



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura

Estimación semiparamétrica del umbral en el análisis de valores
extremos

Juan D. González

Directores: Dra. Daniela Rodríguez y Dra. Mariela Sued

Agosto de 2011

A Laura, Paloma y Salvador

Índice general

1. Introducción	2
2. Fundamentos probabilísticos de la teoría de valores extremos	4
2.1. Motivación	4
2.2. Consideraciones generales y necesidad de reescalamiento	9
2.3. Problema límite de extremos	11
2.3.1. Enfoque histórico	20
2.3.2. Dominios de atracción	21
2.3.3. Una aplicación del Teorema de Fisher, Tippet y Gnedenko	30
3. Función de excesos	31
3.1. Distribuciones límites para la función de excesos	32
3.2. Selección del umbral	33
3.2.1. Mean Residual Plot	34
3.2.2. Estabilización del parámetro de escala y forma	35
3.3. Selección del umbral: nueva propuesta	35
3.3.1. El modelo	35
3.3.2. La propuesta de estimación	38
3.4. Resultados numéricos	39
3.4.1. Modelo uniforme - exponencial	39
3.4.2. Modelo normal - Pareto	44
A. Estimación no paramétrica de la densidad basada en núcleos	50

Agradecimientos

En primer lugar quiero agradecer a mis directoras, Mariela y Daniela, por permitirme trabajar con ellas en el desarrollo de esta tesis. Por darme algo parecido a un tesoro: un problema bien planteado, dentro de una hermosa teoría, y una primer implementación numérica para empezar a “ver” la naturaleza del problema. Por su criterioso apoyo para guiarme con extremada precisión allí donde las cosas se ponían oscuras. Por ayudarme con la teoría, con la programación y con la redacción de esta tesis con total compromiso y dedicación.

En segundo lugar quiero agradecer a Silvia Blanc, por ofrecerme el tiempo y el espacio necesario para el desarrollo de este trabajo, su apoyo, tanto logístico como emocional, fue fundamental para la culminación de esta tesis.

A mis compañeros de trabajo, los “acústicos”, por aguantar mi autismo crónico durante la escritura de la tesis, por los almuerzos y los desayunos compartidos.

A mis compañeros de la facultad, los de fútbol, los del conurbano sur, los de los primeros años, los de los últimos y los que estuvieron siempre.

A mis colegas del CBC Montes de Oca que conocí durante el paso del tiempo.

A mi familia, mis padres y mis hermanos, por bancarme (en todo sentido) cuando los pasos en la carrera no eran tan seguros.

A Pablo Groisman, sus observaciones y comentarios contribuyeron de manera significativa a la claridad de exposición de este trabajo.

Capítulo 1

Introducción

En la teoría de valores extremos el interés principal se encuentra en los valores más bajos o más altos de la variable bajo estudio, es decir, el interés está en los eventos asociados a la cola de la distribución. Por ejemplo, en oceanografía es necesario estudiar el comportamiento de corrientes marinas extremas, en estadística ambiental es necesario analizar niveles altos de ozono en determinada región, en climatología es necesario conocer el comportamiento de velocidades extremas de huracanes o valores extremos de temperatura, etc.

Un enfoque al estudio de valores extremos es a partir del análisis de excedentes sobre umbrales. Es decir, a partir de qué valor, por ejemplo, una temperatura puede considerarse extrema. Por este motivo, es importante poder dar estimadores consistentes para los umbrales. Existen diferentes métodos heurísticos y gráficos que proponen estimadores para los umbrales. Elegir adecuadamente el umbral no es una tarea fácil, si el umbral elegido es bastante alto esto implicará que sólo algunas pocas observaciones se utilicen para estimar la cola de la distribución, aumentando así la varianza de los estimadores. Diferentes autores han estudiado la influencia de la elección del umbral en la estimación de los parámetros de la cola de la distribución (Smith (1987), Frigessi, et.al. (2003), Coles y Tawn (1996), Coles y Powell (1996), Davison y Smith (1990) y Coles y Tawn (1994)).

La teoría desarrollada para el estudio de estos conceptos está íntimamente relacionada con las leyes que rigen el comportamiento asintótico del máximo de n variables aleatorias, apropiadamente reescalado.

En esta tesis desarrollaremos una propuesta de estimación del umbral a través de un abordaje semiparamétrico. Para ello asumiremos que existe un umbral a partir del cual la distribución excedente pertenece a la familia Pareto generalizada. Es decir el modelo propuesto asume que la distribución por debajo del umbral pertenece a una familia con densidades no especificadas mientras que la distribución por encima del umbral pertenece a la familia Pareto Generalizada. Para cada valor de u calcularemos una pseudo-verosimilitud indexada en el umbral, que combina la estimación no paramétrica de la densidad para las observaciones menores a u , con la estimación de máxima verosimilitud

para las observaciones excedentes al umbral. Estimaremos el parámetro de interés como el menor valor a partir del cual se maximiza la pseudo-verisimilitud.

Esta tesis está conformada de la siguiente manera. En el Capítulo 2 estudiaremos la distribución asintótica del máximo, discutiremos la necesidad de rescalamiento, es decir la corrección apropiada respecto a la posición y escala, y caracterizaremos las posibles distribuciones límite. Además presentaremos algunos resultados relativos a los dominios de atracción de las posibles distribuciones límite.

En el Capítulo 3 daremos una introducción a la teoría de excesos presentando el marco teórico necesario para el desarrollo de nuestra propuesta de estimación del umbral. Por otro lado, resumiremos los métodos mencionados en la literatura (ver Coles (2001)) para la selección del umbral. Finalmente, presentamos nuestra propuesta y desarrollamos un estudio de simulación que permite evaluar el comportamiento de la misma.

Respecto de la bibliografía utilizada, los resultados de la sección 2.3 se basaron principalmente en los libros de De Haan y Ferreira (2006), y Teugels *et al.*(2004). La bibliografía no mencionada a lo largo del texto fue consultada para la elaboración de esta tesis.

Capítulo 2

Fundamentos probabilísticos de la teoría de valores extremos

2.1. Motivación

Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas con distribución F , que asumiremos uniforme en el intervalo $[0, 1]$. Queremos inferir el comportamiento de la variable suma S_n

$$S_n = \sum_{i=1}^n X_i.$$

Para esto, supongamos que tenemos m muestras aleatorias, cada una de ellas con cien valores, podemos generar una nueva muestra aleatoria de la variable S_n , es decir $S_n^1, S_n^2, \dots, S_n^m$ como indica el siguiente esquema:

$$\begin{array}{ccccccc} X_1^1 & \dots & X_{100}^1 & \longrightarrow & S_{20}^1, & S_{40}^1, & S_{100}^1 \\ X_1^2 & \dots & X_{100}^2 & \longrightarrow & S_{20}^2, & S_{40}^2, & S_{100}^2 \\ \vdots & & \vdots & & \vdots & \vdots & \vdots \\ X_1^m & \dots & X_{100}^m & \longrightarrow & S_{20}^m, & S_{40}^m, & S_{100}^m. \end{array}$$

Donde S_n^k indica que realizamos la suma de los primeros n valores de la k -ésima muestra, es decir $S_n^k = \sum_{l=1}^n X_l^k$. Si observamos las columnas derechas de este esquema, tenemos que el conjunto $\{S_n^k\}_{k=1}^m$, es una muestra aleatoria con n fijo ($n = 20, 40, 100$). Por ejemplo, para $n = 20$, tenemos una m.a de S_{20} , a saber, $\{S_{20}^1, S_{20}^2, \dots, S_{20}^m\}$ es dicha muestra aleatoria. De ésta forma, podemos conseguir un histograma de S_n , con n fijo, y una estimación de su densidad por núcleos¹ como en la Figura 2.1.

¹En el apéndice se puede encontrar un resumen del estimador no paramétrico de la densidad basado en núcleos.

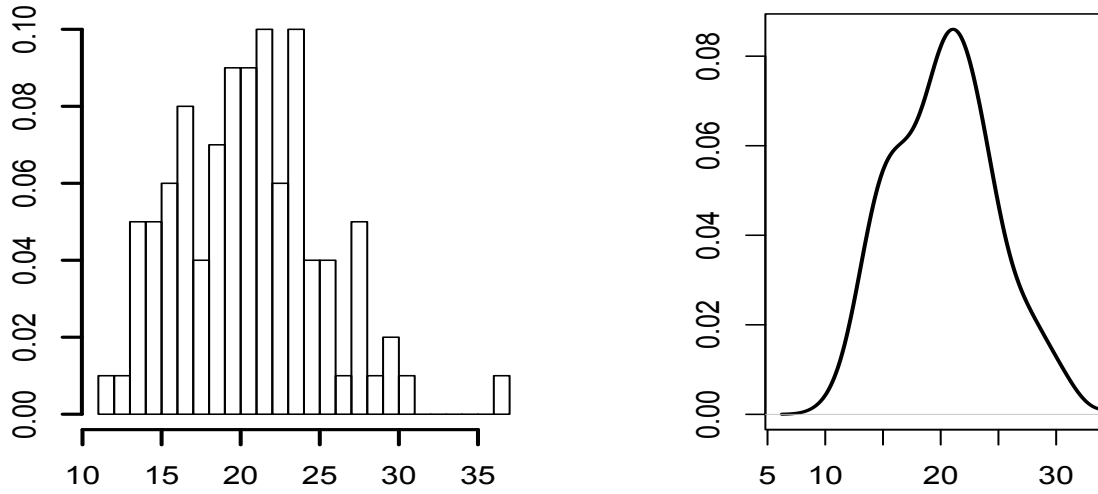


Figura 2.1: Histograma y densidad estimada por núcleos de la variable S_{20} .

Se realizó una simulación de este procedimiento para los distintos valores de n , con $n = 20, 40, 100$, los resultados se muestran en la Figura 2.2, donde observamos que a medida que n crece los estimadores de S_n se asemejan a campanas, cuya amplitud σ_n y posición central μ_n es creciente con n . Esto sugiere que si quisieramos hallar el comportamiento asintótico de S_n nos bastará conocer como es μ_n y σ_n y realizar la transformación

$$Z_n = \frac{S_n - \mu_n}{\sigma_n}.$$

El “Teorema Central del Límite”(TCL) establece cómo elegir éstos valores y cuál es la curva de densidad asintótica que resulta de aplicar esta transformación. En efecto, recordando que nuestra muestra es $U(0, 1)$ con $E(X) = \mu = \frac{1}{2}$ y $Var(X) = \sigma^2 = \frac{1}{12}$, tenemos que $\mu_n = n\mu$ y $\sigma_n = \sigma\sqrt{n}$. En la Figura 2.3 se observa el efecto de esta normalización, y la buena aproximación de la distribución Normal a las variables Z_n . El TCL puede entenderse como una manera de caracterizar el comportamiento asintótico de la distribución de la variable aleatoria S_n mediante un cambio de escala y posición.

Así como resulta de interés la variable suma S_n podemos interesarnos en la variable máximo M_n definida por:

$$M_n = \max\{X_1, X_2, X_3, \dots, X_n\}.$$

Para estudiar el comportamiento de M_n , empezaremos haciendo una simulación para variables i.i.d $\sim \mathcal{E}(1)$, con el mismo esquema utilizado anteriormente. Es decir, tenemos

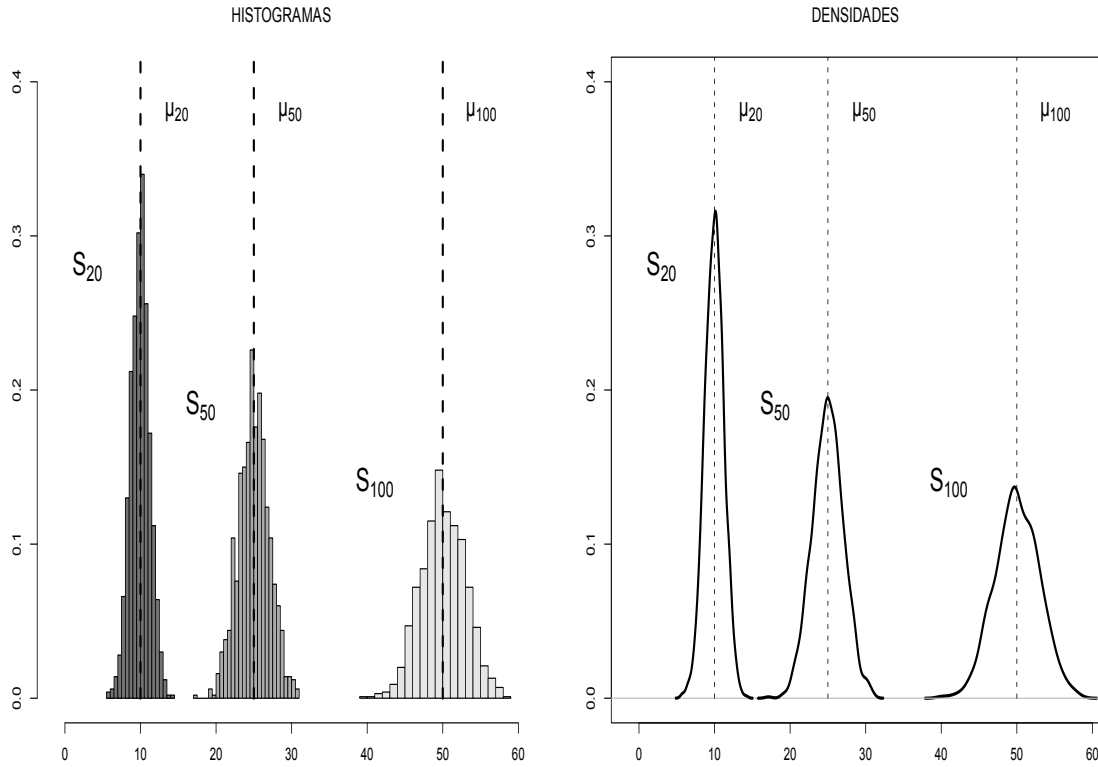


Figura 2.2: Histogramas y densidades estimadas por núcleos de las variables S_{20} , S_{50} y S_{100} .

m muestras de tamaño 1000 y vamos tomando $M_n^k = \max\{X_1^k, X_2^k, X_3^k, \dots, X_n^k\}$,

$$\begin{array}{rcccl}
 X_1^1 & \dots & X_{1000}^1 & \longrightarrow & M_{150}^1, & M_{600}^1, & M_{1000}^1 \\
 X_1^2 & \dots & X_{1000}^2 & \longrightarrow & M_{150}^2, & M_{600}^2, & M_{1000}^2 \\
 \vdots & & \vdots & & \vdots & \vdots & \vdots \\
 X_1^m & \dots & X_{1000}^m & \longrightarrow & M_{150}^m, & M_{600}^m, & M_{1000}^m.
 \end{array}$$

Calculando estimaciones de densidad para distintos tamaños de muestra, análogos a los exhibidos en la Figura 2.2. Obtenemos el gráfico a) de la Figura 2.4.

La Figura 2.4, hace plausible que las variables M_n , tengan forma definida mediante reescalamiento, al igual que las variables S_n . Es decir, continuando con la analogía de S_n , quisieramos averiguar si existirán constantes $a_n > 0$, de escala, y $b_n \in \mathbb{R}$, de posición, tales que la transformación de las variables M_n^* tengan una distribución límite conocida, a saber:

$$M_n^* = \frac{M_n - b_n}{a_n}.$$

La Teoría de los Valores Extremos (TVE) estudia este tipo de problemas. Los primeros

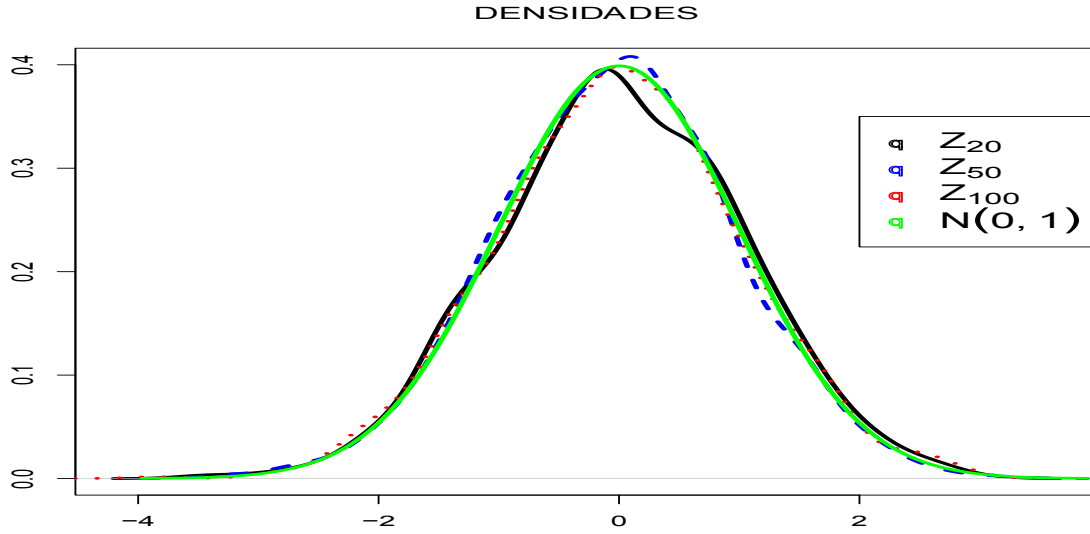


Figura 2.3: Densidades estimadas por núcleos de las variables Z_{20} , Z_{50} , y Z_{100} .

autores en abordar este tema fueron Fisher y Tippet (1928). Gumbel (1958) formalizó y extendió sus ideas. Actualmente esta teoría está en continuo desarrollo y mantiene un alto interés. La cantidad de aplicaciones vinculadas a conocer el comportamiento del máximo de un proceso aleatorio, tiene como consecuencia que la TVE interese no sólo a la comunidad matemática, sino también a especialistas de las más diversas disciplinas. Para adelantar el tipo de resultado que se busca, consideremos el siguiente ejemplo conocido.

