



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ESTIMACIONES DE INTERVALO EN EL
ANÁLISIS DE MÁXIMOS POR BLOQUES.
UN ESTUDIO BASADO EN
SIMULACIONES

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

MATEMÁTICO

PRESENTA:

ALEJANDRO CRUZ MARCELO

DIRECTOR DE TESIS:

DR. JOAQUÍN ORTEGA SÁNCHEZ



2007



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Hoja de Datos del Jurado

1. Datos del alumno

Cruz

Marcelo

Alejandro

57457281

Universidad Nacional Autónoma de México

Facultad de Ciencias

Matemáticas

2. Datos del tutor

Doctor

Joaquín

Ortega

Sánchez

3. Datos del sinodal 1

Doctora

Ana

Meda

Guardiola

4. Datos del sinodal 2

Matemática

Margarita Elvira

Chávez

Cano

5. Datos del sinodal 3

Doctora

Ruth

Fuentes
García

6. Datos del sinodal 4

Actuario
Jaime
Vázquez
Alamilla

7. Datos del trabajo escrito

Estimaciones de intervalo en el análisis de máximos por bloques. Un estudio basado en simulaciones.

45 p.
2007

A mi familia.

Agradecimientos

A la Dra. Eloísa Díaz-Francés y al M.C.Eliud Vilchis por su ayuda para resolver algunos problemas numéricos que aparecieron en los cálculos involucrados en esta tesis. Al M.C. Pablo Galaviz por su apoyo para conseguir el equipo de cómputo que fue requerido para la realización de este trabajo. Las simulaciones fueron realizadas en el Laboratorio de Supercómputo Astrofísico del Departamento de Física, CINVESTAV.

A mi tutor el Dr. Joaquín Ortega Sánchez por su disposición para discutir cualquier aspecto relacionado con la tesis. A los integrantes del jurado por sus comentarios que mejoraron el contenido de este documento. Este trabajo forma parte del proyecto 07-02-K662-044 A02 de Concyteg.

Índice general

Introducción	1
1. Teoría de Valores Extremos	3
1.1. Teorema de Tipos para Extremos	3
1.2. Dominios de Atracción	7
1.2.1. Dominio de Atracción de las <i>DVE</i> Tipo Fréchet (Φ_α)	7
1.2.2. Dominio de Atracción de las <i>DVE</i> Tipo Weibull (Ψ_α)	9
1.2.3. Dominio de Atracción de las <i>DVE</i> Tipo Gumbel (Λ)	9
1.3. Distribución Generalizada de Valores Extremos	12
1.4. Niveles y periodos de Retorno	13
1.5. Dominios de Atracción para Algunas Distribuciones Conocidas	13
2. Métodos de Estimación en el Análisis de Máximos por Bloques	16
2.1. Herramientas Basadas en la Verosimilitud	17
2.1.1. Estimadores de Máxima Verosimilitud	18
2.1.2. Intervalos de Verosimilitud y Aproximación χ^2	20
2.1.3. Verosimilitud Perfil	21
2.1.4. Aplicación a las Distribuciones de Valores Extremos	22
2.1.5. El Método de Máxima Verosimilitud y la Verosimilitud Perfil	24
2.2. Estimadores de Momentos Ponderados por Probabilidad	24
2.2.1. Definición y Distribución Asintótica	25
2.2.2. No Factibilidad de los Estimadores Puntuales	27
3. Estudio Basado en Simulaciones: Descripción y Resultados	29
3.1. Características del Estudio	29
3.2. Análisis de Resultados	31

3.2.1. Fallas en el Cálculo de Estimadores de Intervalo	32
3.2.2. Cobertura	35
3.2.3. Longitud de los Estimadores de Intervalo	35
Conclusiones	41
Bibliografía	44

Introducción

En muchas situaciones es de interés modelar los valores más grandes (o pequeños) que presenta un proceso estocástico. Por ejemplo, en la construcción de una presa interesa conocer el máximo nivel de agua que deberá soportar, o también, la intensidad máxima que podría tener un terremoto en la zona donde estará ubicada la presa. En general, el conocimiento de la distribución de los máximos o mínimos de algún fenómeno relevante es importante dentro de la ingeniería: “El diseño en ingeniería debe estar basado en extremos porque los valores más grandes, tales como cargas, terremotos, viento, inundaciones, olas, etc., y los valores más pequeños como fuerza, fatiga, etc. son los parámetros claves que pueden llevar a una falla en los trabajos de ingeniería.” (Castillo [1]: 4) Más ejemplos de situaciones en las que es de interés considerar eventos extremos pueden consultarse en Coles [2], pág. 1-17, y también Castillo [1], pág. 3-17.

Una forma particular de modelar eventos extremos ocurre cuando el proceso estocástico en cuestión está indexado por el tiempo, lo conforman variables aleatorias independientes e idénticamente distribuidas (i.i.d.), y tenemos como “valores más grandes” a los máximos del proceso obtenidos en periodos cuya duración de tiempo es fija. En la Teoría de Valores Extremos, un análisis en ese contexto se conoce como de máximos por bloques. Sin embargo, existen diferentes métodos que podemos aplicar para hacer inferencia en ese tipo de análisis. El objetivo principal de este documento será comparar tres de esos enfoques a través de un estudio basado en simulaciones.

En el Capítulo 1 presentamos algunos resultados clásicos de la Teoría de Valores Extremos, con los cuales encontraremos que la distribución de máximos se puede aproximar por algún tipo de las distribuciones de valores extremos (DVE) -Fréchet, Gumbel y Weibull-, las cuales podemos representar en una sola familia cuya función de distribución se conoce como distribución generalizada de valores extremos (DGVE).

Por otro lado, en el Capítulo 2 describimos los enfoques estadísticos que consideramos en este trabajo para calcular estimadores de intervalo. Específicamente, describimos el método

de máxima verosimilitud, el método de momentos ponderados por probabilidad, y finalmente, los intervalos de verosimilitud-confianza que se obtienen a partir de la verosimilitud perfil. La presentación de dichos enfoques será, primero, de manera general, pero posteriormente consideramos detalles específicos que aparecen cuando estamos trabajando con la DGVE.

Finalmente, el Capítulo 3 incluye la descripción específica del estudio basado en simulaciones que consideramos en este trabajo, así como los resultados obtenidos. También hacemos algunos comentarios sobre trabajos previos que estudian el desempeño de estimadores para muestras pequeñas en el contexto de un análisis de máximos por bloques. Dichos comentarios nos permitirán enfatizar la contribución del presente trabajo en este tema de investigación.

Capítulo 1

Teoría de Valores Extremos

En este capítulo presentamos algunos resultados clásicos de la Teoría de Valores Extremos. Un resultado fundamental es el Teorema de Tipos para Extremos, el cual, como veremos en la siguiente sección, puede usarse para sustentar métodos estadísticos que se aplican para estimar distribuciones asociadas a máximos. Además, también abordaremos el tema de los Dominios de Atracción y su caracterización.

Los teoremas que aparecen en este capítulo son sólo enunciados pero las demostraciones están dadas por Resnick [14]. Al final de cada teorema agregamos el número de proposición o corolario que le corresponde en ese documento.

1.1. Teorema de Tipos para Extremos

Sea X_1, X_2, \dots , una sucesión de variables aleatorias independientes idénticamente distribuidas (*iid*) con función de distribución común F . La teoría que presentamos a continuación se enfoca al análisis estadístico de máximos asociados a dicha sucesión, de manera específica, al análisis de

$$M_n = \max\{X_1, \dots, X_n\}, \quad n \geq 2.$$

La distribución teórica de M_n está dada por:

$$\begin{aligned} Pr\{M_n \leq z\} &= Pr\{X_1 \leq z, \dots, X_n \leq z\} \\ &= P(X_1 \leq z) \times \dots \times P(X_n \leq z) \\ &= F^n(z), \end{aligned} \tag{1.1}$$

donde $z \in \mathbf{R}$ y $n \in \mathbf{N}$. En la práctica la distribución F es desconocida, y aunque es posible estimarla a partir de los datos, variaciones pequeñas de dicha estimación, particularmente en la cola superior de F , podrían implicar diferencias sustanciales para $F^n(z)$.

