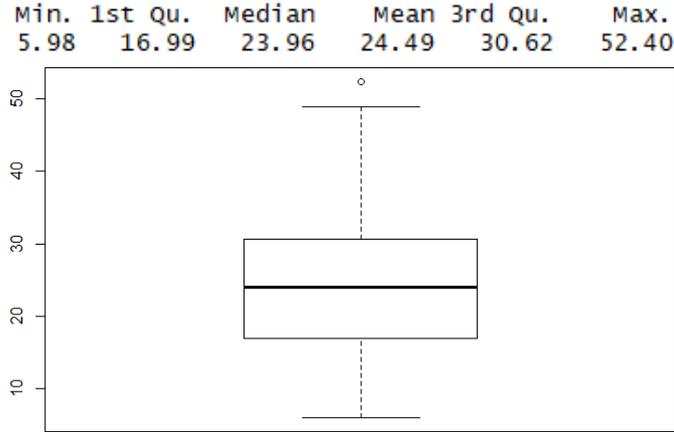


Figura 6: Gráfico de cajas de los caudales

En este caso se van a obtener los estimadores respectivos a la función de distribución del estadístico ordenado mínimo. Para ello se van a seguir los pasos indicados en la introducción del capítulo 3: se cambia el signo de los datos que se quieren estudiar, se obtienen los estimadores del estadístico ordenado máximo para estos datos y se le cambia el signo a la estimación de λ .

Utilizando en primer lugar el paquete de R “evd”, con la función “fgev”, obtenemos los siguientes valores:

$$\begin{aligned}\lambda &= -27.0318 \\ \delta &= 10.4309 \\ \gamma &= -0.4708\end{aligned}$$

También en este caso se obtiene un valor del parámetro γ muy cercano a 0. Además, con la función “confint” se obtienen los respectivos intervalos de confianza para un $\alpha = 0.05$, que son:

$$\begin{aligned}\lambda &\in [-29.2682, -24.7954] \\ \delta &\in [8.7755, 12.0864] \\ \gamma &\in [-0.5739, -0.3676]\end{aligned}$$

Aplicando el método iterativo se repite el fenómeno observado en el caso anterior: en caso de no suponer que $\gamma = 0$, a partir de cierto paso la matriz de información de Fisher pasará a tomar valores muy cercanos a 0, lo que hará que su determinante sea prácticamente 0 y en consecuencia, se obtendrá un error que indica que dicha matriz no tiene inversa. De nuevo, utilizando, el método iterativo para la estimación de solo dos parámetros, con 500 conjuntos de índices elementales distintos considerados a la hora de obtener una estimación inicial, se obtiene la siguiente estimación para λ y δ :

$$\begin{aligned}\lambda &= -29.5954 \\ \delta &= 10.6840\end{aligned}$$

En este caso para δ la diferencia no es tan grande, pero para λ el valor obtenido en el segundo caso incluso sale del intervalo de confianza del obtenido con el método de optimización. Pero además $\gamma = 0$ también se queda fuera del respectivo intervalo de confianza, por lo que quizás esta no sea la mejor elección. Así, he decidido considerar $\gamma \neq 0$ y considerar como estimación obtenida por el método iterativo la última

estimación antes de obtener errores al calcular la inversa de la matriz de información de Fisher, siendo los valores obtenidos:

$$\begin{aligned}\lambda &= -28.5689 \\ \delta &= 9.3192 \\ \gamma &= -0.4353\end{aligned}$$

que esta vez sí que cumplen las condiciones de estar dentro de los intervalos de confianza obtenidos antes. Ahora, si finalmente se considera como estimación definitiva para λ , δ y γ la media de estas dos estimaciones:

$$\lambda = -27.8004, \delta = 9.8751 \text{ y } \gamma = -0.4531$$

Finalmente, como hemos dicho que pretendemos obtener la función de distribución de estadístico ordenado mínimo, las estimaciones definitivas son:

$$\lambda = 27.8004, \delta = 9.8751 \text{ y } \gamma = -0.4531$$

por lo que la función de distribución del caudal mínimo del río Segura en su paso por Abarrán viene dada por:

$$L_{-0.4531}(x) = 1 - \exp\left(-\left(1 + 0.4531 \frac{x-27.8004}{9.8751}\right)^{\frac{1}{0.4531}}\right)$$

con dominio $[\lambda + \frac{\delta}{\gamma}, +\infty) = [6.0059, +\infty)$ y función de densidad de probabilidad:

$$l_{-0.4531}(x) = \frac{1}{9.8751} \left(1 + 0.4531 \frac{x-27.8004}{9.8751}\right)^{\frac{1-0.4531}{0.4531}} \exp\left(-\left(1 + 0.4531 \frac{x-27.8004}{9.8751}\right)^{\frac{1}{0.4531}}\right)$$

con el mismo dominio que $L_{-0.4531}$. Sus representaciones pueden ser vistas en las figuras 7 y 8 respectivamente.

Un concepto importante en hidrología es la **abundancia cortical** que, en general, es el porcentaje medio de un elemento en la corteza terrestre y que en este contexto mide la cantidad de agua que lleva un río en un punto concreto, también conocido como **módulo**. El **módulo relativo** está dado por la fórmula $M_r = \frac{M}{S}$, donde M es el caudal medio dado en l/s en cierto período de tiempo y S es la superficie de la cuenca donde se ha medido el caudal, dada en km^2 . Se considera que si el módulo relativo es inferior a $5 l/s/km^2$ hay escasez, si está entre 5 y $15 l/s/km^2$ son los valores medios y por encima de $15 l/s/km^2$ son valores elevados.