Lema 2.1.1. Sean $X_i \sim \mathcal{E}(1)$ i.i.d's definimos la variable aleatoria M_n^*

$$M_n^* = \frac{M_n - b_n}{a_n}$$

con $a_n = 1$ y $b_n = \ln(n)$ entonces,

$$M_n^* \xrightarrow{\mathcal{D}} W$$

donde W es una variable aleatoria cuya densidad es

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

Demostración Lema 2.1.1. Llamemos $F_{M_n^*}$ a la función de distribución de la variable M_n^* y $F(x) = 1 - e^{-x}$ a la función de distribución correspondiente a $\mathcal{E}(1)$. Dado $z \in \mathbb{R}$ hallemos $F_{M_n^*}(z)$. Claramente

$$F_{M_n^*}(z) = P(M_n^* \leq z) = P(M_n \leq za_n + b_n) = P(M_n \leq z + \ln(n)).$$

Como se verá en (2.2.2), $F_{M_n}(x) = (F(x))^n$ luego,

$$F_{M_n^*}(z) = (F(z + \ln(n)))^n = (1 - e^{-(z + \ln(n))})^n.$$

Reescribiendo la parte derecha de la igualdad tenemos que

$$F_{M_n^*}(z) = \left(1 - \frac{e^{-z}}{n}\right)^n \xrightarrow{n \rightarrow \infty} \exp(-e^{-z}),$$

por lo tanto

$$M_n^* \xrightarrow{\mathcal{D}} W$$

con $F_W(z) = \exp(-e^{-z})$ y cuya función densidad de W está dada por $f_W(z) = \exp\{-z - e^{-z}\}$, que es lo que se quería ver. \square

El lema recién probado, sugiere que para valores grandes de n , la distribución de M_n^* podrá ser aproximada por la distribución de W . La Figura 2.4 da cuenta del reescalamiento y de la efectividad de este ajuste para varios valores de n .

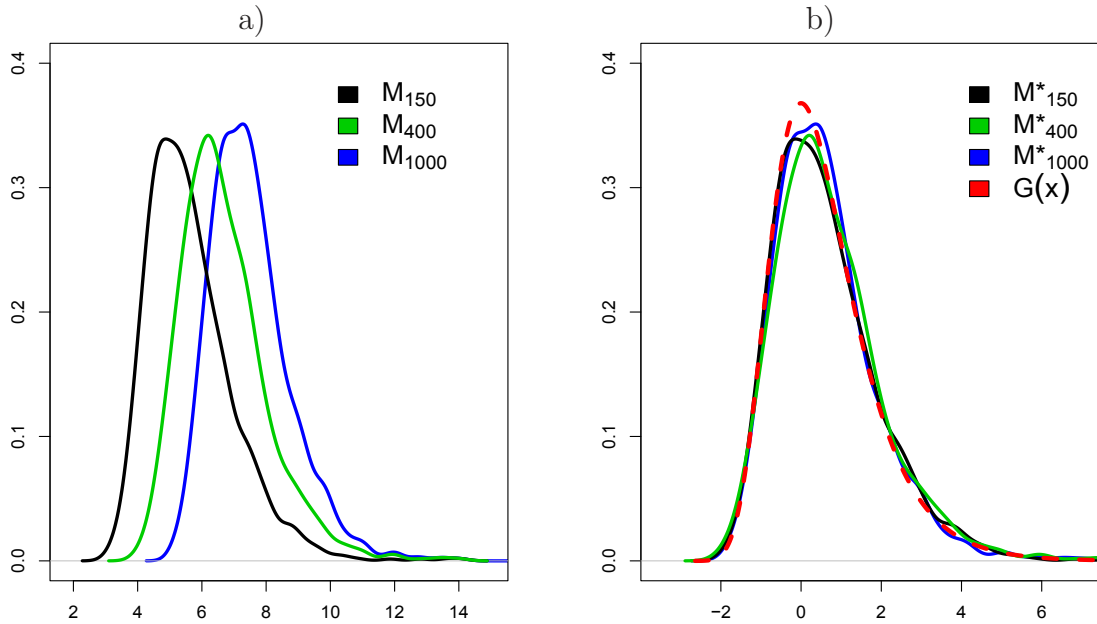


Figura 2.4: a) Densidades estimadas de M_{150} , M_{400} , y M_{1000} b) Densidades estimadas de M_{150}^* , M_{400}^* , y M_{1000}^* y su comparación con $g(x) = \exp\{-x - e^{-x}\}$.

En este caso el conocimiento de la distribución subyacente ($X_i \sim \mathcal{E}(1)$), nos dice exactamente que constantes a_n y b_n tomar, y hacia que curva convergen las variables M_n^* con esta normalización. Afortunadamente, bajo hipótesis razonables, la TVE es lo suficientemente general como para prescindir del conocimiento exacto de la distribución subyacente, proveyendo resultados de convergencia en distribución a familias de variables aleatorias con distribución conocida. Esta es, quizá, su mayor fortaleza.

2.2. Consideraciones generales y necesidad de reescalamiento

Sean X_1, X_2, \dots , variables aleatorias independientes e idénticamente distribuídas (i.i.d's) con distribución F : $X_i \sim F$. Denotaremos con M_n a $M_n = \max\{X_1, \dots, X_n\}$ al máximo de las primeras n variables aleatorias. Nos interesa describir el comportamiento de M_n con $n \rightarrow \infty$, para ello necesitaremos la siguiente definición.

Definición 2.2.1. Sea F una función de distribución acumulada. Se define el punto derecho final x^* como $x^* = \inf\{x : F(x) = 1\}$, donde se adopta la convención $\inf \emptyset = \infty$

Es importante notar que, cuando $x^* \in \mathbb{R}$, tenemos que $P(X_i \leq x^*) = 1$. De hecho, para todo $x > x^*$ se satisface que $F(x) = 1$ y siendo F continúa a derecha, concluimos que $F(x^*) = 1$. En tal caso, podemos también garantizar que $P(M_n \leq x^*) = 1$.

Para poder estudiar la distribución de M_n notemos que

$$\{M_n \leq x\} = \bigcap_{i=1}^n \{X_i \leq x\}. \quad (2.2.1)$$

Bajo las hipótesis mencionadas, de independencia e idéntica distribución, al tomar probabilidad a ambos miembros en la ecuación (2.2.1) se obtiene:

$$\begin{aligned} P(\{M_n \leq x\}) &= P\left(\bigcap_{i=1}^n \{X_i \leq x\}\right) = \prod_{i=1}^n P(\{X_i \leq x\}) \\ &= \prod_{i=1}^n F_{X_i}(x) = \prod_{i=1}^n F(x) = (F(x))^n. \end{aligned}$$

Por lo tanto, la función de distribución de M_n es

$$P(\{M_n \leq x\}) = (F(x))^n, \quad (2.2.2)$$

luego $F_{M_n}(x) = (F(x))^n$, de modo que para cada x fijo, bastará con analizar el límite de $F(x)^n$. Observando que $F(x) \in [0, 1]$, trivialmente ocurre que

$$\lim_{n \rightarrow \infty} F_{M_n}(x) = \begin{cases} 0 & \text{si } F(x) < 1 \\ 1 & \text{si } F(x) = 1. \end{cases}$$

Siguiendo esta idea, puede verse que, cuando $x^* \in \mathbb{R}$, $M_n \xrightarrow{\mathcal{D}} x^*$, es decir, la convergencia de M_n en distribución es a una función de distribución degenerada concentrada en x^* . Vamos a demostrar que, en realidad, M_n converge casi todo punto a x^* , sea o no finito.

Lema 2.2.1. Sean $(X_i)_{i \geq 1}$ i.i.d., con distribución F : $X_i \sim F$. Sea x^* el punto derecho de F , introducido en la Definición 2.2.1. Tenemos entonces que

$$P(M_n \rightarrow x^*) = 1.$$

Demostración Lema 2.2.1. Notemos que $P(X_i \leq x^*) = 1$, y por consiguiente, $P(M_n \leq x^*) = 1$, para todo $n \geq 1$. Para hacer la demostración, debemos diferenciar el caso $x^* \in \mathbb{R}$ del caso $x^* = \infty$. Cuando $x^* \in \mathbb{R}$, tenemos que

$$\{M_n \rightarrow x^*\} = \cap_M \cup_{n_0} \cap_{n \geq n_0} \{|M_n - x^*| < 1/M\}.$$

Luego, basta verificar que el complemento de $\{M_n \rightarrow x^*\}$ tiene probabilidad cero. Es decir queremos ver que cualquiera sea M

$$P((\cap_{n_0} \cup_{n \geq n_0} \{|M_n - x^*| \geq 1/M\})) = 0. \quad (2.2.3)$$

Denotemos con $A_n = \{|M_n - x^*| \geq 1/M\}$, la ecuación 2.2.3 es equivalente a que

$$P(\limsup A_n) = 0.$$

Por Borel-Cantelli, bastará probar que $\sum_{n=1}^{\infty} P(A_n) < \infty$ que se desprende del hecho que

$$P(A_n) = P(M_n \leq x^* - 1/M) = (F(x^* - 1/M))^n.$$

Por definición de x^* , cualquiera sea M , $F(x^* - 1/M) = r < 1$, luego

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} r^n < \infty.$$

Si $x^* = +\infty$, tenemos que $\{M_n \rightarrow +\infty\} = \cap_K \cup_{n_0} \cap_{n \geq n_0} \{M_n > K\}$, luego veremos que este conjunto tiene probabilidad 1. Tomando complemento y haciendo un razonamiento análogo al anterior, alcanzará con estudiar el conjunto

$$\cap_{n_0} \cup_{n > n_0} \{M_n \leq K\}.$$

Otra vez, por Borel-Canteli, llamamos $A'_n = \{M_n \leq K\}$ y queremos ver que $\sum_{n=1}^{\infty} P(A'_n) = \sum_{n=1}^{\infty} (F(K))^n < \infty$. Donde al ser K un número fijo y $x^* = \infty$, $F(K) = r < 1$ pudiendo escribir $\sum_{n=1}^{\infty} P(A'_n) = \sum_{n=1}^{\infty} r^n < \infty$. \square

El Lema precedente muestra que para hacer aproximaciones asintóticas de M_n a una distribución G no degenerada, es decir que no esté concentrada en único punto, es necesario realizar algún tipo de transformación. En vistas de resolver esto se plantea el estadístico M_n^* :

$$M_n^* = \frac{M_n - b_n}{a_n}$$

donde $b_n \in \mathbb{R}$ y $a_n > 0$, son sucesiones hipotéticas, que hacen que

$$M_n^* \xrightarrow{\mathcal{D}} W, \quad (2.2.4)$$

con $F_W(x) = G(x)$ y $G(x)$ una función de distribución no degenerada. La condición (2.2.4), es equivalente a que existan a_n , b_n y una función de distribución no degenerada G , tal que para cada punto x de continuidad de G se satisfaga

$$\lim_{n \rightarrow \infty} (F(a_n x + b_n))^n = G(x). \quad (2.2.5)$$

En adelante, estudiaremos los siguiente dos puntos:

- i) Suponiendo la existencia de sucesiones a_n y b_n adecuadas, hallaremos cuáles son las posibles distribuciones G a las que el máximo reescalado puede converger. Este problema es nombrado por la literatura como el “*Problema Límite de Extremos*” y fue resuelto por Fisher y Tippett (1928), y Gnedenko (1943), y más tarde revivido y perfeccionado por De Haan (1970).
- ii) Dada una posible distribución límite G , determinar cuáles son las condiciones que se deben pedir sobre la distribución subyacente F de las variables X_i para que existan a_n, b_n de manera que valga la condición (2.2.5). En ese caso se dice que F está en el dominio de atracción de G , y se denota $F \in \mathcal{D}(G)$.

2.3. Problema límite de extremos

En esta Sección se caracterizará a las funciones de distribución G que cumplan con la condición (2.2.5). Es decir, bajo el supuesto de que el límite (2.2.5) existe para ciertas sucesiones a_n y b_n , se tratará de saber cómo es la función G .

Sea x un punto de continuidad de G , y además $0 < G(x) < 1$. Tomando logaritmo natural en ambos lados de la igualdad en (2.2.5) tenemos que

$$\lim_{n \rightarrow \infty} n \ln F(a_n x + b_n) = \ln G(x). \quad (2.3.1)$$

Por otro lado, sabemos que $\lim_{n \rightarrow \infty} F(a_n x + b_n) = 1$, de lo contrario, como $\{F(a_n x + b_n)\}_{n=1}^{\infty} \subseteq [0, 1]$ compacto, existe $a \in \mathbb{R}$ tal que, $F(a_{n_k} x + b_{n_k}) \xrightarrow[k \rightarrow \infty]{} a < 1$ y entonces, $(F(a_{n_k} x + b_{n_k}))^{n_k} \xrightarrow[k \rightarrow \infty]{} 0 = G(x)$ que contradice la hipótesis de que $0 < G(x) < 1$.

Asímismo, usando que $\lim_{w \rightarrow 1} \frac{\ln w}{w - 1} = 1$, haciendo la sustitución $w = F(a_n x + b_n)$ y operando algebraicamente en la ecuación (2.3.1) se obtiene

$$\lim_{n \rightarrow \infty} n(1 - F(a_n x + b_n)) = -\ln G(x). \quad (2.3.2)$$

Invirtiendo la ecuación anterior queda

$$\lim_{n \rightarrow \infty} \frac{1}{n(1 - F(a_n x + b_n))} = -\frac{1}{\ln G(x)}. \quad (2.3.3)$$

Este límite puede ser escrito en términos de la función inversa a izquierda. Para eso, debemos recordar su definición y mencionar algunas de sus principales propiedades.

Definición 2.3.1 (Inversa a Izquierda). Dada una función f no decreciente, se define la inversa a izquierda de f en y , $f^{\leftarrow}(y)$, como aquella que cumple

$$f^{\leftarrow}(y) = \inf\{x : f(x) \geq y\}.$$

Observemos que en el caso de que $f = F$ sea una función de distribución, $F^{\leftarrow}(p) = Q(p)$, donde Q es usualmente conocido como el p -cuantil asociado a la distribución F . A continuación enunciaremos algunas propiedades de la función inversa a izquierda.

Propiedades 2.3.1.

- I) f^{\leftarrow} es no decreciente.
- II) Si f es continua, con $f(x) = y$ $f(f^{\leftarrow}(y)) = y$.

La siguiente propiedad, cuya demostración puede verse en De Haan y Ferreira (2006), establece que la inversa a izquierda, preserva convergencia, en el siguiente sentido:

Propiedades 2.3.2. Sean f_n una sucesión de funciones no decrecientes convergiendo puntualmente a una función no decreciente g , para cada punto de continuidad de g , i.e;

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

Sean f_n^{\leftarrow} y g^{\leftarrow} sus respectivas inversas a izquierda, entonces

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

para cada x , punto de continuidad de g^{\leftarrow} .

Sean

$$f_n(x) = \frac{1}{n(1 - F(a_n x + b_n))}, \quad g(x) = -\frac{1}{\ln G(x)},$$

aplicando los resultados presentados en 2.3.2 a la ecuación (2.3.3), obtenemos que para todo x punto de continuidad de g^{\leftarrow}

$$\lim_{n \rightarrow \infty} f_n^{\leftarrow}(x) = g^{\leftarrow}(x). \quad (2.3.4)$$

El próximo objetivo es operar este límite por medio de los “Cuantiles de Cola”, un concepto clave para el desarrollo de la Teoría de Valores Extremos. El mismo está relacionado al comportamiento de los valores cercanos al punto derecho final x^* de la distribución subyacente F .