Otro enfoque consiste en buscar directamente aproximaciones para la distribución de M_n . Un ejemplo de dicho enfoque se explica a continuación y se refiere a la aplicación del Teorema del Límite Central (TLC) para aproximar la distribución de la suma de variables aleatorias. Sean Y_1, Y_2, \dots una sucesión de variables aleatorias *iid* con función de distribución G . Si conocemos a G , entonces podemos obtener la distribución de $\sum_{i=1}^n Y_i$ usando la convolución, sin embargo, en la práctica no tenemos esa información. Por otro lado, suponiendo que Y_1 tiene varianza finita, entonces por el TLC se cumple que

$$\frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0, 1), \quad (1.2)$$

donde $E[Y_1] = \mu$, $Var(Y_1) = \sigma^2$ y \xrightarrow{d} denota convergencia en distribución. Si se cuenta con estimaciones para μ y σ , entonces la convergencia dada en (1.2) puede usarse para aproximar la distribución de la suma pues, con n “grande”, se tiene que

$$\frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}} \approx N(0, 1).$$

Como analogía al caso anterior, trataremos de encontrar distribuciones límites para aproximar a F^n . Primero notemos que, denotando al extremo derecho de F por

$$x_F = \sup\{x \in \mathbf{R} : F(x) < 1\},$$

se cumple que $M_n \xrightarrow{d} x_F$, es decir, la distribución de máximos converge a una distribución degenerada. La afirmación anterior se sigue de notar que para $x < x_F$, se tiene $F(x) < 1$ y

$$P(M_n \leq x) = F^n(x) \rightarrow 0, \quad n \rightarrow \infty,$$

por lo que M_n converge en probabilidad a x_F , y dado que la sucesión $\{M_n\}$ es no decreciente, entonces converge casi seguramente, y en consecuencia, también en distribución.

El argumento anterior nos muestra que es necesario estandarizar a M_n , es decir, estamos buscando garantizar una convergencia débil de la forma

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \rightarrow H(x),$$

con $\{a_n\}$ y $\{b_n\}$ sucesiones de números reales y H una distribución límite no degenerada. El Teorema de tipos para extremos, propuesto por Fisher y Tippett en 1928 y demostrado rigurosamente por Gnedenko en 1943, va en la dirección mencionada. Para enunciar ese teorema requerimos de la siguiente definición.

Definición 1.1 (Familias de distribuciones)

Se dice que dos distribuciones G y H pertenecen a la misma familia de distribuciones, o son del mismo tipo, si para algunas constantes $a > 0$ y $b \in \mathbf{R}$ se cumple

$$H(x) = G(ax + b), \quad x \in \mathbf{R}.$$

En términos de variables aleatorias, tomando $X \sim G$ y $Y \sim H$, la condición anterior es equivalente a que Y y $\frac{X-b}{a}$ tengan la misma distribución.

Teorema 1.1 (Teorema de tipos para extremos)

Supongamos que existen sucesiones de constantes $a_n > 0$ y $b_n \in \mathbf{R}$ tales que

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} H(x),$$

donde H es propia ($H(\mathbf{R}) = 1$) y no está concentrada en un punto, entonces H pertenece a alguna de las siguientes familias de distribuciones.

$$\begin{aligned} \text{Fréchet} & : \quad \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0 \end{cases} & \alpha > 0 \\ \text{Weibull} & : \quad \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha), & x < 0 \\ 1, & x \geq 0 \end{cases} & \alpha > 0 \\ \text{Gumbel} & : \quad \Lambda(x) = \exp\{-\exp\{-x\}\}, & x \in \mathbf{R} \end{aligned}$$

Una demostración puede consultarse en [14], Proposición 0.3.

Definición 1.2 (Distribuciones de valores extremos) Las funciones de distribución Φ_α , Ψ_α y Λ , como se presentaron en el Teorema 1.1, serán llamadas distribuciones de valores extremos estándar, mientras que las funciones de distribución que sean de alguno de esos tipos serán denotadas como distribuciones de valores extremos (DVE).

Nota 1.1 Las familias de funciones de distribución para las DVE se obtienen al incluir parámetros de localización y escala a sus elementos estándar. Las expresiones correspondientes son:

$$\begin{aligned}
\text{Fréchet} & : \quad \Phi_{\alpha,a,b}(x) = \begin{cases} 0, & x \leq b \\ \exp(-(\frac{x-b}{a})^{-\alpha}), & x > b \end{cases} \quad \alpha > 0; \\
\text{Weibull} & : \quad \Psi_{\alpha,a,b}(x) = \begin{cases} \exp(-(-(\frac{x-b}{a}))^\alpha), & x < b \\ 1, & x \geq b \end{cases} \quad \alpha > 0; \\
\text{Gumbel} & : \quad \Lambda_{a,b}(x) = \exp \left\{ - \exp \left\{ - \left(\frac{x-b}{a} \right) \right\} \right\}, \quad x \in \mathbf{R},
\end{aligned}$$

donde $b \in \mathbf{R}$, $a > 0$ y, para las distribuciones Fréchet y Weibull, $\alpha > 0$. El parámetro a es de escala, mientras que el parámetro b es de localización y determina, en los casos Fréchet y Gumbel, el soporte de la distribución. Finalmente, α es un parámetro de forma.

Para demostrar el Teorema 1.1 podemos aplicar el teorema de convergencia a familias de Gnedenko y Khinchin, el cual enunciamos a continuación. Incluir ese resultado es de interés pues con él se garantiza que, dada una distribución de máximos, si existen diferentes sucesiones de constantes normalizadoras con las cuales se puede obtener una distribución límite, entonces esas distribuciones límite son del mismo tipo.

Teorema 1.2 (Teorema de convergencia a familias)

Sean $G(x)$ y $H(x)$ dos distribuciones, ninguna de las cuales está concentrada en un punto. Supongamos que para $n \geq 1$ F_n es una distribución, $a_n \geq 0$, $b_n \in \mathbf{R}$, $\alpha_n \geq 0$, $\beta_n \in \mathbf{R}$ y

$$F_n(a_n x + b_n) \rightarrow G(x), \quad F_n(\alpha_n x + \beta_n) \rightarrow H(x) \quad (1.3)$$

débilmente. Entonces existen constantes $A > 0$ y $B \in \mathbf{R}$ tales que, cuando $n \rightarrow \infty$,

$$\frac{\alpha_n}{a_n} \rightarrow A, \quad \frac{\beta_n - b_n}{a_n} \rightarrow B \quad (1.4)$$

y

$$H(x) = G(Ax + B). \quad (1.5)$$

Por otro lado, si 1.4 es cierta, entonces cualquiera de las relaciones en 1.3 implica la otra y 1.5 es cierta.

Una demostración de este teorema aparece en [14], Proposición 0.2.

1.2. Dominios de Atracción

El teorema enunciado en la sección anterior identifica las distribuciones límite para máximos normalizados. A continuación revisaremos las condiciones que debe cumplir una función de distribución F para que, cuando existe una distribución límite, ésta sea de un tipo específico de las DVE .

Definición 1.3 (*Dominio de atracción*)

Una variable aleatoria X (o su función de distribución F) pertenece al dominio de atracción de la distribución de valores extremos H si existen constantes $a_n > 0$, $b_n \in \mathbf{R}$ tales que

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} H(x).$$

En ese caso se usará la notación $X \in DA(H)$ ($F \in DA(H)$).

Con la definición anterior podemos reformular el objetivo de esta sección: caracterizar el dominio de atracción para cada tipo de DVE .

1.2.1. Dominio de Atracción de las DVE Tipo Fréchet (Φ_α)

Comenzamos definiendo un tipo de funciones llamadas de variación regular y la función cuantil asociada a una función de distribución.

Definición 1.4 (*Funciones de variación regular*)

Una función medible $U : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ se dice que es de variación regular en ∞ con índice $\alpha \in \mathbf{R}$ ($U \in VR_\alpha$), si para $x > 0$

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

Definición 1.5 (*Función cuantil*)

Sea F una función de distribución, su inversa generalizada

$$Q(t) = F^{\leftarrow}(t) = \inf\{x \in \mathbf{R} : F(x) \geq t\}, \quad 0 < t < 1,$$

es llamada la función cuantil de la función de distribución F . El cuantil t de F se define como $x_t = F^{\leftarrow}(t)$.

Con las definiciones anteriores y usando la notación $\bar{F} = 1 - F$, podemos enunciar un teorema que caracteriza el dominio de atracción Fréchet ($DA(\Phi_\alpha)$).