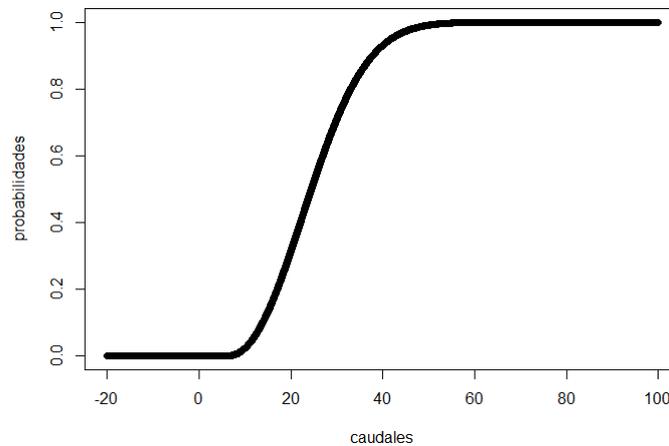


Figura 7: Función de distribución $L_{-0.4531}(x)$

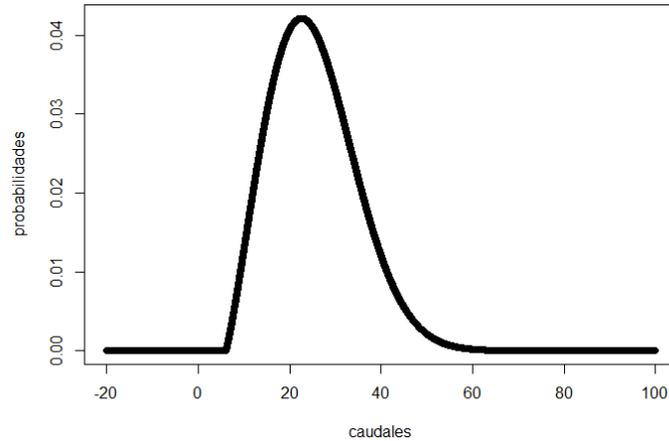


Figura 8: Función de densidad de probabilidad $l_{-0.4531}(x)$

Así, sabiendo que la superficie del tramo del río de donde se han tomado estos datos es de 14936 km^2 , dato obtenido también de la página web del Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente, de la distribución obtenida podemos extraer datos como que el caudal mínimo más probable en un año es de $22.59 \text{ m}^3/\text{s}$ o 22590 l/s , lo que nos da un caudal mínimo más probable de 1.5125 l/s/km^2 , que indica escasez de agua. También nos indica que es casi imposible que el caudal mínimo tenga un valor medio, pues para ello se necesitaría un caudal medio mayor o igual a $74.68 \text{ m}^3/\text{s}$, algo que ocurrirá con una probabilidad de 3.4×10^{-6} .

En este caso el período de retorno nos dice cosas como que lo más normal es tener escasez de agua en este recorrido del río Segura, pues el hecho de tener un módulo relativo inferior a 5 es prácticamente algo que ocurrirá para el caudal medio de cada mes, obteniendo un período de retorno de 1.000003 o también nos dice que se va a tener un módulo relativo inferior a 1, lo que puede ser considerado un déficit extremo de agua, una vez cada 7.6763 meses.

En conclusión, podemos ver que la modelización obtenida nos indica temperaturas máximas elevadas en un verano de Murcia o caudales mínimos mensuales muy reducidos en un tramo del río Segura que transcurre por la Región de Murcia. Son muy conocidos los fenómenos de calor y sequedad dados por las condiciones climáticas de la Región de Murcia, por lo que se puede afirmar que la modelización obtenida al aplicar las distribuciones asintóticas de los estadísticos ordenados nos proporciona información veraz y correspondiente a la realidad.

ANEXO

Se presenta a continuación el código que he utilizado en R para los métodos que he programado. En primer lugar se tiene el código del método que está descrito en la sección 3.3 y que se utiliza para obtener una estimación inicial. Aunque se difiere el caso $\gamma = 0$ del caso $\gamma \neq 0$, en este método los dos casos se tratan en un mismo programa:

```
#Esta funcion es utilizada para comprobar que un triplete  
#generado no forma parte de los ya utilizados.
```

```
comprobarFila<-function(fila , matriz){  
  if(is.matrix(matriz)){  
    n<-dim(matriz)[1] #n. de filas  
    for (i in c(1:n)){  
      if(prod(fila==matriz[i,])==1){  
        return(FALSE)  
      }  
    }  
    return(TRUE)  
  }else{  
    if(prod(fila==matriz)==1){  
      return(FALSE)  
    }  
    return(TRUE)  
  }  
}  
  
#Funciones auxiliares  
pr<-function(r,n){  
  return((r-0.35)/n)  
}  
  
cr<-function(r,n){  
  return(-log(pr(r,n)))  
}  
  
ark<-function(r,k,n){  
  return(cr(r,n)/cr(k,n))  
}  
  
dijk<-function(triplete , info){  
  return((info[triplete[2]] - info[triplete[3]])/  
          (info[triplete[1]] - info[triplete[3]]))  
}
```

```

f<-function(x, triplete ,m, info){
  return((1-ark(triplete [2], triplete [3],m)^-x)/
    (1-ark(triplete [1], triplete [3],m)^-x)-dijk(triplete , info))
}

delta<-function(gamma, triplete ,m, info){
  return(gamma*(info [ triplete [2]] - info [ triplete [1]])/
    (cr(triplete [2],m)^(-gamma)-cr(triplete [1],m)^-gamma))
}

delta0<-function(pareja ,m, info){
  return((info [ pareja [1]] - info [ pareja [2]])/
    (log(cr(pareja [2],m))-log(cr(pareja [1],m))))
}

lambda<-function(gamma, delta , triplete ,m, info){
  return(info [ triplete [1]] -(delta*(cr(triplete [1],m)^
    -gamma-1)/gamma))
}

lambda0<-function(delta , pareja ,m, info){
  return(info [ pareja [1]] + delta*log(cr(pareja [1],m)))
}

#Funcion iterativa que calcula la raiz si gamma no es 0.
encontrarRaiz<-function(x0,y0, triplete ,m, info){
  if (f(x0, triplete ,m, info)*f(y0, triplete ,m, info)<0){
    n=(x0+y0)/2
    if (abs(f(n, triplete ,m, info))<0.00001){
      return(n)
    }
    if (f(x0, triplete ,m, info)*f(n, triplete ,m, info)<0){
      a=x0
      b=n
      return(encontrarRaiz(a,b, triplete ,m, info))
    }else {
      if (f(n, triplete ,m, info)*f(y0, triplete ,m, info)<0){
        a=n
        b=y0
        return(encontrarRaiz(a,b, triplete ,m, info))
      }else {
        return('No')
      }
    }
  }
}