Definición 2.3.2 (Cuantil de cola). Dada una función de distribución F , y un valor $x \in (1, \infty)$ se define el cuantil de cola $U(x) = U_F(x)$ como

$$U(x) = U_F(x) = F^{\leftarrow} \left(1 - \frac{1}{x} \right). \quad (2.3.5)$$

Notemos que a medida que x crece, la probabilidad de encontrar un valor más alto que $U(x)$ decrece. Por ejemplo si $x = 10^n$ entonces $P(X > U(x)) = 10^{-n}$. Así, los valores que toma $U(x)$, dan una idea de como se comporta la distribución para aquellos valores altos que tienen baja probabilidad de ser superados. De esta forma $U(x)$ indica cual es el comportamiento de la cola. En la Figura 2.5 se observa, a modo de comparación, los valores de U para dos distribuciones opuestas, la distribución Normal y la distribución Cauchy. Se ve que U crece muy lentamente, de manera logarítmica, para el caso Normal (Cola Liviana), mientras que lo hace rápidamente, de manera lineal, para el caso donde la distribución es Cauchy (Colas Pesadas).

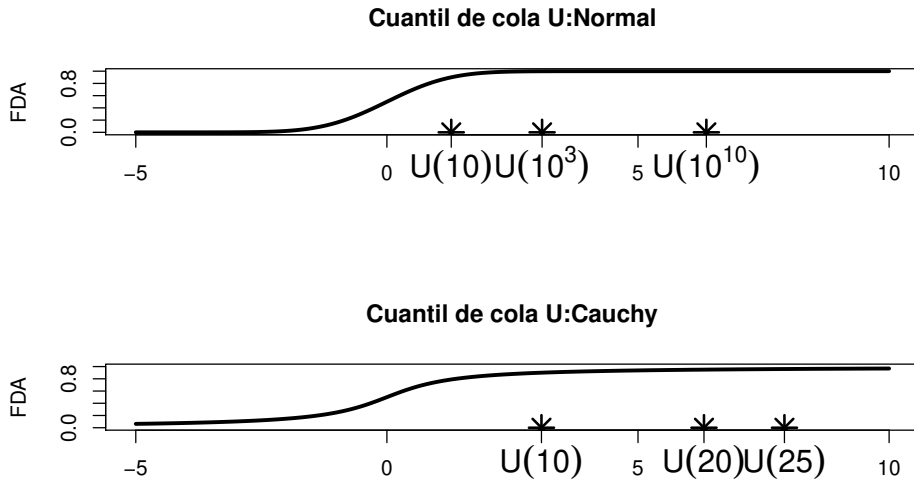


Figura 2.5: Función de distribución acumulada y comportamiento de Cuantiles de Cola.

A continuación probaremos que $f_n^{\leftarrow}(x)$ la función inversa de $f_n(x) = \frac{1}{n(1 - F(a_n x + b_n))}$ que definimos anteriormente, puede expresarse en términos de la función U . Para $x > 0$, tenemos que

$$f_n^{\leftarrow}(x) = \inf\{y : f_n(y) \geq x\} = \inf\left\{y : \frac{1}{n(1 - F(a_n y + b_n))} \geq x\right\}.$$

Llamemos A al conjunto $\{y : \frac{1}{n(1-F(a_n y + b_n))} \geq x\}$. Utilizaremos la siguiente propiedad, sea B un conjunto acotado inferiormente $B \subset \mathbb{R}$, y valores $\alpha > 0$ y $\beta \in \mathbb{R}$

$$\inf\{B\alpha + \beta\} = \inf\{B\}\alpha + \beta.$$

Luego tomando $\alpha = a_n > 0$ y $\beta = b_n$ tenemos que $f_n^\leftarrow(x) = \frac{\inf\{Aa_n + b_n\} - b_n}{a_n}$. Por último observemos que

$$\begin{aligned} \inf\{Aa_n + b_n\} &= \inf\left\{\left\{y : \frac{1}{n(1-F(a_n y + b_n))} \geq x\right\} a_n + b_n\right\} \\ \inf\{Aa_n + b_n\} &= \inf\left\{\left\{y : \frac{1}{(1-F(a_n y + b_n))} \geq nx\right\} a_n + b_n\right\} \\ &= \inf\{a_n y + b_n : \frac{1}{(1-F(a_n y + b_n))} \geq nx\} \\ &= \inf\{w : \frac{1}{(1-F(w))} \geq nx\} \\ &= \inf\{w : F(w) \geq 1 - \frac{1}{nx}\} \\ &= U(nx) \end{aligned}$$

de esta forma obtenemos la relación buscada que vincula a la función inversa generalizada de f_n con el cuantil de cola de la distribución F ,

$$f_n^\leftarrow(x) = \frac{U(nx) - b_n}{a_n}. \quad (2.3.6)$$

A su vez, operando de manera análoga, el término derecho de (2.3.4), puede expresarse como

$$g^\leftarrow(x) = G^\leftarrow(e^{-\frac{1}{x}}) \quad (2.3.7)$$

para $x > 0$. Observemos que $g^\leftarrow(x)$ es no decreciente y continua a derecha. Esta propiedad será muy útil más adelante.

Definición 2.3.3. Dada una función de distribución G , definimos $D = D_G : \mathbb{R}_{>0} \rightarrow \mathbb{R}$ mediante la fórmula

$$D(x) = g^\leftarrow(x) = G^\leftarrow(e^{-\frac{1}{x}}) \quad (2.3.8)$$

siendo G^\leftarrow la función inversa a izquierda de la distribución G .

En consecuencia, de (2.3.7), (2.3.6) y (2.3.4) hemos probado que para cada x punto de continuidad de la función D ,

$$\lim_{n \rightarrow \infty} \frac{U(nx) - b_n}{a_n} = D(x). \quad (2.3.9)$$

El desarrollo hecho hasta ahora puede resumirse en el siguiente teorema que relaciona el límite (2.2.5) puesto en términos de F , con el límite (2.3.9) en términos de U .

Teorema 2.3.1. *Sean $a_n > 0$ y b_n suseciones de números reales y G una función de distribución acumulada no degenerada. Las siguientes afirmaciones son equivalentes:*

I) *Para cada x , punto de continuidad de G con $0 < G(x) < 1$*

$$\lim_{n \rightarrow \infty} (F(a_n x + b_n))^n = G(x) \quad (2.3.10)$$

II) *Para cada $x > 0$ punto de continuidad de D ,*

$$\lim_{n \rightarrow \infty} \frac{U(nx) - b_n}{a_n} = D(x) \quad (2.3.11)$$

En vistas de caracterizar la función G , necesitamos modificar levemente este teorema de manera de tener funciones de parámetro continuo $a(t)$ y $b(t)$ en lugar de sucesiones a_n y b_n .

Lema 2.3.1. *Sean $a_n > 0$ y b_n suseciones de números reales y G una función de distribución acumulada no degenerada. Las siguientes afirmaciones son equivalentes:*

I) *Para cada x punto de continuidad de G , con $0 < G(x) < 1$*

$$\lim_{n \rightarrow \infty} (F(a_n x + b_n))^n = G(x) \quad (2.3.12)$$

II) *Para cada x punto de continuidad de G , con $0 < G(x) < 1$*

$$\lim_{t \rightarrow \infty} t(1 - F(a(t)x + b(t))) = -\ln G(x) \quad (2.3.13)$$

siendo $a(t) := a_{[t]}$ y $b(t) := b_{[t]}$ (donde $[t]$ es la parte entera de t).

III) *Para cada $x > 0$ punto de continuidad de D*

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x) \quad (2.3.14)$$

con $a(t) := a_{[t]}$ y $b(t) := b_{[t]}$

Demostración Lema 2.3.1. $I) \Rightarrow II)$: Previamente se probó la equivalencia entre $I)$ y la ecuación (2.3.2), luego bastará ver la equivalencia entre (2.3.2) y $II)$. La misma se desprende en forma sencilla de la siguiente relación que permite cambiar la variable discreta n por la continua t

$$\begin{aligned} -\ln G(x) &= \lim_{n \rightarrow \infty} n(1 - F(a_n x + b_n)) = \lim_{t \rightarrow \infty} [t](1 - F(a(t)x + b(t))) \\ &= \lim_{t \rightarrow \infty} \frac{[t]}{t} \cdot t(1 - F(a(t)x + b(t))) = \lim_{t \rightarrow \infty} t(1 - F(a(t)x + b(t))). \end{aligned}$$

$II) \Rightarrow III)$: Ya probamos que $I)$ es equivalente a la ecuación (2.3.9), y el hecho que $I) \iff II)$ implica que bastará con ver que (2.3.9) es equivalente a $III)$. Para esto, dado $x > 0$ y $t > 1$, entonces

$$[t]x \leq tx \leq ([t] + 1)x.$$

Al ser U no decreciente y $a(t)$ positivo sucede que

$$\frac{U([t]x) - b(t)}{a(t)} \leq \frac{U(tx) - b(t)}{a(t)} \leq \frac{U(([t] + 1)x) - b(t)}{a(t)}. \quad (2.3.15)$$

Sea ahora $x' > x$, entonces $x' - x > 0$, luego se puede tomar $[t] > t_0 \geq \max\{\frac{x}{x' - x}, 1\}$. Para este valor de t vale que $[t] > \frac{x}{x' - x}$, que es equivalente a

$$[t]x' > [t]x + x = ([t] + 1)x.$$

Luego

$$\frac{U(([t] + 1)x) - b(t)}{a(t)} \leq \frac{U([t]x') - b(t)}{a(t)}.$$

Combinando este último resultado con (2.3.15) se obtiene que para cualesquiera x y x' puntos de continuidad, con $x' > x$

$$\frac{U([t]x) - b(t)}{a(t)} \leq \frac{U([t]x) - b(t)}{a(t)} \leq \frac{U([t]x') - b(t)}{a(t)}.$$

Tomando límite en esta última expresión y haciendo uso del resultado (2.3.9) tenemos que

$$D(x) = \lim_{t \rightarrow \infty} \frac{U([t]x) - b(t)}{a(t)} \leq \lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} \leq \lim_{t \rightarrow \infty} \frac{U([t]x') - b(t)}{a(t)} = D(x').$$

Como esto vale para todo $x' > x$, tomando $x' = x + \epsilon$, se tiene que $\forall \epsilon > 0$

$$D(x) \leq \lim_{t \rightarrow \infty} \frac{U([t]x) - b(t)}{a(t)} \leq D(x + \epsilon) \quad (2.3.16)$$

Por lo tanto, al ser D continua a derecha se obtiene

$$\lim_{t \rightarrow \infty} \frac{U([t]x) - b(t)}{a(t)} = D(x)$$

que es la condición *III*) y a lo que se quería llegar. \square

A continuación estudiaremos el principal resultado de esta Sección, la caracterización de la función $G(x)$. Para ello, introducimos la familia de valores extremos.

Definición 2.3.4. Consideremos la función

$$G_\gamma(x) = \exp\{-(1 + \gamma x)^{-1/\gamma}\} \quad \text{para } 1 + \gamma x > 0 \quad (2.3.17)$$

para $\gamma \neq 0$, mientras que para $\gamma = 0$, se define $G_0(x) = \exp\{-e^{-x}\}$.

La clase de distribuciones de valores extremos está dada por

$$VE = \{G_\gamma(ax + b) : \gamma \in \mathbb{R}, a > 0, b \in \mathbb{R}\}.$$

El parámetro γ de la ecuación (2.3.17) se denomina *índice de valores extremos* (IVE).

Teorema 2.3.2 (Fisher y Tippet(1928), Gnedenko(1943)). Sean $(X_i)_{i \geq 1}$ i.i.d., con distribución F , supongamos existen a_n y b_n tal que

$$M_n^* = \frac{\max\{X_1, \dots, X_n\} - b_n}{a_n} \xrightarrow{\mathcal{D}} W. \quad (2.3.18)$$

con $W \sim G$. Si G es no degenerada entonces G pertenece a la clase de distribuciones de valores extremos. Es decir, existen γ , a y b tal que

$$G(x) = G_\gamma(ax + b).$$

Demostración Teorema 2.3.2. Bajo el supuesto de que existen sucesiones a_n, b_n satisfaciendo (2.3.18), por el Teorema 2.3.1, para $x > 0$ punto de continuidad de la función D definida en (2.3.8), se tiene que

$$D(x) = \lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)}.$$

Supongamos además que 1 es punto de continuidad de la función D y sea $E(x) := D(x) - D(1)$. $E(x)$ puede ser expresada como el siguiente límite para los puntos de continuidad de D

$$E(x) = \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)}. \quad (2.3.19)$$

Notemos que

$$\frac{U(txy) - U(t)}{a(t)} = \frac{U(txy) - U(ty)}{a(ty)} \frac{a(ty)}{at} + \frac{U(ty) - U(t)}{a(t)}. \quad (2.3.20)$$

Se afirma que $\frac{a(ty)}{a(t)}$ tiene límite cuando $t \rightarrow \infty$, para $y > 0$, punto de continuidad de D . En efecto, siendo G no degenerada, existe $x > 0$ punto de continuidad de D con $E(x) \neq 0$, y xy punto de continuidad de D . De la ecuación (2.3.20), obtenemos que

$$\frac{E(xy) - E(y)}{E(x)} = \lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)}.$$

De ésta forma podemos definir, para y punto de continuidad de D la función $A(y) := \lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)}$. Notemos que, cualesquiera sean los puntos x , e y de continuidad de D , la función A cumple

$$A(x \cdot y) = A(x) \cdot A(y). \quad (2.3.21)$$

En efecto,

$$A(xy) = \lim_{t \rightarrow \infty} \frac{a(xyt)}{a(t)} = \lim_{t \rightarrow \infty} \frac{a(xyt)}{a(ty)} \frac{a(ty)}{a(t)} = \lim_{t \rightarrow \infty} \frac{a(xyt)}{a(ty)} \lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)} = A(x)A(y).$$

Además, puede probarse que si A satisface (2.3.21), necesariamente existe $\gamma \in \mathbb{R}$ cumpliendo

$$A(y) = y^\gamma. \quad (2.3.22)$$

Luego, a partir de la ecuación (2.3.20) tenemos que

$$E(xy) = E(x)y^\gamma + E(y), \quad (2.3.23)$$

y por simetría también tenemos

$$E(xy) = E(y)x^\gamma + E(x).$$

Estudiemos primero el caso $\gamma = 0$. En este caso la ecuación (2.3.23) se simplifica en $E(xy) = E(x) + E(y)$ luego para algún $\lambda > 0$

$$E(x) = \lambda \ln x$$

para x punto de continuidad de E . Luego las propiedades de monotonía de E y la continuidad de $\ln x$ garantizan que la igualdad se satisfaga para todo $x > 0$. Por lo tanto, $D(x) = \lambda \ln(x) + D(1)$, así

$$G^{\leftarrow}(e^{-1/x}) = \lambda \ln(x) + D(1).$$

Luego, tomando $z = \lambda \ln(x) + D(1)$ tenemos $x = e^{az+b}$ y sustituyendo en la ecuación anterior obtenemos que

$$G(z) = \exp\{-e^{(az+b)}\}. \quad (2.3.24)$$

En el caso $\gamma \neq 0$, operando con las ecuaciones obtenidas anteriormente se sigue que

$$\begin{aligned} E(x)y^\gamma + E(y) &= E(y)x^\gamma + E(x) \\ E(x)(y^\gamma - 1) &= E(y)(x^\gamma - 1) \end{aligned}$$

fijando $y = y_0 \neq 1, y_0^\gamma - 1 \neq 0$, y renombrando $K = \frac{E(y_0)}{y_0^\gamma - 1}$, se llega a que

$$E(x) = K(x^\gamma - 1)$$

para todo $x > 0$, usando nuevamente la monotonía de la función E al igual que en el caso $\gamma = 0$. Así, como $E(x) = D(x) - D(1)$ y $D(x) = G^\leftarrow(e^{-1/x})$ tenemos que

$$G^\leftarrow(e^{-1/x})(x) = K(x^\gamma - 1) + D(1)$$

tomando $h(x) = e^{-1/x}$

$$G^\leftarrow(h(x)) = K(x^\gamma - 1) + D(1).$$

Como x es punto de continuidad de $D(\cdot)$, y la función $h(t) = e^{-1/t}$ es continua con inversa continua en $t > 0$, $G^\leftarrow(\cdot)$ es continua en $h(x)$. Luego aplicando G a ambos miembros se obtiene:

$$\begin{aligned} h(x) &= G(K(x^\gamma - 1) + D(1)) \\ e^{-1/x} &= G(K(x^\gamma - 1) + D(1)) \end{aligned}$$

Escribiendo

$$\begin{aligned} K &= \frac{C}{\gamma} \\ e^{-1/x} &= G\left(\frac{C}{\gamma}(x^\gamma - 1) + D(1)\right) \end{aligned}$$

Tomando $y = \frac{1}{\gamma}(x^\gamma - 1)$, se tiene que $x = (\gamma y + 1)^{1/\gamma}$. Luego

$$e^{-\frac{1}{(\gamma y + 1)^{1/\gamma}}} = e^{-(\gamma y + 1)^{-1/\gamma}} = G(Ky + D(1)).$$

Esto último implica que, para $\gamma \neq 0$, existen constantes a y b cumpliendo

$$G_\gamma(ax + b) = \exp\{-(\gamma x + 1)_+^{-1/\gamma}\}. \quad (2.3.25)$$

Los dos casos que se acaban de probar, ($\gamma = 0$ y $\gamma \neq 0$) pueden ser puestos en una sola familia de funciones, a costa de adoptar la convención

$$\left\{(\gamma x + 1)_+^{-1/\gamma}\right\}_{\gamma=0} = e^{-x}.$$

Así, hemos probado que existen constantes γ , a y b

$$G(az + d) = e^{-(\gamma y + 1)_+^{-1/\gamma}}$$

donde $\{A\}_+ = \max\{0, A\}$. \square

En vistas de hacer una caracterización de las funciones para las cuales existe una ley asintótica para el máximo. Enunciaremos un teorema que resume las propiedades vistas en esta sección.