Teorema 1.3 (*Caracterización del $DA(\Phi_\alpha)$*)

Sea F una función de distribución y $\alpha > 0$, entonces $F \in DA(\Phi_\alpha)$ si y sólo si $\bar{F} \in VR_{-\alpha}$. Además, si $F \in DA(\Phi_\alpha)$, entonces

$$\frac{M_n}{a_n} \xrightarrow{d} \Phi_\alpha, \quad (1.6)$$

con

$$a_n = Q\left(1 - \frac{1}{n}\right) = \left(\frac{1}{\bar{F}}\right)^{-}(n).$$

Una prueba aparece en [14], Proposición 1.11.

Nota 1.2 Para que $F \in DA(\Phi_\alpha)$ es condición necesaria que $x_F = \infty$. Lo anterior se obtiene notando que (1.6) es equivalente a que

$$F^n(a_n x) \rightarrow \Phi_\alpha(x)$$

débilmente, y por lo tanto, a_n debe ser no acotada.

El siguiente corolario presenta condiciones suficientes para que $F \in DA(\Phi_\alpha)$. El interés de contar con ese tipo de condiciones es que podrían ser más fáciles de verificar.

Corolario 1.1 (*Condición de Von Mises*)

Sea F una función de distribución absolutamente continua con función de densidad f que satisface

$$\lim_{x \rightarrow \infty} \frac{xf(x)}{\bar{F}(x)} = \alpha > 0,$$

entonces $F \in DA(\Phi_\alpha)$

Para indicaciones sobre la demostración ver [14], Proposición 1.15.

1.2.2. Dominio de Atracción de las DVE Tipo Weibull (Ψ_α)

La caracterización del dominio de atracción de Ψ_α , al igual que en el caso Fréchet, está relacionada con funciones de variación regular. Esa similitud es congruente con la estrecha relación que existe entre esas DVE las cuales satisfacen la relación:

$$\Psi_\alpha(-x^{-1}) = \Phi_\alpha(x), \quad x > 0.$$

Teorema 1.4 (*Caracterización del $DA(\Psi_\alpha)$*)

Sea F una función de distribución y $\alpha > 0$, entonces $F \in DA(\Psi_\alpha)$ si y sólo si $x_F < \infty$ y $\bar{F}(x_F - x^{-1}) \in VR_\alpha$. Además, si $F \in DA(\Psi_\alpha)$, entonces

$$\frac{M_n - x_F}{a_n} \xrightarrow{d} \Psi(\alpha), \tag{1.7}$$

con

$$a_n = x_F - Q\left(1 - \frac{1}{n}\right) = x_F - \left(\frac{1}{F}\right)^{\leftarrow}(n).$$

Una demostración de este teorema se puede encontrar en [14], Proposición 1.13.

Como en el caso Fréchet, si tenemos una función de densidad, entonces podemos encontrar condiciones suficientes para que $F \in DA(\Psi_\alpha)$.

Corolario 1.2 (*Condición de Von Mises*)

Sea F una función de distribución absolutamente continua con función de densidad f que es positiva en algún intervalo finito (z, x_F) . Si

$$\lim_{x \uparrow x_F} \frac{(x_F - x)f(x)}{\bar{F}(x)} = \alpha > 0,$$

entonces $F \in DA(\Psi_\alpha)$.

Para una demostración consultar [14], Proposición 1.16.

1.2.3. Dominio de Atracción de las DVE Tipo Gumbel (Λ)

Comenzamos presentando un tipo de funciones llamadas de Von Mises que, como veremos, nos permitirán presentar condiciones suficientes para que una función de distribución pertenezca al $DA(\Lambda)$. Posteriormente, enunciaremos algunos teoremas para caracterizar totalmente a ese dominio de atracción.

Definición 1.6 (Funciones de Von Mises)

Sea F una función de distribución con extremo derecho $x_F \leq \infty$. Supongamos que existe $z < x_F$ tal que F tiene representación

$$\bar{F}(x) = c \exp \left\{ - \int_z^x \frac{1}{a(t)} dt \right\}, \quad z < x < x_F,$$

donde c es alguna constante positiva, $a(\cdot)$ es una función positiva y absolutamente continua (respecto a la medida de Lebesgue) con densidad a' y $\lim_{x \uparrow x_F} a'(x) = 0$. Entonces decimos que F es una función de Von Mises y que $a(\cdot)$ es la función auxiliar de F .

El siguiente teorema afirma que una función de Von Mises pertenece al dominio de atracción de la distribución Gumbel. Además, también presenta condiciones suficientes para que una función de distribución sea de Von Mises.

Teorema 1.5 (Funciones de Von Mises y $DA(\Lambda)$)

a) Sea F una función de distribución. Si F es una función de Von Mises, entonces $F \in DA(\Lambda)$. En ese caso se tiene que

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} \Lambda, \tag{1.8}$$

con $b_n = F^{-1}(1 - \frac{1}{n})$ y $a_n = a(b_n)$, donde a es la función auxiliar de F .

b) Supongamos que F tiene extremo derecho $x_F \leq \infty$ y que existe $z < x_F$ tal que F es dos veces diferenciable en (z, x_F) con densidad positiva $f = F'$ y $F''(x) < 0$ para $z < x < x_F$. Si

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x)F''(x)}{f^2(x)} = -1, \tag{1.9}$$

entonces F es una función de Von Mises y $F \in DA(\Lambda)$. Podemos tomar como función auxiliar

$$a = \frac{\bar{F}}{f}.$$

Inversamente, una función de Von Mises dos veces diferenciable satisface (1.9).

Para una demostración consultar [14], Proposición 1.1.

Como en los otros dominios de atracción, en el caso Gumbel también se cuenta con caracterizaciones completas de su dominio de atracción.

Teorema 1.6 (Caracterización del $DA(\Lambda)$)

Sea F una función de distribución con extremo derecho $x_F \leq 0$. $F \in DA(\lambda)$ si y sólo si existe algún $z < x_F$ tal que F tiene representación

$$\bar{F}(x) = c(x) \exp \left\{ - \int_z^x \frac{g(t)}{a(t)} dt \right\}, \quad z < x < x_F, \quad (1.10)$$

donde c y g son funciones medibles que satisfacen $c(x) \rightarrow c > 0$, $g(x) \rightarrow 1$ cuando $x \uparrow x_F$, y $a(x)$ es una función positiva y absolutamente continua respecto a la medida de Lebesgue con densidad $a'(x)$ que satisface $\lim_{x \uparrow x_F} a'(x) = 0$. Para F con representación (1.10) podemos escoger

$$b_n = F^{\leftarrow} \left(1 - \frac{1}{n} \right) \quad \text{y} \quad a_n = a(b_n)$$

como constantes normalizadoras. Una elección posible para la función a es

$$a(x) = \int_x^{x_F} \frac{\bar{F}(t)}{\bar{F}(x)} dt, \quad x < x_F.$$

Una demostración aparece en [14], Corolario 1.7 y Proposición 1.9.

Nota 1.3 Un procedimiento útil para determinar el dominio de atracción y las constantes normalizadoras usa transformaciones monótonas. Sea X_1, X_2, \dots , una sucesión de variables aleatorias y g una función creciente. Definiendo $\tilde{X} = g(X)$ se tiene que

$$\tilde{M}_n = \max(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n) = g(M_n).$$

Si $X \in DA(\Lambda)$ con

$$\lim_{n \rightarrow \infty} P(M_n \leq a_n x + b_n) = \Lambda(x), \quad x \in \mathbf{R},$$

entonces

$$\lim_{n \rightarrow \infty} P(\tilde{M}_n \leq g(a_n x + b_n)) = \Lambda(x), \quad x \in \mathbf{R},$$

En algunos casos, es posible calcular la expansión de Taylor para g alrededor de b_n , y los términos lineales son suficientes para obtener la distribución límite de \tilde{M}_n considerando como constantes normalizadoras a $\tilde{a}_n = a_n g'(b_n)$ y $\tilde{d}_n = g(b_n)$.

1.3. Distribución Generalizada de Valores Extremos

Las familias de distribuciones Fréchet, Gumbel y Weibull pueden ser combinadas dentro de una familia de modelos que se presenta en la siguiente definición.