```

```

}else{
  if (f(x0, triplete ,m, info)*f(y0, triplete ,m, info)>0){
    return ('No')
  }else{
    if (abs(f(x0, triplete ,m, info))<0.00000001){
      return(x0)
    }
    if (abs(f(y0, triplete ,m, info))<0.00000001){
      return(y0)
    }
  }
}
}
}

#Funcion principal en el caso gamma != 0.
raiz<-function(){
  #data<-as.numeric(unlist(Temperaturas_max))
  data<-as.numeric(unlist(CaudalesSeguraMin))
  data<-sort(data)
  lon<-length(data)
  ndeTripletes<-500
  tripletes<-sort(sample(lon,3))

  if (dijk(tripletes ,data)<ark(tripletes [2], tripletes [3], lon)/
      ark(tripletes [1], tripletes [3], lon)){
    #a<-log(-dijk(tripletes ,data))/log(ark(tripletes [2],
      tripletes [1], lon))

    #b<- -0.0001
    a<--10
    b<-11
    gammas<-encontrarRaiz(a,b, tripletes , lon ,data)
  }else{
    #a<-0.0001
    #b<-log(1-dijk(tripletes ,data))/log(ark(tripletes [2],
      tripletes [3], lon))

    a<--10
    b<-11
    gammas<-encontrarRaiz(a,b, tripletes , lon ,data)
  }

  deltas<-delta(gammas, tripletes , lon ,data)

  lambdas<-lambda(gammas, deltas , tripletes , lon ,data)

  #estimCoh<-lambdas[1]-deltas [1]/gammas[1]

```

```

contador<-2
a<-10
b<-11

while (ndeTripletes >1){
  triplete<-sort(sample(lon,3))
  if (comprobarFila(triplete, tripletes) && encontrarRaiz
      (a,b, triplete, lon, data)!='No'){
    gammas<-cbind(gammas,
                  encontrarRaiz(a,b, triplete, lon, data))
    deltas<-cbind(deltas, delta(gammas[contador],
                                triplete, lon, data))
    lambdas<-cbind(lambdas, lambda(gammas[contador],
                                   deltas[contador], triplete, lon, data))
    #estimCoh<-cbind(estimCoh, lambdas[length(lambdas)]-
                    #deltas[length(deltas)]/gammas[length(gammas)])
    tripletes<-rbind(tripletes, triplete)
    contador<-contador+1
    ndeTripletes<-ndeTripletes-1
  }
}

gammaDef<-median(gammas)
deltaDef<-median(deltas)
#estimCohDef<-median(estimCoh)
#lambdaDef<-estimCohDef+deltaDef/gammaDef
lambdaDef<-median(lambdas)
estim<-c(lambdaDef, deltaDef, gammaDef)
print('(lambda, delta, gamma)=')
print(estim)
return(estim)
}

#Funcion principal en el caso gamma = 0.
raiz0<-function(){
  #data<-as.numeric(unlist(Temperaturas_max))
  data<-as.numeric(unlist(CaudalesSeguraMin))
  data<-sort(data)
  lon<-length(data)
  ndeParejas<-500
  parejas<-sort(sample(lon,2))

  deltas<-delta0(parejas, lon, data)
  lambdas<-lambda0(deltas, parejas, lon, data)

  contador<-2