Teorema 2.3.3. *Las siguientes sentencias son equivalentes.*

1. *Existen $a_n > 0$, y $b_n \in \mathbb{R}$ tales que*

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp(-(1 + \gamma x)^{-1/\gamma}) \quad (2.3.26)$$

para todo x tal que $1 + \gamma x > 0$.

2. *Existe una función $a(t) > 0$ tal que para $x > 0$,*

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(x)}{a(t)} = D_\gamma(x) = \frac{x^\gamma - 1}{\gamma} \quad (2.3.27)$$

donde para $\gamma = 0$ y $D_0(x) := \log(x)$.

3. *Existe una función $a(t) > 0$ tal que para $x > 0$,*

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

para todo x tal que $1 + \gamma x > 0$.

Además, (2.3.26) se satisface con $b_n := U(n)$ y $a_n := a(n)$.

2.3.1. Enfoque histórico

La parametrización dada en el Teorema anterior, se debe a Von Mises (1936) y Jenkinson (1955). Historicamente se separó según el signo de γ en tres casos. Cada uno de ellos correspondiente a una clase de distribuciones. Asimismo observemos que según el signo de γ la función G_γ tiene dominio diferente.

- Para $\gamma > 0$ (Tipo II)

La distribución de valores extremos se denomina de *Tipo II*, que es la clase de Frechet y está asociada a la letra Φ . Es decir, si la ley de valores extremos de F converge a una función de distribución G_γ , con $\gamma > 0$, entonces existen valores $a > 0$ y $b \in \mathbb{R}$ que cumplen $G_\gamma(az + b) = \Phi_{1/\gamma}(z)$ donde $\Phi_\alpha(z)$ se define como

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

Mas aún las transformaciones correspondientes entre Φ_α y G_γ , son tomar $\alpha = 1/\gamma$ y luego hacer $G_\gamma(\frac{z-1}{\gamma}) = \Phi_\alpha(z)$.

- Para $\gamma = 0$ (Tipo I)

En este caso, la distribución es de Tipo I y está asociada a la distribución de Gumbel $\Lambda(z)$. En este caso $\Lambda(z) := G_0(z) = \exp(-\exp\{-x\})$.

- Para $\gamma < 0$ (Tipo III)

En este caso, la distribución es de Tipo III y está asociada a la distribución de Weibull $\Psi(z)$.

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

Con $\alpha > 0$ la relación de G_γ con Ψ_α es que $\Psi_\alpha = G_{-1/\alpha}(-\alpha(z-1))$.

2.3.2. Dominios de atracción

En esta sección se establecerán condiciones que determinen cuando una función de distribución F cumple con la ley asintótica del máximo que se trató en la subsección anterior.

Lema 2.3.2. *Supongamos que existen sucesiones $a_n > 0$, b_n , satisfaciendo,*

$$\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{D}} W$$

con $W \sim G_\gamma(cx + d)$ para $c > 0$ y $d \in \mathbb{R}$. Entonces existen $a'_n > 0$, b'_n

$$\frac{M_n - b'_n}{a'_n} \xrightarrow{\mathcal{D}} W'$$

donde $W' \sim G_\gamma(x)$.

Demostración Lema 2.3.2. La demostración se sigue fácilmente utilizando el Lema de Slutsky, notando que

$$c \cdot \frac{M_n - b_n}{a_n} + d \xrightarrow{\mathcal{D}} W' = cW + d.$$

Luego, la función de distribución de la variable aleatoria W' es $G_\gamma(z)$ y $c \cdot \frac{M_n - b_n}{a_n} + d = \frac{M_n - b'_n}{a'_n}$ con $a'_n = \frac{a_n}{c}$ y $b'_n = b_n - a'_n d$. \square

El lema precedente, indica que las sucesiones a_n y b_n no son únicas y están asociadas a los valores de posición d y escala c que eventualmente toma la función G_γ . De esta manera, se puede pensar que las condiciones que hay que pedir a una función de distribución F para que exista una ley límite para el máximo, están dadas, no ya para una familia de distribuciones G que incluye corrimientos de posición y escala, sino para la función G_γ estandarizada con $c = 1$ y $d = 0$.

Definición 2.3.5. Sea F una función de distribución y $\gamma \in \mathbb{R}$, se define que F está en el dominio de atracción de G_γ , si vale que

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp\{-(1 + \gamma x)^{1/\gamma}\}.$$

En este caso, se escribirá $F \in \mathcal{D}(G_\gamma)$.

Con el fin de analizar las hipótesis necesarias sobre la función de distribución F para que se satisfaga la ley asintótica del máximo, separaremos el estudio en tres casos dependiendo del signo de γ . El signo de este parámetro es el primer indicador de las propiedades de la “cola” de la distribución. Si $\gamma > 0$ la función F tiene cola “pesada” y $x^* = \infty$; si $\gamma < 0$, $x^* < \infty$; y si $\gamma = 0$ la cola es “liviana” y x^* puede ser finito o infinito. La mayoría de las funciones de distribución continuas típicamente utilizadas en estadística están en algún dominio de atracción. Por ejemplo, las distribuciones Normal y Exponencial pertenecen a $\mathcal{D}(G_0)$; la distribución uniforme a $\mathcal{D}(G_\gamma)$ con $\gamma < 0$; y la distribución de Cauchy a $\gamma > 0$.

En esta tesis sólo estudiaremos el caso $\gamma > 0$. Las técnicas utilizadas para este caso, pueden readaptarse al caso $\gamma < 0$, pero no para el caso $\gamma = 0$, este último puede consultarse en De Haan y Ferreira (2006). Es oportuno aclarar que existen casos de distribuciones que no están en ningún dominio de atracción, por ejemplo, la distribución discreta de Poisson no pertenece a ningún dominio de atracción. La Tabla 2.1 resume algunas distribuciones junto a los dominios de atracción al que pertenecen.

El caso de Fréchet-Pareto: $\gamma > 0$

Como primer ejemplo se tomará la función de distribución de *Pareto estricta* $Pa(\alpha)$, es decir su función de distribución esta dada por, $F(x) = 1 - x^{-\alpha}$ con $x \geq 1$ donde α se

Distribución	$F(x)$	$F \in D(G_\gamma)$
$N(\mu, \sigma^2)$	$\Phi\left(\frac{x-\mu}{\sigma}\right)$	G_0
$\mathcal{E}(\lambda)$	$F(x) = 1 - e^{-\lambda x}$	G_0
$\beta(a, b)$	$\int_0^x \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1} (1-w)^{b-1} dw$	$G_{-1/a}$
$U(a, b)$	$F(x) = \frac{x-a}{b-a}, x \in [a, b]$	G_{-1}
Cauchy (a, b)	$F(x) = \frac{1}{\pi} \arctan\left(\frac{x-a}{b}\right) + \frac{1}{2}$	G_1
$ T_n $	$F(x) = \int_{-\infty}^x 2 \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} (1 + n^{-1}w^2)^{-(n+1)/2} dw$	$G_{1/n}$

Cuadro 2.1: Distribuciones usuales junto a los dominios de atracción al que pertenecen. Φ es la función de distribución respecto a una variable normal estandar.

denomina el *índice de Pareto*. Para esta distribución tenemos que $U(x) = x^\gamma$. Con $\gamma = \frac{1}{\alpha}$. Entonces planteando la condición sobre los cuantiles de cola, tenemos que

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

Tomando $a(x) = \gamma x^\gamma$ obtenemos que

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

Luego si $F \sim Pa(\alpha)$, entonces $F \in \mathcal{D}(G_\gamma)$. La ecuación anterior vale solo para un caso particular. Pero la cuenta detallada da idea de como relajar las condiciones para extender este resultado a funciones mas generales que aquellas asociadas a $Pa(\alpha)$. Antes de comenzar definiremos las funciones de variación regular.

Definición 2.3.6. (Funciones de variación Regular) Una función $g : A \subseteq \mathbb{R} \rightarrow B \subseteq \mathbb{R}$ es de variación regular de índice α si

$$\lim_{x \rightarrow \infty} \frac{g(xt)}{g(x)} = t^\alpha.$$

Si $\alpha = 0$, entonces g se dice de variación lenta, a las funciones de variación lenta las notaremos con la letra ℓ .

Asintóticamente, si una función g es de variación regular de parámetro α , una amplificación t en su argumento x , tiene como respuesta una amplificación en el valor de la función de orden α , es decir, $g(xt) \approx g(x)t^\alpha$. Por ello puede pensarse que las funciones con $\alpha = 0$, son insensibles a amplificaciones en el argumento pues vale $g(xt) \approx g(x)$. Consecuencia de esto es que si $\alpha = 0$, la función g varíe muy lentamente, y de allí su denominación como función de variación lenta. Un ejemplo de una función de variación regular de orden α es x^α y un ejemplo de función de variación lenta es $\log(x)$.

Proposición 2.3.1. $g(x)$ es de variación regular de parámetro $\alpha \Leftrightarrow g(x) = x^\alpha \ell(x)$, donde $\ell(x)$ es de variación lenta.

Demostración Proposición 2.3.1. (\Rightarrow): $g(x)$ es de variación regular de parámetro α , entonces vale

$$\lim_{x \rightarrow \infty} \frac{g(xt)}{g(x)} = t^\alpha \tag{2.3.28}$$

Luego, definimos $\ell(x) := \frac{g(x)}{x^\alpha}$, entonces valen las siguientes dos afirmaciones,

- $g(x) = x^\alpha \ell(x)$

■ $\ell(x)$ es de variación lenta:

$$\frac{\ell(xt)}{\ell(x)} = \frac{g(xt) x^\alpha}{(xt)^\alpha g(x)} \quad (2.3.29)$$

$$= \frac{1}{t^\alpha} \frac{g(xt)}{g(x)} \quad (2.3.30)$$

$$\lim_{x \rightarrow \infty} \frac{\ell(xt)}{\ell(x)} = \frac{1}{t^\alpha} \lim_{x \rightarrow \infty} \frac{g(xt)}{g(x)} = \frac{1}{t^\alpha} t^\alpha = 1$$

De ésta forma, los dos ítems anteriores, concluyen la implicación en el sentido que se quería probar.

(\Leftarrow): Ahora, suponiendo que $g(x) = x^\alpha \ell(x)$, con $\ell(x)$ de variación lenta, puede verse directamente que g es de variación regular de parámetro α aplicando la definición.

$$\lim_{x \rightarrow \infty} \frac{g(xt)}{g(x)} = \lim_{x \rightarrow \infty} \frac{(tx)^\alpha \ell(tx)}{x^\alpha \ell(x)} = t^\alpha \lim_{x \rightarrow \infty} \frac{\ell(xt)}{\ell(x)} = t^\alpha$$

El siguiente teorema impone condiciones de suficiencia sobre F de manera que $F \in D(G_\gamma)$.

Teorema 2.3.4. *Sea F una función de distribución, tal que*

$$U_F(x) = U(x) = x^\gamma \ell_{U_F}(x) \quad (2.3.31)$$

con $\gamma > 0$ y ℓ_{U_F} una función de variación lenta entonces $F \in D(G_\gamma)$.

Demostración Teorema 2.3.4. Para probar el teorema, basta probar que se satisface la condición dada en (2.3.27) que se desprende directamente del siguiente hecho

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{U(xt) - U(x)}{a(x)} &= \lim_{x \rightarrow \infty} \frac{(xt)^\gamma \ell_U(xt) - x^\gamma \ell_U(x)}{a(x)} \\ &= \lim_{x \rightarrow \infty} \frac{\ell_U(x) x^\gamma}{a(x)} \left(\frac{\ell_U(xt)}{\ell_U(x)} t^\gamma - 1 \right) \end{aligned}$$

Sea $a(x) = \gamma \ell_U(x) x^\gamma$ obtenemos que

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

que es lo que queríamos ver. \square

Observación 2.3.1. *La condición $U(x) = x^\gamma \ell_U(x)$ implica que si $U(x)$ es de variación regular de parámetro $\gamma > 0$ entonces $F \in D(G_\gamma)$.*

En adelante se probará que la condición suficiente (2.3.31) también es necesaria. Con esto tendremos que para $\gamma > 0$ el dominio de atracción $\mathcal{D}(G_\gamma)$ quedará completamente caracterizado por el comportamiento asintótico de las funciones cuantiles de cola U_F de las distribuciones F que están en dicho dominio. Antes de continuar enunciaremos el siguiente resultado que probaremos más adelante y utilizaremos en la prueba de que la condición suficiente también es necesaria.

Lema 2.3.3. *Si $F \in \mathcal{D}(G_\gamma)$, con $\gamma > 0$, entonces*

$$\lim_{t \rightarrow \infty} U(t) = \infty.$$

Observemos que este lema implica que si $\gamma > 0$, entonces $x^* = \infty$. Ya que por las definiciones de U y x^* es fácil ver que

$$U(t) \leq x^*.$$

Teorema 2.3.5. *Sea $F \in \mathcal{D}(G_\gamma)$, con $\gamma > 0$, entonces*

$$U_F(x) = x^\gamma \ell_{U_F}(x)$$

Demostración Teorema 2.3.5. Como $F \in \mathcal{D}(G_\gamma)$ entonces existe $a(t) > 0$ tal que

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

y además en (2.3.22) vimos que

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

Sea ahora z un número arbitrario tal que $z > 1$ para $k \in \mathbb{N}$, se tiene que

$$\frac{U(z^{k+1}) - U(z^k)}{U(z^k) - U(z^{k-1})} = \frac{U(z^{k+1}) - U(z^k)}{a(z^k)} \cdot \left(\frac{U(z^k) - U(z^{k-1})}{a(z^{k-1})} \right)^{-1} \cdot \frac{a(z^k)}{a(z^{k-1})}. \quad (2.3.32)$$

En el término de la derecha, tomando límite en k , reemplazando z^{k+1} por $z \cdot z^k$ y que haciendo la sustitución $t = z^k$ ($k \rightarrow \infty \Rightarrow t \rightarrow \infty$, por ser $z > 1$) tenemos

$$\lim_{k \rightarrow \infty} \frac{U(z \cdot z^k) - U(z^k)}{U(z^k)} = \lim_{t \rightarrow \infty} \frac{U(z \cdot t) - U(t)}{U(t)} = \frac{z^\gamma - 1}{\gamma}.$$

Análogamente $\lim_{k \rightarrow \infty} \frac{a(z^k)}{a(z^{k-1})} = \lim_{k \rightarrow \infty} \frac{a(z^{k-1}z)}{a(z^{k-1})} = z^\gamma$. Luego tomando límite en (2.3.32) obtenemos

$$\lim_{k \rightarrow \infty} \frac{U(z^{k+1}) - U(z^k)}{U(z^k) - U(z^{k-1})} = z^\gamma.$$

Así, para $\varepsilon > 0$ se tiene que existe $k_0(\varepsilon)$ tal que:

$$\left| \frac{(U(z^{k+1}) - U(z^k)) - z^\gamma (U(z^k) - U(z^{k-1}))}{U(z^k) - U(z^{k-1})} \right| < \varepsilon.$$

Entonces,

$$-\varepsilon (U(z^k) - U(z^{k-1})) < (U(z^{k+1}) - U(z^k)) - z^\gamma (U(z^k) - U(z^{k-1})) < \varepsilon (U(z^k) - U(z^{k-1}))$$

Reorganizando los términos

$$(U(z^k) - U(z^{k-1})) z^\gamma (1 - \varepsilon) < U(z^{k+1}) - U(z^k) < (U(z^k) - U(z^{k-1})) z^\gamma (1 + \varepsilon)$$

y sumando término a término cada miembro de la desigualdad desde k_0 hasta N se obtiene

$$(U(z^N) - U(z^{k_0-1})) z^\gamma (1 - \varepsilon) < U(z^{N+1}) - U(z^{k_0}) < (U(z^N) - U(z^{k_0-1})) z^\gamma (1 + \varepsilon).$$

Ahora dividiendo por $U(z^N)$ obtenemos

$$\left(1 - \frac{U(z^{k_0-1})}{U(z^N)}\right) z^\gamma (1 - \varepsilon) < \frac{U(z^{N+1})}{U(z^N)} - \frac{U(z^{k_0})}{U(z^N)} < \left(1 - \frac{U(z^{k_0-1})}{U(z^N)}\right) z^\gamma (1 + \varepsilon). \quad (2.3.33)$$