Definición 1.7 (*Distribución generalizada de valores extremos*)

Consideremos la familia de distribuciones con parámetros $\mu \in \mathbf{R}$, $\sigma > 0$ y $\xi \in \mathbf{R}$ dada por

$$G(z) = \begin{cases} \exp -[1 + \xi(\frac{z-\mu}{\sigma})]^{-\frac{1}{\xi}}, & \xi \neq 0; \\ \exp[-\exp -(\frac{z-\mu}{\sigma})], & \xi = 0, \end{cases} \quad (1.11)$$

definida en

$$\begin{cases} \{z : 1 + \xi(\frac{z-\mu}{\sigma}) > 0\}, & \xi \neq 0; \\ -\infty < z < \infty, & \xi = 0. \end{cases}$$

Ésta es la familia de distribuciones de valores extremos generalizada y nos referiremos a cualquiera de sus elementos como una distribución generalizada de valores extremos (*DGVE*).

Una *DGVE* tiene tres parámetros: un parámetro de localización, μ ; un parámetro de escala, σ ; y un parámetro de forma, ξ . Además, para $\xi \neq 0$ el soporte está determinado por una combinación de dichos parámetros, en particular, para $\xi > 0$ la distribución tiene un extremo izquierdo finito, mientras que para $\xi < 0$ el extremo derecho es finito.

Sea G una *DGVE* con parámetros (μ, σ, ξ) . Si $\xi > 0$, entonces G es de tipo Fréchet, para $\xi < 0$ es de tipo Weibull, y finalmente, si $\xi = 0$, entonces es de tipo Gumbel. Además, es posible determinar los parámetros (a, b, α) que le corresponden a G cuando se escribe como una *DVE*, ésta última parametrizada como aparece en la nota 1.1.

$$\begin{aligned} \text{Cuando } \xi > 0 : & \begin{cases} b = \mu - \frac{\sigma}{\xi}; \\ a = \frac{\sigma}{\xi}; \\ \alpha = \frac{1}{\xi}. \end{cases} \\ \text{Cuando } \xi < 0 : & \begin{cases} b = \mu - \frac{\sigma}{\xi}; \\ a = -\frac{\sigma}{\xi}; \\ \alpha = -\frac{1}{\xi}. \end{cases} \end{aligned}$$

Y, finalmente, para $\xi = 0$ se tiene $b = \mu$ y $a = \sigma$.

En consecuencia, además de identificar con el signo de ξ el tipo de *DVE* a la que corresponde una *DGVE*, también podemos determinar el elemento estándar con el que está

relacionada.

$$\begin{cases} \xi = \alpha^{-1} > 0, & \text{corresponde a una distribución de tipo Fréchet } \Phi_\alpha, \\ \xi = 0, & \text{corresponde a una distribución de tipo Gumbel } \Lambda, \\ \xi = -\alpha^{-1} < 0, & \text{corresponde a una distribución de tipo Weibull } \Psi_\alpha. \end{cases} \quad (1.12)$$

1.4. Niveles y periodos de Retorno

Dentro del análisis de eventos extremos aparece la siguiente pregunta: ¿Cuál es el tiempo de espera promedio para que ocurra un evento extremo específico?

Sea X_1, X_2, \dots , una sucesión de variables aleatorias *iid* con función de distribución continua F y $u \in \mathbf{R}$ un umbral dado. Consideraremos el evento $\{X_i > u\}$ como un evento extremo y definimos el tiempo de la primera excedencia como

$$T(u) = \min\{i \geq 1 : X_i > u\}.$$

Con el modelo anterior, la respuesta a la pregunta planteada es $E[T(u)]$. Además, tomando $p = 1 - F(u)$, obtenemos

$$P(T(u) = k) = (1 - p)^{k-1}p, \quad k = 1, 2, \dots,$$

por lo tanto, $T(u)$ es una variable aleatoria geométrica y $E[T(u)] = p^{-1}$. Definimos a u como el **nivel de retorno** con un **periodo de retorno** p^{-1} para los eventos $\{X_i > u\}$. Nótese que u es el cuantil de F con un nivel de probabilidad $1 - p$.

Aunque en el modelo anterior partimos de un umbral que definía un evento extremo al cual se le calculó su tiempo promedio de ocurrencia, también podemos proceder de manera inversa. Supongamos que buscamos los eventos extremos del tipo $\{X_i > u\}$ que tarden en ocurrir en promedio un tiempo igual a p^{-1} , entonces basta tomar como nivel de retorno a $u = x_p$, donde $F(x_p) = 1 - p$.

1.5. Dominios de Atracción para Algunas Distribuciones Conocidas

En los Cuadros 1.1, 1.2 y 1.3, aparecen clasificadas algunas familias de distribuciones según el tipo de dominio de atracción al que pertenecen. Para los tipos Fréchet y Weibull es

Nombre	Distribución F o densidad f	Parámetros	Soporte	ξ
Burr	$F(x) = 1 - \left(\frac{\kappa}{\kappa+x\tau}\right)^\alpha$	$\alpha, \kappa, \tau > 0$	$x > 0$	$\frac{1}{\alpha\tau}$
Pareto	$F(x) = 1 - \left(\frac{\kappa}{\kappa+x}\right)^\alpha$	$\alpha, \kappa > 0$	$x > 0$	$\frac{1}{\alpha}$
Loggamma	$f(x) = \frac{\alpha^\beta}{\Gamma(\beta)} (\ln(x))^{\beta-1} x^{-\alpha-1}$	$\alpha, \beta > 0$	$x > 1$	$\frac{1}{\alpha}$
Fréchet	$F(x) = \exp\{-x^{-\alpha}\}$	$\alpha > 0$	$x > 0$	$\frac{1}{\alpha}$
Cauchy	$f(x) = \left[\pi a \left(1 + \left(\frac{x-b}{a}\right)^2\right)\right]^{-1}$	$b \in R, a > 0$	$-\infty < x < \infty$	1

Cuadro 1.1: Distribuciones que pertenecen al dominio de atracción Fréchet

posible determinar el parámetro α que determina a Φ_α y Ψ_α , sin embargo, se usó la transformación (1.12) y presentamos el valor para ξ correspondiente. Finalmente, hacemos algunos comentarios sobre la notación que usamos. Dentro de las distribuciones consideradas, la que denotamos como “Weibull_m” hace referencia a la distribución Weibull de dos parámetros para mínimos -la notación que usamos en este caso coincide con la de Castillo [1]-. Por otro lado, denotamos como “Weibull (DVE)” a la distribución Weibull que corresponde a una de las DVE.

Demostraciones de que las funciones de distribución que aparecen en los Cuadros 1.1, 1.2 y 1.3 pertenecen al dominio de atracción señalado, están dadas por Embrechts, Klüpelberg y Mikosch [5]. Además, información general sobre dichas distribuciones puede encontrarse en Hogg y Klugman [8], o bien, para un análisis más detallado, se puede consultar Johnson y Kotz [10] y [11].

Nombre	Distribución F o densidad f	Parámetros	Soporte
Weibull $_m$	$F(x) = 1 - \exp\{-cx^\tau\}$	$c, \tau > 0$	$x > 0$
Lognormal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right\}$	$\mu \in R, \sigma > 0$	$x > 0$
Exponencial	$F(x) = 1 - \exp\{-\lambda x\}$	$\lambda > 0$	$x > 0$
Gamma	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}$	$\alpha, \beta > 0$	$x > 0$
Logística	$F(x) = 1 - [1 + \exp\{\frac{x-b}{a}\}]^{-1}$	$b \in R, a > 0$	$-\infty < x < \infty$
Normal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$	$\mu \in R, \sigma > 0$	$-\infty < x < \infty$

Cuadro 1.2: Distribuciones que pertenecen al dominio de atracción Gumbel

Nombre	Distribución F o densidad f	Parámetros	Soporte	ξ
Beta	$f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1} (1-x)^{\beta-1}$	$\alpha, \beta > 0$	$0 < x < 1$	$-\frac{1}{\beta}$
Weibull (DVE)	$F(x) = \exp\{-(-x)^\alpha\}$	$\alpha > 0$	$x > 0$	$-\frac{1}{\alpha}$
Uniforme	$F(x) = \frac{x-a}{b-a}$	$a < b$	$a \leq x \leq b$	-1

Cuadro 1.3: Distribuciones que pertenecen al dominio de atracción Weibull

Capítulo 2

Métodos de Estimación en el Análisis de Máximos por Bloques

Sea $\{X_i\}_{i \geq 1}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas (*iid*), cuya distribución común no es conocida. Definimos

$$M_n = \text{máx} \{X_1, \dots, X_n\}, \quad n \geq 2.$$

El problema de interés es determinar el comportamiento estadístico de M_n .