```

```

while (ndeParejas > 1){
  pareja<-sort(sample(lon,2))
  if (comprobarFila(pareja , parejas )){
    deltas<-cbind(deltas , delta0(pareja , lon , data))
    lambdas<-cbind(lambdas , lambda0(deltas [ contador ] , pareja ,
                                     lon , data))

    parejas<-rbind(parejas , pareja)
    contador<-contador+1
    ndeParejas<-ndeParejas-1
  }
}

deltaDef<-median(deltas)
lambdaDef<-median(lambdas)
estim<-c(lambdaDef , deltaDef)
return(estim)
}

```

Por otra parte, se tiene el método iterativo descrito en la sección 3.2. Aquí sí que se han utilizados dos programas distintos para los casos $\gamma = 0$ y $\gamma \neq 0$. Así, el código del programa utilizado en el caso $\gamma = 0$ es:

```

#Funciones auxiliares.
gradiente<-function(estim , data){
  lon<-length(data)
  lambdaDers<-c()
  deltaDers<-c()
  gammaDers<-c()

  for (i in c(1:lon)){
    const<-1+estim[3]-exp(-log(1+estim[3]*(data[i]-estim[1])/
                                     estim[2]))/estim[3])

    lambdaDer<-const/(estim[2]+estim[3]*(data[i]-estim[1]))
    lambdaDers<-c(lambdaDers , lambdaDer)
    deltaDer<- -1/estim[2]+(data[i]-estim[1])*const/
               (estim[2]^2+estim[2]-estim[3]*(data[i]-estim[1]))
    deltaDers<-c(deltaDers , deltaDer)
    gammaDer<- -log(1+estim[3]*(data[i]-estim[1])/estim[2])/
               estim[3]+const*(log(1+estim[3]*(data[i]-estim[1])/
               estim[2]))/(estim[2]^2-(data[i]-estim[1])/(estim[3]*
               estim[2]+(estim[3]^2)*(data[i]-estim[1])))
    gammaDers<-c(gammaDers , gammaDer)
  }
}

```

```

derLambda<-sum(lambdaDers)
derDelta<-sum(deltaDers)
derGamma<-sum(gammaDers)

y<-c(derLambda,derDelta,derGamma)
M<-matrix(y,1,3,T)
vec<-t(M)

return(vec)
}

inversaFisherInfo<-function(estim,data){
n<-length(data)
p<-(1+estim[3])^2 * gamma(1+2*estim[3])
q<-gamma(2+estim[3])*(digamma(1+estim[3])+(1+estim[3])/

eulerMascheroni<- -digamma(1)

m11<-n*p/(estim[2]^2)
m22<-n*(1-2*gamma(2+estim[3])+p)/(estim[2]*estim[3])^2
m33<-n*(pi*pi/6+(1-eulerMascheroni+1/estim[3])^2-2*q/
estim[3]+p/estim[3]^2)/estim[3]^2
m12<-n*(p-gamma(2+estim[3]))/(estim[3]*estim[2]^2)
m21<-m12
m13<-n*(q-p/estim[3])/(estim[2]*estim[3])
m31<-m13
m23<-n*(1-eulerMascheroni+(1-gamma(2+estim[3]))/
estim[3]-q+p/estim[3])/(estim[2]*estim[3]^2)
m32<-m23

x<-c(m11,m12,m13,m21,m22,m23,m31,m32,m33)
I<-matrix(x,3,3,T)

if(abs(det(I))>0.00001){
inverseI<-solve(I)
M=inverseI
}else{
print('La matriz no tiene inversa')
M=0
}
#print(M)
return(M)
}