Como k_0 está fijo, utilizando el resultado del Lema 2.3.3 tenemos $U(z^N) \xrightarrow{N \rightarrow \infty} \infty$ y por

lo tanto $\frac{U(z^{k_0-1})}{U(z^N)} \xrightarrow{N \rightarrow \infty} 0$. De este hecho y tomando límite en N en (2.3.33) se obtiene

$$z^\gamma (1 - \varepsilon) < \lim_{N \rightarrow \infty} \frac{U(z^{N+1})}{U(z^N)} < z^\gamma (1 + \varepsilon)$$

para todo ε . Luego

$$\lim_{N \rightarrow \infty} \frac{U(z^{N+1})}{U(z^N)} = z^\gamma.$$

Un razonamiento inductivo generaliza la última expresión para $l \in \mathbb{N}$

$$\lim_{N \rightarrow \infty} \frac{U(z^{N+l})}{U(z^N)} = z^{l\gamma}. \quad (2.3.34)$$

Ahora, para $x > 1$ definamos el número $n(x) \in \mathbb{N}$ como aquél que cumple

$$z^{n(x)} < x \leq z^{n(x)+1}.$$

Para $t, x > 1$ se tiene que

$$z^{n(t)} z^{n(x)} < tx < z^{n(t)+1} z^{n(x)+1},$$

como U es no decreciente

$$\begin{aligned} U(z^{n(t)} z^{n(x)}) &< U(tx) < U(z^{n(t)+1} z^{n(x)+1}) \\ \frac{U(z^{n(t)} z^{n(x)})}{U(z^{n(t)+1})} &< \frac{U(tx)}{U(t)} < \frac{U(z^{n(t)+1} z^{n(x)+1})}{U(z^{n(t)})}. \end{aligned} \quad (2.3.35)$$

Tomando la parte izquierda de la desigualdad, se tiene que

$$\frac{U(z^{n(t)} z^{n(x)})}{U(z^{n(t)+1})} = \frac{U(z^{n(t)} z^{n(x)})}{U(z^{n(t)})} \frac{U(z^{n(t)})}{U(z^{n(t)+1})} \quad (2.3.36)$$

y aplicando (2.3.34) a (2.3.36)

$$\lim_{t \rightarrow \infty} \frac{U(z^{n(t)} z^{n(x)})}{U(z^{n(t)+1})} = z^{\gamma n(x)} z^{-\gamma} = (z^{n(x)})^\gamma z^{-\gamma} \geq x^\gamma z^{-2\gamma}. \quad (2.3.37)$$

Donde la última desigualdad vale por la definición de $n(x)$

$$z^{n(x)+1} \geq x \Rightarrow z^{n(x)} \geq x/z \Rightarrow (z^{n(x)})^\gamma \geq (x/z)^\gamma \Rightarrow (z^{n(x)})^\gamma z^{-\gamma} \geq x^\gamma z^{-2\gamma}.$$

Análogamente se ve que la parte derecha de (2.3.35) satisface

$$\lim_{t \rightarrow \infty} \frac{U(z^{n(t)+1} z^{n(x)+1})}{U(z^{n(t)})} < x^\gamma z^{2\gamma}. \quad (2.3.38)$$

De esta forma utilizando (2.3.37) en (2.3.35), tomando límite inferior en t tenemos

$$\liminf_t \frac{U(tx)}{U(t)} \geq x^\gamma z^{-2\gamma}$$

y cuando $z \searrow 1$ obtenemos

$$\liminf_t \frac{U(tx)}{U(t)} \geq x^\gamma.$$

De la misma manera usando (2.3.38) en (2.3.35),

$$\limsup_t \frac{U(tx)}{U(t)} \leq x^\gamma z^{2\gamma} \Rightarrow \limsup_t \frac{U(tx)}{U(t)} \leq x^\gamma.$$

Luego hemos probado que

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma. \quad (2.3.39)$$

La ecuación 2.3.39 implica que el cuantil de cola U es de variación regular de parámetro γ , esto según la propiedad 2.3.1 es equivalente a la existencia de una función de variación lenta ℓ_{U_F} cumpliendo

$$U_F(x) = x^\gamma \ell_{U_F}(x) \quad (2.3.40)$$

Que es lo que se quería ver. \square

Demostración Lema 2.3.3. En principio observemos que

$$U(z^j) - U(z^{j-1}) = U(z^{n_0}) - U(z^{n_0-1}) \frac{U(z^{n_0+1}) - U(z^{n_0})}{U(z_0^n) - U(z^{n_0-1})} \cdots \frac{U(z^j) - U(z^{j-1})}{U(z^{j-1}) - U(z^{j-2})}.$$

Es decir,

$$U(z^j) - U(z^{j-1}) = U(z_0^n) - U(z^{n_0-1}) \prod_{k=n_0}^j \frac{U(z^{k+1}) - U(z^k)}{U(z^k) - U(z^{k-1})}. \quad (2.3.41)$$

Luego utilizando (2.3.33) de la demostración del teorema anterior dado $0 < \varepsilon < 1 - z^\gamma$ existe n_0 tal que para todo $k > n_0$

$$\frac{U(z^{k+1}) - U(z^k)}{U(z^k) - U(z^{k-1})} > (1 - \varepsilon)z^\gamma.$$

Por lo tanto en (2.3.41) si $j \geq n_0$

$$\begin{aligned} U(z^j) - U(z^{j-1}) &> U(z_0^n) - U(z^{n_0-1}) \prod_{k=n_0}^j (1 - \varepsilon)z^\gamma \\ &> (U(z_0^n) - U(z^{n_0-1})) ((1 - \varepsilon)z^\gamma)^{j-n_0+1}. \end{aligned}$$

Entonces

$$\begin{aligned} \sum_{j=n_0}^N U(z^j) - U(z^{j-1}) &> (U(z_0^n) - U(z^{n_0-1})) \sum_{j=n_0}^N ((1 - \varepsilon)z^\gamma)^{j-n_0+1} \\ U(z^N) - U(z^{n_0-1}) &> (U(z_0^n) - U(z^{n_0-1})) \sum_{j=n_0}^N ((1 - \varepsilon)z^\gamma)^{j-n_0+1}. \end{aligned}$$

Por la elección de ε efectuada, $(1 - \varepsilon)z^\gamma > 1$, entonces al tomar límite en N , la parte derecha de la última ecuación es ∞ , por lo tanto también es $\lim_{N \rightarrow \infty} U(z^N) = \infty$. Finalmente, el hecho de que U es no decreciente implica que $\lim_{t \rightarrow \infty} U(t) = \infty$. \square

El teorema anterior establece que las funciones de distribución F que están en el dominio de atracción $D(G_\gamma)$, con $\gamma > 0$ son exactamente las mismas funciones cuyos cuantiles de cola son de variación regular de parámetro γ . Intuitivamente la condición $F \in D(G_\gamma)$, es equivalente a que su función U_F no crezca arbitrariamente, sino que este controlada, por una función de tipo x^γ

Es posible, a partir de la caracterización en términos de la función U_F , dar una caracterización en términos de la función F .

Teorema 2.3.6. Sea F una función de distribución acumulada y $\bar{F} = 1 - F$ entonces

$$F \in \mathcal{D}(G_\gamma) \text{ con } \gamma > 0 \text{ si y sólo si } \bar{F}(x) = x^{-1/\gamma} \ell_F(x)$$

donde $\ell_F(x)$ una función de variación lenta.

El teorema anterior es equivalente a pensar que las variables aleatorias que satisfacen la ley del máximo para $\gamma > 0$, son aquellas que tienen una distribución tal que \bar{F} , la cola de la distribución, cae como $1/x^\alpha$, para x suficientemente grande. La demostración de este hecho puede consultarse en Teugels *et al.* (2004).

2.3.3. Una aplicación del Teorema de Fisher, Tippet y Gnedenko

En esta sección mostraremos la importancia del Teorema 2.3.2 estudiado anteriormente a fin de desarrollar un procedimiento de estimación. Consideremos una muestra aleatoria de tamaño N fijo $\{X_1, \dots, X_N\}$ y sea M_N el máximo de las X_i . Estamos interesados en dar una estimación de $P(M_N \leq z)$ asumiendo que F está en $D(G_\gamma)$, es decir, existen sucesiones $a_n > 0$ y $b_n \in \mathbb{R}$ tales que

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq z\right) = G_\gamma(cz + d)$$

para ciertos valores de posición d y escala c .

Este límite brinda una potente idea para estimar $P(M_N \leq z)$, una idea informal para hacerlo es el siguiente hecho,

$$P(M_N \leq z) = P\left(\frac{M_N - b_N}{a_N} \leq \frac{z - b_N}{a_N}\right) \approx G_\gamma\left(c \cdot \left(\frac{z - b_N}{a_N}\right) + d\right).$$

El argumento de la parte derecha de la aproximación, cumple una relación lineal en z , que puede ser escrito como $\sigma_N z + u_N$, que olvidando la dependencia con N podemos escribirlo de la forma $\sigma z + u$. Así tenemos que la parte derecha de la aproximación puede ser expresada como $G_{\gamma, u, \sigma}(z) = G_\gamma(\sigma z + u)$. En definitiva

$$P(M_N \leq z) \approx G_{\gamma, u, \sigma}(z)$$

La función $G_{\gamma, u, \sigma}$ se denomina función de valores extremos generalizada. Y se concluye que M_N puede ser aproximada por una función de valores extremos generalizada. Es importante recalcar que el procedimiento utilizado prescinde del conocimiento de los valores de escala y posición a_N y b_N .

Capítulo 3

Función de excesos

El tiempo de duración de un componente electrónico puede ser representado por una variable aleatoria X con distribución F . Para una duración u , dada en días, interesa saber cuál es la probabilidad de que el componente electrónico dure y días más. Dicha probabilidad puede ser calculada de la siguiente manera

$$P(X \geq u + y | X \geq u) = \frac{1 - F(u + y)}{1 - F(u)}.$$

El término de la derecha está relacionado con la función de distribución de excesos F_u

$$F_u(y) = P(X - u \leq y | X \geq u) = 1 - \frac{1 - F(u + y)}{1 - F(u)},$$

utilizando la notación $\bar{F} = 1 - F$ tenemos,

$$\bar{F}_u(y) = \frac{1 - F(u + y)}{1 - F(u)}.$$

Una propiedad conocida es que $X \sim \mathcal{E}(\lambda)$ si y sólo si, $F_u(x)$ no depende de u , esto significa que $F_u(x) = H(x)$, para alguna función H . Las distribuciones que cumplen con esta propiedad son asociadas al desgaste cero.

En general las funciones de exceso cumplen una ley asintótica, para $u \rightarrow \infty$. La teoría utilizada para obtenerla está extremadamente relacionada con las leyes para el máximo de una muestra aleatoria. Así, resultados de la teoría de valores extremos se utilizan para estimaciones de funciones de excesos y viceversa. En la literatura, los primeros trabajos que dan cuenta de una situación acabada de esta relación son de Pickands (1975) y Balkema y de Haan (1974). Ambos obtuvieron sus resultados independientemente. Pickands trató de estimar el parámetro γ de la distribución de valores extremos, para ello desarrolló las principales propiedades asintóticas de la función de excedencias de una distribución F . Balkema y De Haan estudiaron el problema de caracterizar directamente la función de

excedencias por sobre un umbral, para valores altos de u , desarrollando la relación existente entre la teoría de valores extremos y la función de excedencias.

En el capítulo anterior se probó que el máximo de una muestra converge en distribución a una función en la familia $G_\gamma(x)$, con $\gamma \in \mathbb{R}$, donde G_γ quedó completamente determinada, salvo valores de posición y escala. Análogamente, se verá que \bar{F}_u , con $u \rightarrow \infty$, puede ser aproximada por la función generalizada de Pareto.

3.1. Distribuciones límites para la función de excesos

A continuación definiremos la familia de distribuciones Pareto. Esta familia cumple un papel análogo a la función de valores extremos G_γ , vista en la sección anterior. También presentaremos una serie de resultados que evidencian la analogía mencionada. La demostración de los mismos puede verse en Balkema, De Haan (1974) y Pickands (1975).

Definición 3.1.1. Decimos que una variable aleatoria es Pareto $H_{\sigma,\gamma}$, con $\sigma > 0$ y $\gamma \in \mathbb{R}$ si su función de distribución está dada por

$$H_{\sigma,\gamma}(z) = 1 - \left(1 + \frac{z\gamma}{\sigma}\right)_+^{-1/\gamma},$$

para $\gamma \neq 0$, mientras que para $\gamma = 0$ se define $H_{\sigma,0}(z) = 1 - e^{-\frac{z}{\sigma}}$.

Esta familia de distribuciones satisface las siguientes propiedades.

Propiedades 3.1.1.

1. *Estabilidad.* Si existe u_0 de forma tal que $F_{u_0} = H_{\sigma(u_0),\gamma}$, entonces, para todo $u \geq u_0$ se tiene que $F_u = H_{\sigma(u),\gamma}$ con $\sigma(u) = \sigma(u_0) + \gamma(u - u_0)$.
2. Si $Y \sim H_{\sigma,\gamma}$ con $\gamma < 1$, entonces $E[Y] = \frac{\sigma}{1-\gamma}$.

Teorema 3.1.1.

1. *Balkema, De Haan (1974).* Supongamos que existen funciones $a(u) > 0$, $b(u) \in \mathbb{R}$ y W tal que

$$\lim_{u \rightarrow x^*} F_u(a(u)x + b(u)) = W(x)$$

para todo x tal que $0 < W(x) < 1$. Entonces, existen $\sigma > 0, \gamma \in \mathbb{R}$ para los cuales $W = H_{\sigma,\gamma}$.

2. *Pickands (1975).* Si $F \in D(G_\gamma)$ $\lim_{u \rightarrow x^*} \sup_{0 \leq x < \infty} |F_u(x) - H_{(\sigma(u),\gamma)}(x)| = 0$.

Observación 3.1.1. El punto 1 del teorema anterior indica que la única posibilidad de que haya una función límite para F_u , con $u \rightarrow \infty$ es que W sea Pareto, mientras que el punto 2 da una idea de la relación existente entre la ley asintótica para el máximo y

la ley asintótica para la función distribución de excesos. En primer lugar establece que aquellas distribuciones que están en el dominio de atracción de G_γ , cumplen tener una distribución límite para la función F_u ; y en segundo lugar, indica que el parámetro γ es el mismo en la función G_γ y en la función $H_{\sigma,\gamma}$.

Cabe mencionar también que la propiedad de estabilidad es un concepto importante que tiene la distribución Pareto. Coloquialmente significa que si la familia Pareto ajusta bien a la función de excesos para un valor u_0 , lo mismo ocurrirá para todo $u > u_0$. Finalmente la segunda propiedad se utilizará más adelante para determinar el valor de u que mejor ajusta a una muestra aleatoria dada.

En el capítulo anterior vimos que la función de distribución del máximo puede ser aproximada por la función de extremos generalizada $G_{\gamma,\mu,\sigma}$. A continuación desarrollaremos en forma heurística el teorema anterior que da una idea de porqué la función generalizada Pareto es adecuada para estimar la función de excesos sobre un umbral alto u .

Notemos que $F^n(z) \approx G_{\gamma,\mu,\sigma}(z) = e^{-(1+\gamma\frac{z-\mu}{\sigma})^{-1/\gamma}}$, luego tomando logaritmo tenemos $n \ln F(z) \approx -(1+\gamma\frac{z-\mu}{\sigma})^{-1/\gamma}$. Ahora si $z \rightarrow x^*$ entonces $F(z) \rightarrow 1$ por lo tanto $\ln(F(z)) \approx F(z) - 1$. De donde se obtiene

$$1 - F(z) \approx \frac{1}{n} \left(1 + \gamma \frac{z - \mu}{\sigma} \right)^{-1/\gamma}.$$

Así, para z suficientemente grande tenemos que

$$\begin{aligned} \bar{F}_u(z) = \frac{1 - F(u+z)}{1 - F(u)} &\approx \frac{\left(1 + \gamma \frac{z+u-\mu}{\sigma} \right)^{-1/\gamma}}{\left(1 + \gamma \frac{u-\mu}{\sigma} \right)^{-1/\gamma}} = \left(\frac{\gamma \frac{z}{\sigma} + 1 + \gamma \frac{u-\mu}{\sigma}}{1 + \gamma \frac{u-\mu}{\sigma}} \right)^{-1/\gamma} \\ &\approx \left(1 + \gamma z \cdot \frac{1}{\sigma(1 + \gamma \frac{u-\mu}{\sigma})} \right)^{-1/\gamma} = \left(1 + \gamma \cdot \frac{z}{\sigma + \gamma(u-\mu)} \right)^{-1/\gamma}. \end{aligned}$$

Por lo tanto

$$\bar{F}_u(z) \approx \left(1 + \frac{\gamma z}{\tilde{\sigma}(u)} \right)^{-1/\gamma} \quad (3.1.1)$$

con $\tilde{\sigma}(u) = \sigma + \gamma(u - \mu)$.