El Teorema 1.1 afirma que las únicas distribuciones límite posibles para M_n son las *DVE*. Es posible adoptar un enfoque basado en el uso de esas distribuciones límite como aproximaciones a un nivel finito. Específicamente, supongamos que existen sucesiones de constantes $\{a_n > 0\}$ y $\{b_n\}$ de modo que

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow G(x)$$

débilmente, con G una *DVE*. Para n suficientemente grande

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \approx G(x),$$

o equivalentemente, con $z = a_n x + b_n$,

$$\begin{aligned} P(M_n \leq z) &\approx G\left(\frac{z - b_n}{a_n}\right) \\ &= G_n^*(z), \end{aligned} \tag{2.1}$$

donde G_n^* es otra *DVE* del mismo tipo que G , pero con diferentes parámetros de localización y escala. Por lo tanto, se ha encontrado que la distribución de M_n se puede aproximar por una *DVE*. Además, dado que los parámetros de la distribución serán estimados, es irrelevante si los parámetros de la distribución G son diferentes de los de G_n^* , y por lo tanto, no es un problema que sean desconocidas las constantes de normalización.

El argumento presentado nos conduce al siguiente enfoque para modelar extremos en la práctica. Sea x_1, x_2, \dots, x_m una muestra de alguna variable aleatoria X . Dividimos la muestra en bloques de n observaciones formando los vectores

$$\begin{aligned} x^{(1)} &= (x_1^1, \dots, x_n^1), \\ x^{(2)} &= (x_1^2, \dots, x_n^2), \\ &\vdots \\ x^{(k)} &= (x_1^k, \dots, x_n^k). \end{aligned}$$

La muestra *iid* con respecto a la que realizaremos inferencias consiste de

$$M_{n,i} = \max \{x_1^i, \dots, x_n^i\}, \quad i = 1, \dots, k.$$

El enfoque descrito se conoce en la literatura como análisis de máximos por bloques. En ciertas áreas, como en el análisis de datos ambientales, es usual que los bloques correspondan con un periodo de tiempo igual a un año. En ese caso, n es el número de observaciones consideradas en un año y los máximos por bloques son llamados máximos anuales.

Una vez que contamos con la muestra de máximos, podemos aplicar herramientas estadísticas para obtener inferencias, en particular, son de interés aquellas que se refieren a cuantiles altos de la distribución de máximos. En este capítulo describimos tres tipos de estimadores para dichos cuantiles los cuales están basados, respectivamente, en el método de máxima verosimilitud, la verosimilitud perfil, y los estimadores de momentos ponderados por probabilidad.

2.1. Herramientas Basadas en la Verosimilitud

En esta sección describimos el **método de máxima verosimilitud** y cómo obtener estimaciones basadas en la **verosimilitud perfil**. Para ello, requerimos de algunos preliminares los cuales presentamos, primero, en un contexto general (Secciones 2.1.1, 2.1.2 y 2.1.3), y luego, para el caso particular en que el modelo paramétrico de interés es la distribución de valores extremos (Sección 2.1.4).

2.1.1. Estimadores de Máxima Verosimilitud

Sea x_1, \dots, x_n una muestra de variables aleatorias *iid* cuya distribución común F_θ es desconocida y que pertenece a una familia paramétrica

$$\mathcal{F} = \{F_\theta : \theta \in \Theta\},$$

donde $\Theta \subset \mathbf{R}^k$, para algún $k \in \mathbf{N}$. Supongamos que las distribuciones F_θ tienen densidades $f(x; \theta)$. Para dar una estimación puntual del valor de θ podemos usar el estimador de máxima verosimilitud (*emv*) que definimos a continuación.

Cada valor de $\theta \in \Theta$ define un modelo potencial, y en consecuencia, determina una función de densidad conjunta posible de la cual x_1, \dots, x_n es una realización. Asociando a cada vector θ con el número real que se obtiene al evaluar la densidad conjunta correspondiente en (x_1, \dots, x_n) , definimos una relación que se denota como **función de verosimilitud**. Por la independencia de las realizaciones, se sigue que esa función tiene la forma

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta),$$

donde $\mathbf{x} = (x_1, \dots, x_n)$. Valores de θ que tienen una verosimilitud alta corresponden a modelos que le dan alta plausibilidad a los datos observados. El principio de estimación puntual por máxima verosimilitud consiste en adoptar el modelo con la verosimilitud más alta, por lo tanto, el **estimador de máxima verosimilitud** para θ es cualquier elemento de Θ , $\hat{\theta}$, que maximiza L . Muchas veces la estimación $\hat{\theta}$ puede obtenerse resolviendo el sistema de **ecuaciones de verosimilitud**,

$$\left. \frac{\partial \log L}{\partial \theta_i} \right|_{\theta=\hat{\theta}} = 0, \quad (i = 1, \dots, k),$$

y confirmando que la solución $\hat{\theta}$ maximiza L . Por otro lado, también es usual tomar logaritmos y trabajar con la **función de log-verosimilitud**

$$\ell(\theta) = \sum_{i=1}^n \log f(x_i; \theta).$$

Dado que la función logaritmo es monótona es equivalente maximizar ℓ para obtener $\hat{\theta}$.

Bajo ciertas condiciones de regularidad, los estimadores de máxima verosimilitud tienen algunas propiedades asintóticas como se muestra en el siguiente teorema. En lo que sigue incluiremos en la notación de los estimadores de máxima verosimilitud un subíndice n el cual hace referencia al tamaño de la muestra.

Teorema 2.1 Sea $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$ una familia de funciones de distribuciones, donde Θ es un subconjunto abierto de \mathbf{R} . Supongamos que las distribuciones F_θ poseen funciones de densidad $f(x; \theta)$ y que la familia \mathcal{F} satisface las siguientes condiciones de regularidad.

1. Para cada $\theta \in \Theta$, las derivadas

$$\frac{\partial \log [f(x, \theta)]}{\partial \theta}, \quad \frac{\partial^2 \log [f(x, \theta)]}{\partial \theta^2}, \quad \frac{\partial^3 \log [f(x, \theta)]}{\partial \theta^3}$$

existen para toda x .

2. Para cada $\theta_0 \in \Theta$, existen funciones $g(x)$, $h(x)$, y $m(x)$ (que posiblemente dependen de θ_0) tales que para θ en una vecindad $B(\theta_0)$ las relaciones

$$\left| \frac{\partial \log [f(x, \theta)]}{\partial \theta} \right| \leq g(x), \quad \left| \frac{\partial^2 \log [f(x, \theta)]}{\partial \theta^2} \right| \leq g(x), \quad \left| \frac{\partial^3 \log [f(x, \theta)]}{\partial \theta^3} \right| \leq g(x)$$

son ciertas para toda x , y

$$\int_{\mathbf{R}} g(x) dx < \infty, \quad \int_{\mathbf{R}} h(x) dx < \infty, \quad \int_{\mathbf{R}} m(x) f(x; \theta) dx < \infty$$

para $\theta \in B(\theta_0)$.

3. Para cada $\theta \in \Theta$,

$$0 < E_\theta \left\{ \left(\frac{\partial \log [f(x; \theta)]}{\partial \theta} \right)^2 \right\} < \infty.$$

Bajo las condiciones anteriores, consideremos observaciones iid de F_θ , para θ un elemento de Θ . Entonces, con probabilidad 1, las ecuaciones de verosimilitud admiten una sucesión de soluciones $\{\hat{\theta}_n\}$ tales que

$$(\hat{\theta}_n - \theta) \sqrt{n \iota_\theta} \xrightarrow{d} N(0, 1),$$

donde $\iota_\theta = -E_\theta(\partial^2 \log f(\theta; x_i) / \partial \theta^2)$ es la información esperada de una sola observación, y $\mathcal{I}_\theta = n \iota_\theta$ es la función de información esperada de Fisher. Es decir, el estimador de máxima verosimilitud es asintóticamente normal.

Además, con apropiadas generalizaciones de las condiciones anteriores, el resultado sigue siendo válido para el caso multidimensional. Si suponemos $\theta = (\theta_1, \dots, \theta_k)$, entonces, con probabilidad 1, las ecuaciones de verosimilitud admiten una sucesión de soluciones $\{\hat{\theta}_n\}$ tales que

$$(\hat{\theta}_n - \theta) \sqrt{n} \xrightarrow{d} N_k(0, \iota(\theta)^{-1}),$$

donde $\iota_\theta = -E_\theta \{ \partial^2 \log f(\theta; x_i) / \partial \theta_j \partial \theta_m \}_{k \times k}$ es la matriz de información esperada para una sola observación y $\mathcal{I}_\theta = n \iota_\theta$ es la matriz de información esperada de Fisher.