```

```

#Metodo que calcula el error absoluto entre dos estimaciones.
errorAbsEstim<-function(estim1,estim2){
  n<-length(estim1)
  m<-length(estim2)
  if(n==m){
    dists<-c()
    for(i in c(1:n)){
      x<-(estim1[i]-estim2[i])^2
      dists<-c(dists,x)
    }
    return(sqrt(sum(dists)))
  }else{
    print('Las dimensiones de los vectores no coinciden.')
    return(0)
  }
}

#Este es el calculo de cada paso del metodo iterativo.
metodoF<-function(estim,data){
  invI<-inversaFisherInfo(estim,data)
  grad<-gradiente(estim,data)

  nEstim<-t(estim)+invI %* %grad
  return(t(nEstim))
}

#Funcion principal del programa.
metodoiterativo<-function(){
  #data<-as.numeric(unlist(Temperaturas_max))
  data<-as.numeric(unlist(CaudalesSeguraMin))
  data<-sort(data)
  vectorEstim<-raiz()
  estim0<-matrix(vectorEstim,1,3,T)
  print(estim0)
  estim1<-metodoF(estim0,data)
  print(estim1)
  estim2<-metodoF(estim1,data)
  while(errorAbsEstim(estim0,estim1)>0.00001){
    estim0<-estim1
    estim1<-metodoF(estim0,data)
  }
  print('La estimacion es:')
  print(estim1)
}

```

Y finalmente el método iterativo para el caso $\gamma = 0$:

```
#Funciones auxiliares.
gradiente0<-function(estim , data){
  lon<-length(data)
  deltaDers<-c()
  lambdaDers<-c()

  for (i in c(1:lon)){
    const<-1-exp(-((data[i]-estim[1])/estim[2]))

    lambdaDer<-const/estim[2]
    lambdaDers<-c(lambdaDers , lambdaDer)

    deltaDer<- -1/estim[2]+const*(data[i]-estim[1])/estim[2]^2
    deltaDers<-c(deltaDers , deltaDer)
  }

  derLambda<-sum(lambdaDers)
  derDelta<-sum(deltaDers)

  y<-c(derLambda , derDelta)
  M<-matrix(y , 1 , 2 , T)
  vec<-t(M)

  return(vec)
}

inversaFisherInfo0<-function(estim , data){
  n<-length(data)
  eulerMascheroni<- -digamma(1)

  m11<-n/estim[2]^2
  m22<-n*(6-12*eulerMascheroni+6*eulerMascheroni^2+n^2)/

  m12<-n*(eulerMascheroni-1)/estim[2]^2
  m21<-m12

  x<-c(m11 , m12 , m21 , m22)
  I<-matrix(x , 2 , 2 , T)

  if(abs(det(I))>0.000001){
    inverseI<-solve(I)
    M<-inverseI
  }else{
    print('La matriz no tiene inversa')
    M=0
  }
}
```

```

    }
    return (M)
}

#Funcion que calcula el error absoluto entre dos estimaciones.
errorAbsEstim0<-function(estim1,estim2){
  n<-length(estim1)
  m<-length(estim2)
  if(n==m){
    dists<-c()
    for(i in c(1:n)){
      x<-(estim1[i]-estim2[i])^2
      dists<-c(dists,x)
    }
    return(sqrt(sum(dists)))
  }else{
    print('Las dimensiones de los vectores no coinciden.')
    return(0)
  }
}
}
#Este es el calculo que se hace en cada paso del metodo
#iterativo.
metodoF0<-function(estim,data){
  invI<-inversaFisherInfo0(estim,data)
  grad<-gradiente0(estim,data)

  nEstim<-t(estim)+invI%*%grad
  return(t(nEstim))
}
#Funcion principal del programa.
mainFunction0<-function(){
  #data<-as.numeric(unlist(Temperaturas_max))
  data<-as.numeric(unlist(CaudalesSeguraMin))
  data<-sort(data)
  vectorEstim<-raiz0()
  estim0<-matrix(vectorEstim,1,2,T)
  print(estim0)
  estim1<-metodoF0(estim0,data)
  print(estim1)
  while(errorAbsEstim0(estim0,estim1)>0.00001){
    estim0<-estim1
    estim1<-metodoF0(estim0,data)
  }
  print('La estimacion es:')
  print(estim1)
}
}