3.2. Selección del umbral

En la ecuación (3.1.1) se muestra que para algún valor de u se tendrá la siguiente aproximación $F_u \approx H_{\sigma,\gamma}$. La pregunta natural es determinar el valor de u para el cuál esto sucede en el caso en que la F es desconocida. Para resolver este problema se presentan algunos métodos sugeridos por la bibliografía.

3.2.1. Mean Residual Plot

Utilizando la Propiedad 3.1.1 se tiene que si $F_u = H_{\sigma(u),\gamma}$ para algún $u = u_0$, entonces

$$e(u) = E[X - u | X > u] = \frac{\sigma(u_0) + \gamma u}{1 - \gamma}. \quad \forall u \geq u_0$$

Por lo tanto la esperanza del exceso debe ser una función lineal en u . Dada una muestra aleatoria, la esperanza del exceso puede ser estimada, para cada u utilizando

$$\hat{e}(u) = \frac{\sum_{i=1}^n (X_i - u) I_{\{X_i > u\}}}{n_u}$$

donde $n_u = \sum_{i=1}^n I_{\{X_i > u\}}$. El gráfico mean residual plot está dado por el conjunto $A \subseteq \mathbb{R}^2$

$$A = \{(u, \hat{e}(u)) : u < X_{(n)}\}.$$

La idea del método es seleccionar el menor valor de u para el cual el gráfico del conjunto A es aproximadamente lineal.

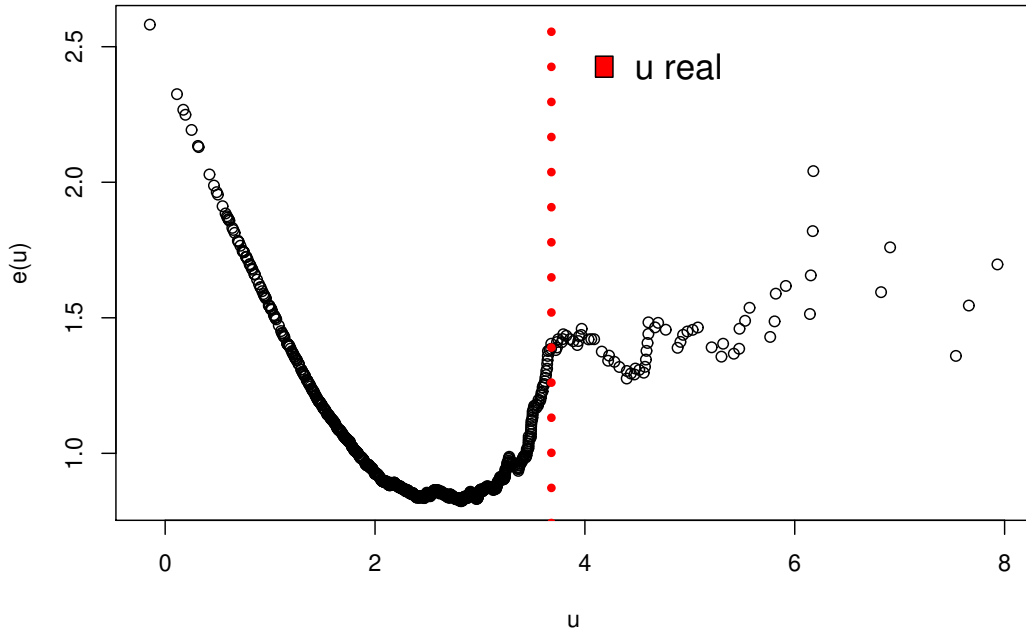


Figura 3.1: Ejemplo de Mean Residual Plot. Los datos fueron generados de manera que sigan la distribución de un variable aleatoria X cumpliendo $X - u_0 |_{X > u_0} \sim H_{\sigma_0, \gamma_0}$. Los valores son $u_0 = 3.480633$; $\sigma_0 = 1,1$ y $\gamma_0 = 0,3$. En rojo, el verdadero valor del umbral.

3.2.2. Estabilización del parámetro de escala y forma

Por la misma Propiedad 3.1.1 utilizada anteriormente, si $F_{u_0} = H_{\sigma_0, \gamma}$ para algún $u = u_0$, entonces $F_u = H_{\sigma_0 + \gamma(u - u_0), \gamma}$. Luego $\sigma_u - \gamma u = \sigma^*$ para todo $u \geq u_0$, mientras que el parámetro γ no se ve modificado. Esto sugiere que la estimación de σ^* y γ debe ser constante a partir de algún u donde F_u sea aproximadamente Pareto. De esta forma, dado un valor de u pueden estimarse $\hat{\sigma}^*(u)$ y $\hat{\gamma}(u)$. Definamos los siguientes conjuntos $B_{\sigma^*}, B_\gamma \subseteq \mathbb{R}^2$ dados por

$$B_{\sigma^*} = \{(u, \hat{\sigma}^*(u)) : u < X_{(n)}\} \quad B_\gamma = \{(u, \hat{\gamma}(u)) : u < X_{(n)}\}$$

Así una buena elección de u consiste en elegir el menor valor de u para el cual los gráficos son aproximadamente constantes.

3.3. Selección del umbral: nueva propuesta

En esta sección desarrollaremos un método de estimación del umbral basado en una propuesta semiparamétrica. El nuevo método de estimación depende de la utilización de un estimador no paramétrico de la densidad, razón por la cual hemos incluido un apéndice que resume el método de estimación de densidad basado en núcleos. Los enfoques anteriores no son automáticos en el sentido que requieren de la elección y el criterio del usuario para determinar el valor del umbral estimado. El método que proponemos depende de la elección del parámetro de suavizado h del estimador por núcleos. Es plausible que desarrollos futuros eliminen este requerimiento, siendo el valor de h calculado a partir de los datos.

3.3.1. El modelo

A fin de proponer un nuevo criterio de selección del umbral supondremos que F satisface la siguiente condición

$$\exists u \text{ tal que } F_u(x) = H_{\sigma, \gamma}(x) \text{ para todo } x \geq u.$$

En otras palabras supondremos que $F \in \Upsilon$ donde Υ es la familia de distribuciones definida por

$$\Upsilon = \left\{ G : G \text{ es función de distribución} \wedge \exists u \text{ tal que } \frac{G(y+u)}{G(u)} = H_{\sigma, \gamma}(y) \right\}.$$

Observación 3.3.1. Si X es v.a., y $X \sim G \in \Upsilon$, entonces debe existir un valor de u para el cual $X - u |_{X > u} \sim H_{\sigma, \gamma}$

En el caso de trabajar con datos empíricos, difícilmente valga esta hipótesis. No obstante, al suponer que los datos reales satisfacen que F_u es aproximadamente como $H_{\sigma,\gamma}(x)$, se espera que el error cometido al hacer la suposición de igualdad sea pequeño en comparación con la ventaja que representa hacer estimaciones en este caso.

En el marco en que $F \in \Upsilon$, las funciones de distribución admiten una escritura particular. Esta escritura permitirá luego deducir, naturalmente estimadores adecuados.

Sea $X \sim F$ notemos que F puede ser escrita de la siguiente forma,

$$P(X < x) = \underbrace{P(X < x|X \leq u)P(X \leq u)}_A + \underbrace{P(X < x|X > u)P(X > u)}_B.$$

Analizaremos dos casos $x > u$ y $x \leq u$.

- ($x \leq u$) En este caso solo quedará el término $A = P(X < x)$ siendo el término B igual a 0.
- ($x > u$) En este caso el término $A = P(X \leq u)$ mientras que para el término B se tiene que

$$B = P(X < x|X > u)P(X > u).$$

Restando u de ambos lados

$$B = P(X - u < x - u|X > u)P(X > u).$$

Del hecho que $F \in \Upsilon$ tenemos

$$B = H_{\sigma,\gamma}(x - u)P(X > u).$$

De esta forma $F(x)$ puede escribirse como

$$F(x) = F(x)\mathbf{1}_{\{x \leq u\}} + [P(X \leq u) + H_{\sigma,\gamma}(x - u)P(X > u)]\mathbf{1}_{\{x > u\}}$$

llamando $p_u = P(X \leq u)$

$$F(x) = F(x)\mathbf{1}_{\{x \leq u\}} + [p_u + H_{\sigma,\gamma}(x - u)(1 - p_u)]\mathbf{1}_{\{x > u\}}.$$

Si F admite densidad $(F)' = f$ y $(H_{\sigma,\gamma})' = h_{\sigma,\gamma}$ obtenemos que

$$f(x) = f(x)\mathbf{1}_{\{x \leq u\}} + [h_{\sigma,\gamma}(x - u)(1 - p_u)]\mathbf{1}_{\{x > u\}},$$

que puede escribirse como

$$f(x) = p_u \frac{f(x)}{p_u} \mathbf{1}_{\{x \leq u\}} + [h_{\sigma,\gamma}(x - u)(1 - p_u)]\mathbf{1}_{\{x > u\}}.$$

Observemos que $f_1(x) = \frac{f(x)}{p_u} \mathbf{1}_{\{x \leq u\}}$ es una función de densidad soportada en el conjunto $\{x \in \text{Sop}(f) : x \leq u\}$, de donde se obtiene

$$f(x) = p_u f_1(x) \mathbf{1}_{\{x \leq u\}} + (1 - p_u) h_{\sigma, \gamma}(x - u) \mathbf{1}_{\{x > u\}}.$$

A su vez $f_2(x) = h_{\sigma, \gamma}(x - u) \mathbf{1}_{\{x > u\}}$ también es una función de densidad soportada en $\{x : x > u\}$

De esta manera, hemos caracterizado a las funciones de distribución F que pertenecen a la familia Υ y que admiten densidad, escribiéndolas como una combinación convexa de dos densidades f_1 y f_2 . Un punto importante es la información relativa a f_2 , que nos permite garantizar que pertenece a una familia Pareto. En particular la función f_2 esta determinada salvo parámetros. En cambio, el conocimiento de f_1 es nulo, sólo sabemos que $\text{Sop}(f_1) \cap \text{Sop}(f_2) = \emptyset$. Hecha la aclaración de los soportes podemos escribir

$$f(x) = p_u f_1(x) + (1 - p_u) f_2(x).$$

Es importante recalcar que el soporte de f_1 y de f_2 depende de u , de esta manera hemos probado la siguiente propiedad.

Propiedades 3.3.1. *Sea F una función de distribución admitiendo densidad f*

$$F \in \Upsilon \Leftrightarrow f(x, u) = p_u f_1(x, u) + (1 - p_u) f_2(x, u)$$

donde f_1 y f_2 son densidades tales que $f_2(x, u) = h_{\sigma, \gamma}(x - u) \mathbf{1}_{\{x > u\}}$ y $\text{Sop}(f_1) \cap \text{Sop}(f_2) = \emptyset$.

Notemos que si $X \sim F \in \Upsilon$, entonces existe un umbral u tal que $X - u|_{X > u} \sim H_{\sigma, \gamma}$, no obstante esta propiedad no determina unívocamente el valor del umbral. Esto se debe a que, bajo estas condiciones, para cualquier $\delta > 0$ se tiene que $X - (u + \delta)|_{X > u + \delta} \sim H_{\tilde{\sigma}, \gamma}$ como se prueba en el siguiente resultado.

Definición 3.3.1. Dada $X \sim F \in \Upsilon$ se dice que u realiza el modelo si $X - u|_{X > u} \sim H_{\tilde{\sigma}, \tilde{\gamma}}$, para algún $\tilde{\sigma}, \tilde{\gamma}$

Proposición 3.3.1. *Sea $X \sim F \in \Upsilon$ y supongamos que u realiza el modelo entonces para todo $\delta > 0$, $u + \delta$ también realiza el modelo.*

Demostración Proposición 3.3.1. Si u realiza el modelo, entonces para ciertos $\sigma = \sigma_0$ y $\gamma = \gamma_0$ se tiene que $P(X - u > z | X > u) = 1 - H_{\sigma_0, \gamma_0}(z)$. Queremos ver que $u + \delta$ también realiza el modelo, para esto veamos que

$$P(X - (u + \delta) > z | X > (u + \delta)) = 1 - H_{\tilde{\sigma}, \tilde{\gamma}}(z).$$

Notemos que

$$\begin{aligned}
 P(X - (u + \delta) > z | X > (u + \delta)) &= \frac{1 - F(z + u + \delta)}{1 - F(u + \delta)} \\
 &= \frac{1 - F(z + u + \delta)}{1 - F(u)} \frac{1 - F(u)}{1 - F(u + \delta)} \\
 &= \frac{1 - F(z + u + \delta)}{1 - F(u)} \bigg/ \frac{1 - F(u + \delta)}{1 - F(u)}. \quad (3.3.1)
 \end{aligned}$$

Como u realiza el modelo, se tienen las siguientes identidades

$$\frac{1 - F(z + u + \delta)}{1 - F(u)} = 1 - H_{\sigma_0, \gamma_0}(z + \delta)$$

y que

$$\frac{1 - F(u + \delta)}{1 - F(u)} = 1 - H_{\sigma, \gamma}(\delta).$$

Entonces de (3.3.1) se desprende

$$P(X - (u + \delta) > z | X > (u + \delta)) = \frac{1 - H_{\sigma_0, \gamma_0}(z + \delta)}{1 - H_{\sigma_0, \gamma_0}(u + \delta)} \quad (3.3.2)$$

como $1 - H_{\sigma_0, \gamma_0}(z) = \left(1 + \frac{z\gamma_0}{\sigma_0}\right)^{-1/\gamma_0}$ el lado derecho de la igualdad se traduce en

$$\frac{\left(1 + \frac{(z+\delta)\gamma_0}{\sigma_0}\right)^{-1/\gamma_0}}{\left(1 + \frac{\delta\gamma_0}{\sigma_0}\right)^{-1/\gamma_0}} = \left(\frac{1 + \frac{\delta\gamma_0}{\sigma_0} + \frac{z\gamma_0}{\sigma_0}}{1 + \frac{\delta\gamma_0}{\sigma_0}}\right)^{-1/\gamma_0} = \left(1 + \frac{z\gamma_0}{\sigma_0 + \delta\gamma_0}\right)^{-1/\gamma_0} = H_{\sigma_0 + \gamma_0\delta, \delta}(z).$$

Por lo tanto de (3.3.2) tenemos $P(X - (u + \delta) > z | X > (u + \delta)) = H_{\sigma_0 + \gamma_0\delta, \delta}(z)$, luego $u + \delta$ realiza el modelo con $\tilde{\sigma} = \sigma_0 + \gamma_0\delta$ y $\tilde{\gamma} = \gamma_0$. \square

La proposición anterior motiva la siguiente definición.

Definición 3.3.2. Dada $X \sim F \in \Upsilon$ se dice que u_0 es el *umbral inicial* de F si u_0 es el ínfimo de los valores u que realizan el modelo.

$$u_0 = u_0(F) = \inf\{u : X - u |_{X > u} \sim H_{\tilde{\sigma}, \tilde{\gamma}}\}$$

3.3.2. La propuesta de estimación

En esta Sección daremos una propuesta para estimar el umbral inicial u_0 , basada en una muestra aleatoria $\{X_1, \dots, X_n\}$ tal que $X_i \sim F \in \Upsilon$. La estimación “natural” que surge de la sección anterior es considerar un estimador no paramétrico para f_1 y un estimador paramétrico para f_2 . Se elegirá el menor valor de u que ajuste a los datos basado en una función de pseudoverosimilitud $L(u)$ según se detalla en el siguiente procedimiento.

Paso 1: Construcción de la función de pseudoverosimilitud $L(u)$

1.1: Consideremos \hat{f} un estimador no paramétrico de la densidad basado en X_1, \dots, X_n .

1.2: Para cada u sean

- $C_u = \int_{-\infty}^u \hat{f}(x) dx$
- $n_u = \sum_{i=1}^n 1_{\{X_i \leq u\}}$ y $\hat{p}_u = \frac{n_u}{n}$
- $\hat{f}_{1u}(x) = \frac{\hat{f}(x) 1_{\{x \leq u\}}}{C_u}$
- $\hat{\gamma}_0(u), \hat{\sigma}_0(u)$ los estimadores de los parámetros de la distribución Pareto basados en las observaciones $X_i - u$ tales que $X_i > u$.

1.3: Para cada u , basados en los estimadores definidos anteriormente, calculamos un estimador semiparamétrico la densidad f ,

$$\hat{f}_u(x) = \hat{p}_u \hat{f}_{1u}(x) + (1 - \hat{p}_u) h_{\hat{\sigma}_0(u), \hat{\gamma}_0(u)}(x - u).$$

1.4: Construimos la pseudoverosimilitud $L(u) = \prod_{i=1}^n \hat{f}_u(X_i)$

Paso 2: Finalmente calculamos un estimador de u_0 , \hat{u} como el mínimo del conjunto donde L se maximiza. Es decir,

$$\hat{u} = \min\{u : \text{tal que } L(u) = \max_z L(z)\}.$$

Observación 3.3.2. Notemos que en el paso **1.4** estamos evaluado \hat{f}_u en el punto X_i y esto puede aumentar el sesgo en la estimación de la pseudo verosimilitud. Una alternativa a ser considerada sería calcular $\hat{f}_{u,i}(X_i)$ en lugar $\hat{f}_u(X_i)$ donde $\hat{f}_{u,i}(\cdot)$ se construye excluyendo a la observación X_i tanto en la estimación paramétrica como en la no paramétrica.