Una demostración puede consultarse en Serfling [15], Sección 4.4.2, pág. 152.

Aunque una familia paramétrica F pueda ser indexada por un parámetro θ , puede ocurrir que el interés sea estimar alguna función $\psi = h(\theta)$, donde ψ puede tener una dimensión distinta a θ . Los siguientes dos resultados permiten usar las inferencias para θ y obtener las correspondientes para ψ .

Teorema 2.2 *Si $\hat{\theta}$ es el estimador de máxima verosimilitud para θ , y $\psi = h(\theta)$ es una función escalar, entonces el estimador de máxima verosimilitud de ψ está dado por $\hat{\psi} = h(\hat{\theta})$.*

Teorema 2.3 *Bajo las condiciones de regularidad del Teorema 2.1, el estimador de máxima verosimilitud $\hat{\theta}$ es asintóticamente normal y denotamos a su matriz de covarianza asintótica como V_θ . Si $\psi = h(\theta)$ es una función escalar, entonces el estimador de máxima verosimilitud $\hat{\psi} = h(\hat{\theta})$ tiene una distribución aproximada $N(\psi, V_\psi)$, donde*

$$V_\psi = \nabla \psi^t V_\theta \nabla \psi$$

con

$$\nabla \psi = \left[\frac{\partial \psi}{\partial \theta_1}, \dots, \frac{\partial \psi}{\partial \theta_k} \right]$$

evaluado en $\hat{\theta}$.

La demostración de este resultado es una aplicación directa del método delta, una prueba puede encontrarse en Shao [16], pág. 61.

2.1.2. Intervalos de Verosimilitud y Aproximación χ^2

Para comparar la plausibilidad asociada a los diferentes valores de θ podemos usar la **función de verosimilitud relativa** que se define como

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta}_n)}.$$

Al conjunto $\{\theta : R(\theta) \geq l\}$, con $l \in (0, 1)$, lo llamaremos **region de verosimilitud**, mientras que a l lo denotaremos como **nivel de verosimilitud**. Para el caso en que la dimensión de Θ es igual a 1 esas regiones reciben el nombre de **intervalos de verosimilitud**.

Las afirmaciones que hagamos usando regiones de verosimilitud se refieren sólo a la plausibilidad relativa para los valores de θ , y no a la incertidumbre del intervalo. Aunque, en algunos casos, es posible añadir a las regiones de verosimilitud la probabilidad de que contengan el parámetro verdadero. Una forma de hacerlo es usando la función

$$D_n(\theta) = 2(\ell(\hat{\theta}_n) - \ell(\theta)) = -2 \log(R(\theta)),$$

porque, como se muestra en el siguiente teorema, bajo ciertas condiciones de regularidad es posible determinar su distribución asintótica.

Teorema 2.4 *Sea x_1, \dots, x_n una muestra de variables aleatorias iid con función de distribución F_θ que pertenece a una familia $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$, donde Θ es un subconjunto abierto de \mathbf{R}^k . Supongamos que se satisfacen las condiciones de regularidad del Teorema 2.1. Entonces*

$$D_n(\theta) \xrightarrow{d} \chi_k^2.$$

Una demostración aparece en Serfling [15], Sección 4.4.3, pág. 155.

Del Teorema 2.4 se sigue que una región de confianza con probabilidad aproximada de $(1 - \alpha)$ está dada por

$$C_\alpha = \{\theta : D(\theta) \leq c_\alpha\}, \quad (2.2)$$

donde c_α es el cuantil $(1 - \alpha)$ de la distribución χ_d^2 . (ver, por ejemplo, Shao [16] pág. 497) Escribiendo (2.2) en términos de la verosimilitud relativa, obtenemos

$$C_\alpha = \{\theta : R(\theta) \geq \exp(-\frac{1}{2}c_\alpha)\},$$

y por lo tanto, cuando es posible aplicar el Teorema 2.4, se cumple que una región de verosimilitud con nivel $\exp(-\frac{1}{2}c_\alpha)$ tiene una probabilidad de cobertura aproximada de $(1 - \alpha)$. En el Cuadro 2.1 se presentan algunos ejemplos de esa relación para el caso $d = 1$.

2.1.3. Verosimilitud Perfil

Consideremos un espacio parametral Θ de dimensión k con elementos denotados por $\theta = (\theta_1, \dots, \theta_d)$. Supongamos que θ se particiona en dos componentes $(\theta^{(1)}, \theta^{(2)})$ y deseamos trabajar sólo con los parámetros incluidos en $\theta^{(1)}$. En ese caso, podemos usar la verosimilitud perfil que se define como

$$L_p(\theta^{(1)}) = \max_{\theta^{(2)} | \theta^{(1)}} L(\theta^{(1)}, \theta^{(2)}),$$

Nivel de Verosimilitud		Nivel de Confianza
0.0362	-	0.99
0.1465	-	0.95
0.2585	-	0.90
0.7965	-	0.50

Cuadro 2.1: Correspondencia entre niveles de verosimilitud y confianza usando la aproximación χ^2 con $d = 1$.

donde L es la función de verosimilitud. En palabras, la verosimilitud perfil se obtiene maximizando la función de verosimilitud evaluada en los elementos de Θ con $\theta^{(1)}$ fijo. Denotando por $\hat{\theta}_n$ al estimador de máxima verosimilitud, entonces definimos a la verosimilitud relativa perfil para $\theta^{(1)}$ como

$$R_p(\theta^{(1)}) = \frac{L_p(\theta^{(1)})}{L(\hat{\theta}_n)}.$$

Con la anterior formamos los conjuntos $\{\theta : R_p(\theta^{(1)}) \geq l\}$ que llamaremos regiones de verosimilitud para $\theta^{(1)}$ de nivel l , a las cuales, de modo similar al caso anterior, es posible asociar un nivel de confianza usando el siguiente teorema.

Teorema 2.5 *Sea x_1, \dots, x_n una muestra de variables aleatorias iid con función de distribución F_θ que pertenece a una familia $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$ con espacio parametral Θ de dimensión k . Supongamos que se satisfacen las condiciones del Teorema 2.1. Sea $\hat{\theta}_n$ el estimador de máxima verosimilitud para el parámetro del modelo $\theta = (\theta^{(1)}, \theta^{(2)})$, donde $\theta^{(1)}$ es un subconjunto de tamaño r de las componentes de θ . Entonces*

$$-2 \log(R_p(\theta^{(1)})) \xrightarrow{d} \chi_r^2.$$

Una demostración aparece en Serfling [15], Sección 4.4.4, pág. 158.

2.1.4. Aplicación a las Distribuciones de Valores Extremos

La función de log-verosimilitud bajo la *DGVE*, con la parametrización (1.11), esta dada por

$$\ell(\mu, \sigma, \xi; \mathbf{x}) = \begin{cases} \left(-n \log(-\xi\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log(y_i) - \sum_{i=1}^n (y_i)^{-\frac{1}{\xi}}\right) \mathbf{I}_{\{1+\xi y_i > 0\}}, & \xi \neq 0; \\ -n \log(\sigma) - \sum_{i=1}^n \exp\{-y_i\} - \sum_{i=1}^n y_i, & \xi = 0, \end{cases} \quad (2.3)$$

con $y_i = \left(\frac{x_i - \mu}{\sigma}\right)$.

Como el soporte de la función de densidad depende de los parámetros, entonces no se satisfacen las condiciones usuales de regularidad que garantizan consistencia, eficiencia y normalidad asintótica de los estimadores de máxima verosimilitud. Sin embargo, Smith [18] demostró que, a pesar de que no se satisfacen las condiciones usuales, los estimadores de máxima verosimilitud tienen las propiedades mencionadas cuando $\xi > -0.5$. Además, para $-1 < \xi < -0.5$, los estimadores se pueden obtener pero no tienen esas propiedades asintóticas. Por último, cuando $\xi < -1$ es posible que no se puedan obtener los estimadores de máxima verosimilitud.