```

Referencias

- [1] Bovier, Anton (n.d.). Extreme values of random processes, Lecture Notes. 9-12, Recuperado en 03/2017 de https://wt.iam.uni-bonn.de/fileadmin/WT/Inhalt/people/Anton_Bovier/lecture-notes/extreme.pdf
- [2] Bücher, Axel y Segers, Johan (2016). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. 5-6 y 34-35, Ruhr-Universität Bochum, Recuperado en 05/2017 de <https://arxiv.org/pdf/1511.07613.pdf>
- [3] Castillo, Enrique, Hadi, Ali S., Balakrishnan, Narayanaswamy y Sarabia, Jose M. (2004). *"Extreme Value and Related Models with Applications in Engineering and Science"*, Wiley Series in Probability and Statistics.
- [4] Castillo Gutiérrez, Sonia y Lozano Aguilera, Emilio D. (2007). Q-Q Plot Normal. Los puntos de posición gráfica. *Revista Electrónica*, 3-5, Universidad de Jaén, Recuperado en 05/2017 de <http://revistaselectronicas.ujaen.es/index.php/ininv/article/view/259/241>
- [5] David, Herbert A. y Edwards, Anthony W. F (2001). *"Annotated Readings in the History of Statistics"* Springer-Verlag, New York.
- [6] David, Herbert A. y Nagaraja, Haikady N. (2003). *"Order Statistics, 3rd Edition"*, Wiley.
- [7] Gnedenko, Boris V. (1988). *"Theory of Probability, Sixth Edition" (I. Ushakov, trad.)*, Gordon and Breach science publishers, India.
- [8] Gumbel, Emil J. (2004). *"Statistics of Extremes"*, Dover Publications, Inc, New York.
- [9] Gut, Allan (2013). *"Probability: A Graduate Course"*, Springer-Verlag, New York.
- [10] de Haan, Laurens (1970). *"On regular variation and its application to the weak convergence of sample extremes"*, Mathematical Centre Tracts, Amsterdam.
- [11] de Haan, Laurens y Ferreira, Ana (2006). *"Extreme Value Theory: an Introduction"*, Springer-Verlag, New York.
- [12] Kotz, Samuel y Nadarajah, Saralees (2000), *"Extreme Value Distributions: Theory and Applications"*, Imperial College Press, Londres.
- [13] Maximenko, Egor y Moctezuma Salazar, Mario A. (n.d.). Función cuantil asociada a una medida de probabilidad, Recuperado en 04/2017 de http://esfm.egormaximenko.com/prob/quantile_function_es.pdf

- [14] Prescott, Phil y Walden, Andrew T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, Oxford University Press, **67**, Recuperado en 04/2017 de http://www.jstor.org/stable/2335152?seq=1#page_scan_tab_contents
- [15] Resnick, Sidney I. (1987). "*Extreme Values, Regular Variation and Point Processes*", Springer-Verlag, New York.
- [16] Rider, Paul R. (1962). The Negative Binomial Distribution and the Incomplete Beta Function. *The American Mathematical Monthly*, **69**, Recuperado en 04/2017 de https://www.jstor.org/stable/2312952?seq=2#page_scan_tab_contents
- [17] UAH, University of Alabama in Huntsville (n.d.). Order Statistics, Recuperado en 02/2017 de <http://www.math.uah.edu/stat/sample/OrderStatistics.html>.
- [18] Weiping, Zhang (2012). Properties of MLE: consistency, asymptotic normality. Fisher information. *University of Science and Technology of China*, Recuperado en 05/2017 de <http://staff.ustc.edu.cn/~zwp/teach/Math-Stat/MLE-consistent-asynormal.pdf>