3.4. Resultados numéricos

En esta sección desarrollaremos un estudio de simulación con el objetivo de evaluar el método propuesto en la sección anterior. El análisis de las propiedades de la función $L(u)$ y su comportamiento asintótico es objeto de trabajos en curso, pero su desarrollo excede el alcance de esta tesis.

3.4.1. Modelo uniforme - exponencial

En el primer modelo que estudiaremos supondremos conocido el parámetro de forma mas aún consideraremos $\gamma = 0$. En este caso, la función $h_{\sigma, \gamma}$ queda dentro de la familia exponencial,

$$h_{\sigma, \gamma}(x - u) = \frac{1}{\sigma} \exp\left\{\frac{-1}{\sigma}(x - u)\right\} 1_{\{x > u\}}.$$

Además asumiremos que f_1 pertenece a la familia uniforme en el intervalo $[0, u_0]$ con $u_0 = 5$. De esta manera el modelo simplificado corresponde a

$$f(x) = p_{u_0} f_1(x) + (1 - p_{u_0}) \lambda e^{-\lambda(x-u_0)} 1_{\{x > u_0\}}$$

con $\lambda_0 = 0,5$ y $p_{u_0} = 0,95$.

Se generaron 1000 muestras de tamaño n , con $n = 500, 1000, 1500, 2000$ para cada una de las muestras se calculo la pseudoverosimilitud $L(u)$ según el procedimiento descrito en la sección anterior. Para ello en el Paso 1.1 consideramos un estimador noparamétrico de la densidad basado en núcleos. En este caso utilizamos el núcleo Epanechnikov¹, y consideramos diferentes valores de ventana h ($h=0.1, 0.25, 0.5, 0.75, 1$). En el Paso 1.2 para estimar el parámetro λ_0 de la cola de la distribución consideramos el estimador de máxima verosimilitud basado en las observaciones mayores que el valor u . Es decir,

$$\hat{\lambda}_0(u) = \frac{\sum 1_{\{X_i > u\}}}{\sum (X_i - u) 1_{\{X_i > u\}}}.$$

Así la función de pseudoverosimilitud queda definida por $L(u) = \prod_{i=1}^n \hat{f}_u(X_i)$ donde

$$\hat{f}_u(x) = \hat{p}_u \hat{f}_{1u}(x) + (1 - \hat{p}_u) \hat{\lambda}_0(u) e^{-\hat{\lambda}_0(u)(x-u)} 1_{\{x > u\}}.$$

Es importante aclarar que estamos suponiendo que el verdadero valor de u_0 corresponde a un punto para el cual la distribución de la cola sigue aproximadamente una distribución Pareto. Así, en el problema general esperaríamos que los valores de u_0 considerados sean suficientemente grandes. En este sentido el valor $1 - p_{u_0}$, el cual indica qué proporción de datos siguen la distribución Pareto, debe ser chico. Esto se condice con el hecho de generar modelos con valores de p_{u_0} cercanos a 1, como consideramos en el esquema de simulación. Por tal motivo, para optimizar la función L consideramos valores de u entre los estadísticos de orden $X^{([0,75n])}$ y $X^{(n)}$.

Se programó un algoritmo que compute la función L , es decir, dada una grilla equiespaciada de t puntos, $\mathbf{U} = \{X^{([0,75n])} = u_1 < u_2 < \dots < u_{t-1} < u_t = X^{(n)}\}$, se computó el arreglo correspondiente a la función L ,

$$\mathbf{L} = (l_1, \dots, l_t) = (L(u_1), \dots, L(u_t)).$$

Cualitativamente, los posibles resultados de esta rutina se muestran gráficamente en la Figura 3.2. Cada uno de los casos, puede corresponder al conjunto de datos generados o a distintas elecciones del valor del parámetro de suavizado h utilizado en el estimador de densidad por núcleos.

En el caso A) el valor u donde la función L alcanza su máximo es $u = 4,971$, éste da una buena aproximación al valor verdadero $u_0(F)$ ($u_0(F) = 5$), mientras que en el caso

¹Ver Apéndice

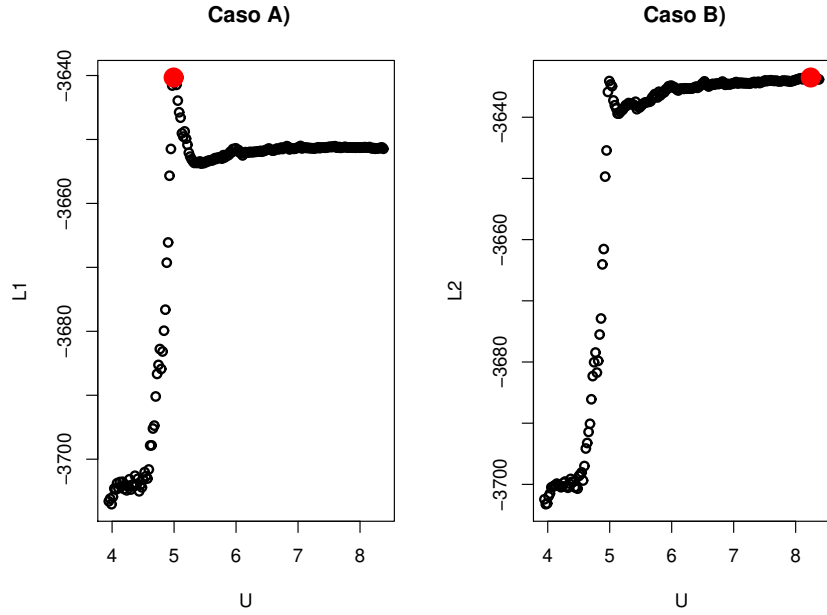


Figura 3.2: En rojo el valor máximo de L

B) no. En ambos casos hay un valor \bar{u} a partir del cual la función crece muy lentamente respecto del crecimiento para valores anteriores a \bar{u} .

Esta lentitud de crecimiento, refleja la propiedad 3.3.1. Es decir, la función de pseudo-verosimilitud L evaluada en u , puede ser entendida como una medida de cuán bien se ajustan los datos al modelo, suponiendo que el valor real es u . Dado $u > u_0(F)$ esperamos que un corrimiento en $u + \delta$, no provoque grandes diferencias en el valor de L debido a que los datos debieran ajustarse a los dos modelos (en virtud de la propiedad 3.3.1). Por ello suponemos que $L(u) \approx L(u + \delta)$. Luego, el valor de \bar{u} será un buen estimador de umbral inicial $u_0(F)$.

Ante la necesidad de computar aproximadamente el valor de \bar{u} , se utilizó una estrategia heurística, consistente en reconocer el valor de u para el cual la función es aproximadamente constante por medio de la comparación con su valor medio parcial \mathbf{S} .

Dado el arreglo \mathbf{L} definimos un nuevo arreglo formado por el valor medio parcial $\mathbf{S} = \{\mathbf{S}_j : 1 \leq j \leq t\}$, donde \mathbf{S}_j es definido como

$$\mathbf{S}_j = \frac{1}{j} \sum_{i=1}^j \mathbf{L}_i.$$

Llamamos \mathbf{D} al arreglo de diferencias que en la posición j , con $1 \leq j \leq t - 1$ queda

definido por,

$$\mathbf{D}_j = \mathbf{L}_{j+1} - \mathbf{S}_j.$$

La idea es que la diferencia entre \mathbf{S} y \mathbf{L} se maximiza en los lugares donde L tiene un máximo, o de lo contrario allí donde hay un cambio abrupto de crecimiento.

De esta forma, obtenemos dos posibles estimadores \hat{u}_{0A} y \hat{u}_{0B} para estimar $u_0(F)$, definidos de la siguiente manera.

a) En sintonía con la definición 3.3.2 y el gráfico A) en la Figura 3.2 tomamos J_0 como

$$J_{0A} = \min\{j : \mathbf{L}_j = \max_{1 \leq k \leq t} \mathbf{L}_k\}$$

y definimos

$$\hat{u}_{0A} = \mathbf{U}_{J_{0A}}.$$

b) Alternativamente, utilizando la heurística propuesta, tomamos el argumento máximo de \mathbf{D} :

$$J_{0B} = \min\{j : \mathbf{D}_i \leq \mathbf{D}_j, \quad 1 \leq i \leq t-1\}$$

y definimos

$$\hat{u}_{0B} = \mathbf{U}_{J_{0B}}.$$

La Figura 3.3 muestra como la heurística propuesta permite detectar sistemáticamente, tanto en el caso A) como en el caso B) de la Figura 3.2, el valor donde la función de pseudoverosimilitud L cambia de forma drástica su comportamiento.

Para resumir los resultados obtenidos presentamos en las tablas **I)** y **II)** el error cuadrático medio (ECM) correspondiente a cada uno de los estimadores calculados en las 1000 replicaciones.

Tabla I)

$ECM(\hat{u}_{0A})$	N=500	N=1000	N=1500	N=2000
$h = 0,10$	5,7592	14,0654	10,0045	14,0350
$h = 0,25$	0,7297	4,4454	7,8544	11,5357
$h = 0,50$	0,1674	0,7462	0,9716	1,0456
$h = 0,75$	0,0561	0,0492	0,0638	0,0415
$h = 1$	0,0129	0,0115	0,0020	0,0021

Tabla II)

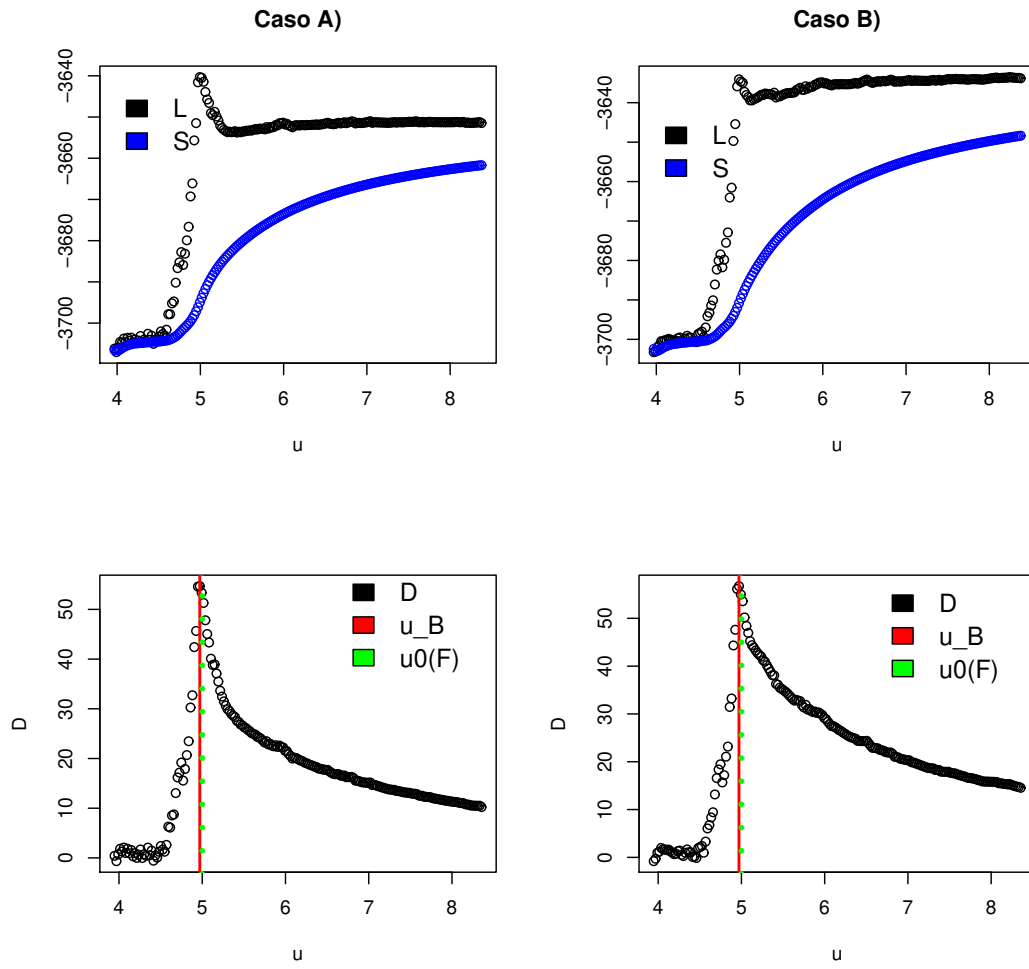


Figura 3.3: Comportamiento de la heurística utilizada.

$ECM(\hat{u}_{0B})$	N=500	N=1000	N=1500	N=2000
$h = 0,10$	0,8095	0,1175	0,0148	0,0055
$h = 0,25$	0,0298	0,0101	0,0053	0,0063
$h = 0,50$	0,0173	0,0043	0,0054	0,0066
$h = 0,75$	0,0121	0,0043	0,0056	0,0068
$h = 1,0$	0,0148	0,0044	0,0056	0,0068

En la tabla I) podemos ver que el estimador \hat{u}_{0A} , se comporta mal para valores pequeños de h . Esto se debe a que, en este caso, se producen grandes apartamientos del valor real de $u_0(F)$, arruinando el error cuadrático medio. Estos grandes apartamientos se deben a que el caso B) de la Figura 3.2 ocurre con mayor frecuencia si tomamos valores de h chicos a lo largo de las 1000 replicaciones. Este efecto no es subsanado al aumentar el tamaño de muestra. Para valores de h grande ($h = 1$), la comparación entre las Tablas I) y II), muestran que los estimadores \hat{u}_{0A} y \hat{u}_{0B} arrojan valores pequeños de ECM, resultando comparables. Finalmente en la tabla II) observamos que para valores pequeños de ventana el estimador \hat{u}_{0B} tiene un buen comportamiento.

En la Figura 3.4 se ve el comportamiento de la densidad estimada, suponiendo distintos valores de u_1 y un valor de $h = 1$. El gráfico correspondiente de la función L es el de la parte A) de la Figura 3.2.

1. Caso $u_1 < u_0(F)$. Ejemplificamos con $u_1 = 3$, aquí el ajuste de los datos es relativamente bueno si se observan valores de $x \leq 3$, pero malo para $x > 3$. En este caso el suponer que el modelo exponencial es válido, cuando en realidad vale el modelo uniforme, provoca un gran desajuste que se ve reflejado en valores chicos de la función de pseudoverosimilitud L , siendo $L(u_1) < L(u_0(F))$.
2. Caso $u_1 > u_0(F)$. Ejemplificamos con $u_1 = 7$, donde el ajuste de los datos es relativamente bueno para los valores de $x \geq 7$, pero no tan bueno para valores de x entre 5 y 7. Esto es debido a que un valor de h grande da un mal estimador en el sector donde la densidad real de los datos cambia bruscamente. Paradojicamente, usar un estimador no paramétrico que funcione defectuosamente de manera local, colabora a que los datos no se ajusten bien al modelo suponiendo valores de u mayores al verdadero umbral $u_0(F)$, teniendo como consecuencia que la función de pseudoverosimilitud en esos valores sea más chica que en el verdadero umbral. Siendo también $L(u_1) < L(u_0(F))$.