Con las funciones de verosimilitud anteriores podemos obtener estimadores de máxima verosimilitud, así como funciones de verosimilitud perfil para los parámetros. Adicionalmente, aplicando el Teorema 2.2 también podemos obtener estimadores de máxima verosimilitud para niveles de retorno. Para ello, notemos que la función de distribución de la *DGVE* es invertible, y aplicando dicha inversa a $1-p$, con $p \in (0, 1)$, obtenemos el estimador de máxima verosimilitud para el nivel de retorno asociado a un periodo de retorno p^{-1} . La expresión específica para la inversa de la distribución, evaluadas en $(1-p)$, es

$$\begin{cases} x_p = \mu + \frac{\sigma}{\xi} \left([-\log(1-p)]^{-\xi} - 1 \right), & \xi \neq 0; \\ x_p = \mu - \sigma \log(-\log(1-p)), & \xi = 0. \end{cases}$$

Usando la expresión anterior también podemos reparametrizar las funciones de verosimilitud y obtener verosimilitudes perfiles para los niveles de retorno. Por ejemplo, para $\xi \neq 0$, y después de escoger un periodo de retorno p^{-1} , podemos despejar σ de la expresión para la inversa de la distribución obteniendo

$$\sigma = \xi (x_p - \mu) \left([-\log(1-p)]^{-\xi} - 1 \right)^{-1}.$$

Al sustituir la anterior en (2.3) obtenemos una parametrización en términos de (μ, x_p, ξ) .

2.1.5. El Método de Máxima Verosimilitud y la Verosimilitud Perfil

Dentro de los enfoques para hacer inferencia que están basados en los conceptos y resultados presentados en esta sección, uno que aparece usualmente en la literatura, y que siguiendo a Castillo [1] denominaremos **método de máxima verosimilitud**, tiene las siguientes características. Supongamos que tenemos una muestra de máximos por bloques. Para ajustar una DGVE utilizamos el estimadores de máxima verosimilitud $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})$. Evaluando la inversa de la DGVE en $1 - p$, donde $p \in [0, 1]$, obtenemos el estimador de máxima verosimilitud para el nivel de retorno con periodo de retorno igual a p^{-1} . Por otro lado, para obtener estimadores de tipo intervalo tanto para los parámetros como para los niveles de retorno, utilizamos la normalidad asintótica de los estimadores de máxima verosimilitud. Como ejemplo, consideremos un nivel de retorno $u_p = G^{-1}(1 - p; \theta)$, donde G es la DGVE. En ese caso, con el método de máxima verosimilitud obtenemos el estimador puntual $\hat{u}_p = G^{-1}(1 - p; \hat{\theta})$, y usando el Teorema 2.3 con $h = G^{-1}$, obtenemos como **intervalo de confianza aproximada** del $(1 - \alpha)100\%$ a

$$\left(\hat{u}_p \pm z_{\alpha/2} \sqrt{\nabla u_p^t V_{\theta} \nabla u_p} \right),$$

donde

$$\nabla x_p = \begin{pmatrix} \frac{\nabla u_p}{\nabla \mu} & \frac{\nabla u_p}{\nabla \sigma} & \frac{\nabla u_p}{\nabla \xi} \end{pmatrix}$$

y $z_{\alpha/2}$ es el $(1 - \alpha/2)$ cuantil de la distribución normal estándar.

Una manera alternativa para encontrar estimadores de intervalo consiste en utilizar los intervalos de verosimilitud que se obtienen a partir de la función de verosimilitud perfil. Cuando se combinan dichos intervalos con la aproximación χ^2 obtenemos **intervalos de verosimilitud confianza**. Como menciona Coles [2], ésta aproximación es usualmente más exacta que la basada en la normalidad asintótica del estimador de máxima verosimilitud.

2.2. Estimadores de Momentos Ponderados por Probabilidad

Los estimadores de momentos ponderados por probabilidad, propuestos por Greenwood et. al. [7], son una generalización de los obtenidos usando el método de momentos. En esta sección mostramos un breve resumen de la teoría sobre dichos estimadores y cómo pueden ser usados para obtener estimaciones puntuales de parámetros y cuantiles de la DGVE. Además,

estos estimadores tienen la propiedad de ser asintóticamente normales, por lo que es posible obtener estimadores de intervalo siguiendo un razonamiento análogo al que se usa bajo el método de máxima verosimilitud.

2.2.1. Definición y Distribución Asintótica

Los momentos ponderados por probabilidad de una variable aleatoria X , con función de distribución $F(x) = P[X \leq x]$, son definidos como

$$M_{p,r,s} = E[X^p \{F(X)\}^r \{1 - F(X)\}^s],$$

donde p, r , y s son numeros reales. Para el método de momentos ponderados por probabilidad, consideramos los momentos de la forma

$$\beta_r = M_{1,r,0} = E[X \{F(X)\}^r]$$

con $r = 0, 1, 2, \dots$. La simplificación anterior coincide con la usada por Hosking et al. [9], quien argumenta que los momentos de la forma $M_{1,r,s}$ son preferibles para estimar los parámetros de la distribución de X por que, en ese caso, sólo consideramos la primera potencia de X , lo que significa que la relación entre los parámetros y momentos podría tener una forma más simple. Además, cuando r y s son números enteros, $F^r(1 - F)^s$ puede ser expresado como potencias de F ó $(1 - F)$, y por lo tanto, es natural considerar los momentos $M_{1,r,0}$ con $r = 0, 1, 2, \dots$ ó $M_{1,0,s}$ con $s = 0, 1, 2, \dots$. Por otro lado, dada una muestra aleatoria x_1, x_2, \dots, x_n de la distribución F , Landwehr et al. [12] probaron que la estadística

$$b_r = n^{-1} \sum_{j=1}^n \frac{(j-1)(j-2) \cdots (j-r)}{(n-1)(n-2) \cdots (n-r)} x_j$$

es un estimador insesgado de β_r . Finalmente, para calcular los estimadores de momentos ponderados por probabilidad, procedemos de modo similar al conocido método de momentos, específicamente, denotando por k la dimensión del parametro de interés, los estimadores de momentos ponderados por probabilidad son obtenidos resolviendo el sistema $b_r = \beta_r$, con $r = 0, 1, \dots, (k-1)$.

Ahora consideramos el caso particular en que F es la distribución generalizada de valores extremos (DGVE) con la parametrización dada en (1.11). Los momentos ponderados por probabilidad para $\xi \neq 0$ tienen la forma

$$\beta_r = (r+1)^{-1} [\mu - \sigma \{1 - (r+1)^\xi \Gamma(1 - \xi)\} / \xi] \quad (2.4)$$

cuando $\xi < 1$. Para $\xi \geq 1$ ninguno de los momentos β_r existen. Una prueba de la relación (2.4) aparece en el Apéndice A en Hosking et al. [9]. La DGVE tiene tres parámetros por lo que consideramos el sistema de 3 ecuaciones

$$\beta_0 = \mu - \sigma \{1 - \Gamma(1 - \xi)\} / \xi, \quad (2.5)$$

$$2\beta_1 - \beta_0 = \sigma \Gamma(1 - \xi)(2^\xi - 1) / \xi, \quad (2.6)$$

y

$$(3\beta_2 - \beta_0) / (2\beta_1 - \beta_0) = (1 - 3^\xi) / (1 - 2^\xi). \quad (2.7)$$

Los estimadores de momentos ponderados por probabilidad, $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ son la solución de las ecuaciones (2.5)-(2.7) cuando reemplazamos β_r por b_r . Para estimar ξ debemos resolver la ecuación

$$(3b_2 - b_0) / (2b_1 - b_0) = (1 - 3^\xi) / (1 - 2^\xi).$$

La solución exacta puede obtenerse usando métodos iterativos, aunque una alternativa es usar la aproximación

$$\hat{\xi} = 7.8590c + 2.9554c^2, \quad c = \frac{2b_1 - b_0}{3b_2 - b_0} - \frac{\log 2}{\log 3} \quad (2.8)$$

pues, como menciona Hosking et al. [9], el error en $\hat{\xi}$ al usar (2.8) es menor que 0.0009 cuando $\xi \in (-1/2, 1/2)$. Finalmente, dado $\hat{\xi}$, usamos las ecuaciones (2.5) y (2.6) para obtener los estimadores de los parámetros de localización y escala

$$\hat{\sigma} = \frac{-(2b_1 - b_0)\hat{\xi}}{\Gamma(1 - \hat{\xi})(1 - 2^\xi)}, \quad \hat{\mu} = b_0 - \hat{\sigma} \left\{ \Gamma(1 - \hat{\xi}) - 1 \right\} / \hat{\xi}. \quad (2.9)$$