3.4.2. Modelo normal - Pareto

Los datos fueron simulados con una distribución de probabilidad tal que su densidad sea $f(x)$ con

$$f(x) = p_u f_1(x) + (1 - p_u) h_{\sigma, \gamma}(x - u) 1_{\{x > u\}}$$

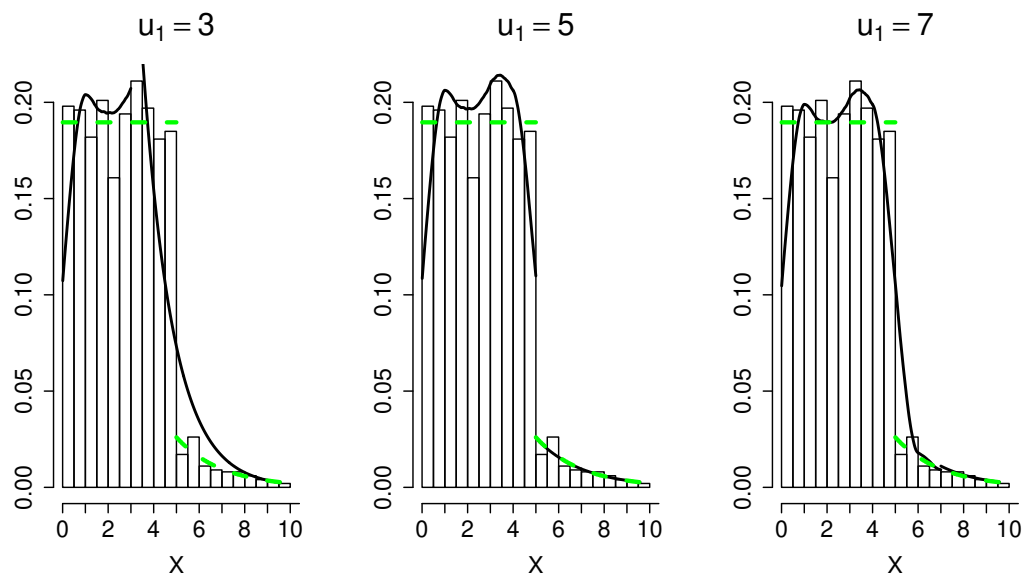


Figura 3.4: En línea punteada la densidad con que fue generada la muestra, en línea sólida la función $\hat{f}(x, u_1) = \hat{p}_{u_1} \hat{f}_1(x, h) \mathbf{1}_{\{x \leq u_1\}} + (1 - \hat{p}_{u_1}) \hat{\lambda} e^{-\hat{\lambda}(x-u_1)} \mathbf{1}_{\{x > u_1\}}$, el valor real es $u = 5$ y con $\hat{f}_1(x, h)$ el estimador de densidad por núcleos, con $h = 1$.

En este caso consideramos como f_1 una función Normal truncada en el umbral u es decir

$$f_1(x, a, b) = \frac{g(x, a, b)}{\int_{-\infty}^u g(s, a, b) ds} 1_{\{x \leq u\}}$$

con $g \sim N(a, b^2)$.

Los parámetros asignados fueron los consignados en la siguiente tabla:

u	σ	γ	p_u	a	b
3,480633	1,1	0,3	0,93	2,3	0,8

El valor de u asignado se hizo de manera que valga que $P(Z \leq u) = 1 - p_u$, donde $Z \sim N(a, b^2)$. Se generaron 1000 muestras de tamaño n , con $n = 500, 1000, 1500, 2000$ para cada una de las muestras se calculo la pseudoverosimilitud $L(u)$ según el procedimiento descrito en la sección 3.3.2. En el Paso 1.1 consideramos un estimador noparamétrico de la densidad basado en núcleos al igual que en el caso anterior. En el Paso 1.2 para estimar los parámetros γ_0 y σ de la cola de la distribución, consideramos el estimador de máxima verosimilitud basado en las observaciones mayores que el valor u . Para realizar el cálculo de estos parámetros se utilizó el paquete *evir*, desarrollado por Mc Neil (2011), que incluye una función para estimarlos. La misma se basa en maximizar la función de verosimilitud utilizando algoritmos de optimización continua para funciones de varias variables. Este paquete también fue utilizado para generar variables aleatorias cuya distribución sea Pareto.

El caso considerado también tiene discontinuidades en la densidad, como puede observarse en el gráfico de la misma en la Figura 3.5. En la figura también se observa el histograma de una de las muestras generadas de tamaño $n = 2000$. Entendemos que esta discontinuidad no es reconocible a simple vista, como ocurre en el caso Uniforme - Exponencial. Afortunadamente, los resultados obtenidos son equivalentes a los de la sección anterior. Observamos que para \hat{u}_{0A} el error cuadrático medio es pequeño si el valor de ventana h es grande, mientras que el estimador \hat{u}_{0B} se comporta satisfactoriamente con anchos de banda h chicos.

La Figura 3.4.2 destaca que en este caso, y en general para muestras donde la distribución es Pareto, la función de pseudoverosimilitud L tiende a ser más inestable que en el modelo Uniforme - Exponencial.

Por último, la Figura 3.7 da cuenta del buen ajuste ofrecido por el modelo.

Tabla III)

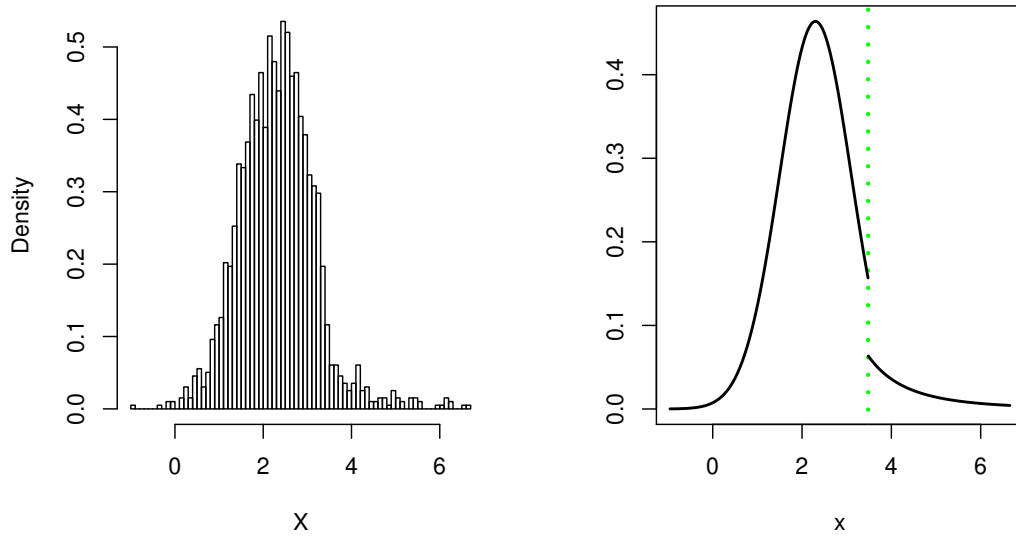


Figura 3.5: Histograma de la muestra generada y densidad real

$ECM(\hat{u}_{0A})$	n=500	n=1000	n=1500	n=2000
$h = 0,10$	0,3733	2,5130	5,0649	7,8893
$h = 0,25$	0,1926	1,6979	3,9043	6,2763
$h = 0,5$	0,0569	0,2759	0,3820	0,4942
$h = 0,75$	0,0374	0,0182	0,0070	0,0090
$h = 1,0$	0,0404	0,0215	0,0106	0,0076

Tabla IV)

$ECM(\hat{u}_{0B})$	n=500	n=1000	n=1500	n=2000
$h = 0,10$	0,1451	1,0510	1,8743	2,4655
$h = 0,25$	0,0858	0,2767	0,3065	0,3029
$h = 0,5$	0,0628	0,1215	0,0510	0,0631
$h = 0,75$	0,0540	0,0543	0,0389	0,0381
$h = 1,0$	0,0554	0,0516	0,0414	0,0387

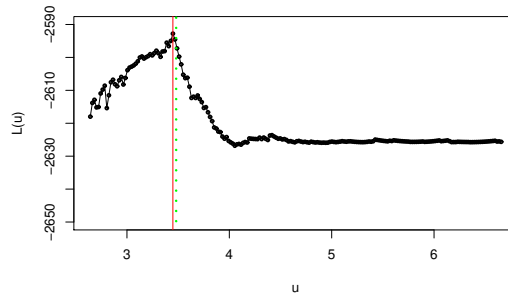


Figura 3.6: Inestabilidades en la función L

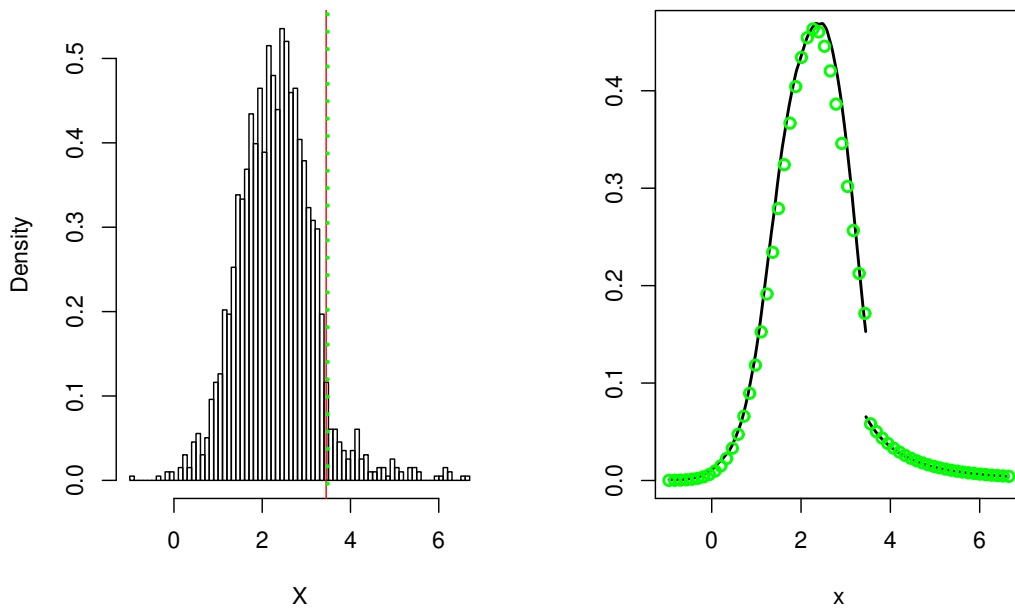


Figura 3.7: **Izquierda:** Histograma de la muestra, $u_1 = \widehat{u_{0A}} = 3,44862$ (en rojo, línea sólida), siendo el verdadero valor $u_0(F) = 3,480633$ (en verde, línea de puntos). No se muestra el valor de $\widehat{u_{0B}}$. ($\widehat{u_{0B}} = 3,428449$) **Derecha:** Ajuste provisto por el estimador semiparamétrico: en línea sólida la densidad con que fue generada la muestra y en línea punteada la función $\widehat{f}(x, u_1) = \widehat{p}_{u_1} \widehat{f}_1(x, h) 1_{\{x \leq u_1\}} + (1 - \widehat{p}_{u_1}) h_{\widehat{\sigma}(u_1) \widehat{\gamma}(u_1)}(x - u_1) 1_{\{x > u_1\}}$. El estimador de densidad por núcleos es con $h = 0,5$.

Apéndice A

Estimación no paramétrica de la densidad basada en núcleos

Sea X_1, \dots, X_n una muestra aleatoria con función de densidad $f(x)$ estamos interesados en estimar $f(x)$ a partir de las observaciones. Los métodos de estimación no paramétricos han surgido con el objetivo de dar una respuesta a este problema y han sido ampliamente estudiados. A continuación estudiaremos una propuesta para la estimación de la función de densidad basada en núcleos que fue introducida por Rosenblatt (1956) y Parzen (1962). Comenzaremos dando una idea intuitiva del estimador propuesto.

Si X es una variable aleatoria con densidad f , luego dado $h > 0$

$$\frac{1}{2h}P(x-h < X < x+h) = \frac{1}{2h} \int_{x-h}^{x+h} f(x)dx \rightarrow f(x)$$

si $h \rightarrow 0$.

Un estimador natural de $P(x-h < X < x+h)$ es simplemente considerar la proporción de la muestra que cae en el intervalo $(x-h, x+h)$ luego para un h suficientemente pequeño podemos deducir el siguiente estimador de $f(x)$,

$$\tilde{f}(x) = \frac{1}{2h} \frac{\# \{X_i : X_i \in (x-h, x+h)\}}{n}.$$

Más formalmente, podemos escribir el estimador anterior de la siguiente manera,

$$\tilde{f}(x) = \sum_{i=1}^n \frac{1}{hn} w\left(\frac{x - X_i}{h}\right) \quad (\text{A.0.1})$$

donde $w(x) = \frac{1}{2}I_{(|x|<1)}$. Es conveniente observar que la función w satisface las siguientes propiedades:

1. $w \geq 0$

$$2. \int w(s)ds = 1$$

$$3. \text{ Para cada } 1 \leq i \leq n, w\left(\frac{x-X_i}{h}\right) = \frac{1}{2} \text{ si y solo si } X_i \in (x-h, x+h).$$

Observemos que la propiedad 3 nos indica que la función w le otorga un peso uniforme a cada observación X_i en el entorno $(x-h, x+h)$. De esta forma, una manera de generalizar (A.0.1) sería reemplazar la función de peso o núcleo w por una función K no negativa que verifique la condición $\int K(x)dx = 1$. De esta forma, si consideramos una función de pesos K con mayor suavidad obtendríamos un estimador más suave.

De esta manera obtenemos el estimador definido por Rosenblatt (1956) y Parzen (1962) que definimos de la siguiente manera,

$$\hat{f}(x) = \frac{1}{hn} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \quad (\text{A.0.2})$$

donde K es una función núcleo, $h = h_n$ es llamado parámetro de suavizado o ancho de ventana y satisface $h_n \rightarrow 0$ si $n \rightarrow \infty$.

Este estimador constituye uno de los estimadores no paramétricos más estudiados. Dentro de sus propiedades más destacables podemos resaltar en primer lugar que el estimador es efectivamente una densidad, por otro lado bajo ciertas condiciones se puede obtener el desarrollo de su sesgo y varianza. Asimismo, también se ha estudiado su comportamiento asintótico obteniendo su consistencia y distribución asintótica. Todas las demostraciones de estas propiedades no fueron objeto de estudio de esta tesis.

El proceso de cálculo de este estimador involucra la elección de dos parámetros, el núcleo y el parámetro de suavizado. En general la función de pesos o núcleos utilizadas decrecen de manera suave, dándole así menor pesos a las observaciones más alejadas del punto x . Algunas opciones posibles de núcleos, podrían ser

Núcleo Gaussiano:

$$K(t) = e^{-\frac{1}{2}t^2}.$$

Núcleo Epanechnikov:

$$K(t) = \begin{cases} \frac{3}{4}(1-t^2) & |t| \leq 1 \\ 0 & |t| > 1. \end{cases}$$

Núcleo Tricúbico:

$$K(t) = \begin{cases} (1-|t|^3)^3 & |t| \leq 1 \\ 0 & |t| > 1. \end{cases}$$

Típicamente la elección de la función de pesos no suele involucra mayores problemas, ni ser determinante en el proceso de estimación. Sin embargo, el parámetro de suavizado suele ser un punto crucial en el procedimiento de estimación, ya que como su nombre lo indica se encuentra altamente relacionado con el nivel de suavización que se introduce en la estimación. Respecto a la selección de este parámetro existe una gran cantidad de literatura que escapa al alcance de esta tesis.

Bibliografía

- [1] Balkema, A. A., de Haan, L., (1974) Residual life time at great age. The Annals of Probability. **2** , 792–804
- [2] Beirlant, J., De Waal, D., Goegebeur Y., Teugels, J., Segers, J. (2004). Statistics Of Extremes, John Wiley & Sons Inc.
- [3] Coles, S. G. and Tawn, J. A. (1994) Statistical methods for multivariate extremes: An application to structural design. Applied Statistics, **43**, 1-48.
- [4] Coles, S. G. and Tawn, J. A. (1996) A Bayesian analysis of extreme rainfall data. Applied Statistics, **45**, (4), 463-478.
- [5] Coles, S. G. and Powell, E. A. (1996) Bayesian methods in extreme value modelling: a review and new developments. International Statistical Review, **64**, (1), 119-136.
- [6] Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer.
- [7] Davison, A. C. and Smith, R. L. (1990) Models for exceedances over high thresholds. Journal of the Royal Statistical Society B, **52**, 393-442.
- [8] De Haan, L. (1970), On regular variation and its application to the weak convergence of sample extremes. Mathematical Centre tracts.32. Math. Centrum Amsterdam
- [9] De Haan, L., Ferreira, A. (2006). Extreme Value Theory. Springer.
- [10] Fisher, R. y Tippett, L., (1928). On the estimation of the frequency distribution of the largest or smallest member of a sample. Proc. Cambridge Philos. Soc. 24, 180–190.
- [11] Frigessi, A., Haug, O. and Rue, H. (2003) Tail estimation with Generalized Pareto Distribution without threshold selection Extremes, **5** (3), 219 - 236, 2003.
- [12] Gnedenko, B.V. (1943) Sur la distribution limite du terme maximum d'une serie aleatoire. Annals of Mathematics. **44**, 423-453.
- [13] Gumbel, E. (1958), Statistics of Extremes, Columbia University Press.

- [14] Mc Neil, A., (2011). *evir: Extreme values in R. S original (Evis) by Alexander Mc Neil and R Port by Alec Stephenson*
- [15] Ortega, J. (2010). *Notas del curso Valores Extremos*.
Disponibile en: <http://www.cimat.mx/~jortega/extremos10.html>
- [16] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065–1076.
- [17] Pickands, J, (1975). Statistical Inference using extreme order statistics. *The Annals of Statistics.* **3** 119–131
- [18] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832–837
- [19] Smith, R. (1987). Estimating Tails of Probability Distributions. *The Annals of Statistics.* **15**, 1174-1207.
- [20] von Mises, R.(1936). *La distribution de la plus grande de n valeurs*. [Reprinted in *Selected Papers Vol. II*, American Mathematical Society, Providence, R.I., 1954, 271-294]