Los estimadores de momentos ponderados por probabilidad tienen la propiedad de ser asintóticamente normales. Dicha propiedad es una consecuencia de la normalidad asintótica del vector $b = (b_0, b_1, b_2)^T$ con media $\beta = (\beta_0, \beta_1, \beta_2)^T$ y matriz de covarianza $n^{-1}V$. Hosking et al. [9] mostraron que, para $\xi < 1/2$, los elementos de V están dados por

$$\begin{aligned} v_{rr} &= \mu^2 \xi^{-2} (r+1)^{2\xi} \left[\Gamma(1 - 2\xi) G(r/(r+1)) - \{\Gamma(1 - \xi)\}^2 \right], \\ v_{r,r+1} &= \frac{1}{2} \mu^2 \xi^{-2} \left[(r+2)^{2\xi} \Gamma(1 - 2\xi) G(r/(r+2)) \right. \\ &\quad \left. + (r+1)^\xi \{(r+1)^\xi - 2(r+2)^\xi\} \Gamma(1 - \xi)^2 \right], \end{aligned}$$

$$v_{r,r+s} = \frac{1}{2}\mu^2\xi^{-2} [(r+s+1)^{2\xi}\Gamma(1-2\xi)G(r/(r+s+1)) - (r+s)^\xi\Gamma(1-2\xi)G((r+1)/(r+s)) + 2(r+1)^\xi \times \{(r+s)^\xi - (r+s+1)^\xi\} \Gamma(1-\xi)^2], \quad s \geq 2,$$

donde G denota la función hipergeométrica de Gauss $G(x) = F(-\xi, -2\xi; 1-\xi; -x)$. Por otro lado, sean $\theta = (\mu, \sigma, \xi)^T$ y $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})^T$. Usando las ecuaciones (2.8) y (2.9) obtenemos la relación $\hat{\theta} = f(b)$, y usando el método delta, concluimos que los estimadores de momentos ponderados por probabilidad, $\hat{\theta}$, son asintóticamente normales con media $f(\beta) = \theta$ y matriz de covarianza $n^{-1}JVJ^T$, donde $J = (J_{ij})$ y $J_{ij} = \partial f_i / \partial b_j$. Finalmente, usando el método delta y la inversa de la DGVE, obtenemos estimadores asintóticamente normales para los niveles de retorno. Es posible obtener intervalos de confianza aproximada para dichos estimadores de una manera análoga al cálculo descrito en el método de máxima verosimilitud.

2.2.2. No Factibilidad de los Estimadores Puntuales

La DGVE con $\xi \neq 0$ tiene soporte acotado, de manera específica,

$$x \geq \mu - \frac{\xi}{\sigma}, \quad \text{cuando } \xi > 0,$$

ó

$$x \leq \mu - \frac{\xi}{\sigma}, \quad \text{cuando } \xi < 0.$$

Los estimadores obtenidos con el método de momentos ponderados por probabilidad tienen el problema de no garantizar que las estimaciones serán consistentes con los datos observados. Es decir, dada una muestra aleatoria x_1, x_2, \dots, x_n de la DGVE, y usando los estimadores (2.8) y (2.9), encontramos que el estimador de la cota es $\hat{\mu} - \hat{\xi}/\hat{\sigma}$, sin embargo, no está garantizado que

$$x_i \geq \hat{\mu} - \frac{\hat{\xi}}{\hat{\sigma}} \quad \forall x_i, \quad i = 1, \dots, n, \quad \text{cuando } \hat{\xi} > 0,$$

ó

$$x_i \leq \hat{\mu} - \frac{\hat{\xi}}{\hat{\sigma}} \quad \forall x_i, \quad i = 1, \dots, n, \quad \text{cuando } \hat{\xi} < 0.$$

Dupuis [4] usó un experimento basado en simulaciones para aproximar la probabilidad de obtener estimadores de momentos ponderados por probabilidad no factibles cuando se estiman parámetros de la DGVE. Simulando 10,000 muestras aleatorias para diferentes combinaciones de tamaños de muestra y valores para el parámetro de forma, fue registrado el

n	ξ									
	-0.500	-0.400	-0.300	-0.200	-0.100	0.100	0.200	0.300	0.400	0.500
10	5.67	4.06	2.10	1.09	.67	.29	.17	.29	.27	.28
50	13.19	8.86	4.79	2.16	.50	.02	.01	.11	.41	.53
100	15.82	10.36	5.30	1.82	.33	.00	.01	.09	.32	.51
200	17.66	10.86	4.90	1.42	.25	.00	.00	.02	.16	.53
500	19.91	11.45	4.31	.90	.06	.00	.00	.00	.03	.29
1000	20.37	11.07	3.87	.56	.00	.00	.00	.00	.04	.24
5000	19.92	10.33	2.53	.11	.00	.00	.00	.00	.02	.07
10,000	20.10	9.55	2.14	.05	.00	.00	.00	.00	.01	.05

Cuadro 2.2: Porcentaje observado de la no factibilidad de los estimadores de momentos ponderados por probabilidad de los parámetros de la Distribución Generalizada de Valores Extremos. n denota tamaño de muestra y ξ el parámetro de forma. Para cada escenario fueron simuladas 10,000 muestras. Datos tomados de Dupuis [4], Tabla I.

porcentaje de muestras que originaron estimadores de momentos ponderados por probabilidad no factibles. Un extracto de los resultados de dicho estudio aparece en el Cuadro(2.2) (ver Dupuis [4], Tabla I). El porcentaje de no factibilidad presenta valores pequeños cuando $\xi > -0.2$ (es menor a 1% cuando $\xi > 0$), sin embargo, para el resto de los casos dichos porcentajes presentan valores que pueden alcanzar el 20%.

Capítulo 3

Estudio Basado en Simulaciones: Descripción y Resultados

En el capítulo anterior explicamos tres tipos de enfoques para obtener estimaciones de intervalo: el método de máxima verosimilitud, intervalos de verosimilitud confianza basados en la verosimilitud perfil y el método de momentos ponderados por probabilidad. Para comparar esos enfoques en el contexto del análisis de máximos por bloques realizamos un estudio basado en simulaciones. Las características de dicho estudio las explicamos en la Sección 3.1, mientras que los resultados obtenidos aparecen en la Sección 3.2.

3.1. Características del Estudio

Para evaluar el desempeño de los diferentes enfoques, compararemos la precisión de las estimaciones que se refieren a cuantiles altos pues, como mencionan Coles y Dixon, esas cantidades son, en muchas ocasiones, el principal componente en un análisis de valores extremos (Coles y Dixon [3], 7). Existen trabajos previos que abordan este problema, en particular, sobresale el realizado por Hosking et al. [9] en el cual analizaron el comportamiento de estimaciones *puntuales* para los parámetros y cuantiles de la distribución generalizada de valores extremos (DGVE). Los métodos que consideraron fueron el de máxima verosimilitud y el de momentos ponderados por probabilidad, respectivamente. Primero analizaron las distribuciones asintóticas de dichos estimadores y calcularon el sesgo, varianza y eficiencia asintótica, ésta última calculada como el cociente entre los determinantes de las matrices de covarianzas asintóticas. (ver Figuras 1-4 en Hosking et al. [9]). Después, para estudiar el

desempeño de los estimadores con muestras “pequeñas”, elaboraron un experimento basado en simulaciones con las siguientes características. Las simulaciones fueron realizadas para muestras de tamaño $n = 15, 25, 50, 100$ considerando parámetros de forma ξ con valores $-0.4, -0.2, 0, 0.2, 0.4$. Además, dado que los métodos son invariantes ante transformaciones lineales de los datos, entonces, sin pérdida de generalidad, usaron $\mu = 0$ y $\sigma = 1$. Para cada combinación de los valores n y ξ , 10,000 simulaciones aleatorias fueron generadas de la DGVE, y para cada una se obtuvieron los estimadores puntuales de máxima verosimilitud y de momentos ponderados por probabilidad, respectivamente. Finalmente, para evaluar el desempeño de dichos estimadores, calcularon su sesgo y la desviación estándar. Con los resultados obtenidos, Hosking et al. [9] concluyen que los estimadores de máxima verosimilitud y de momentos ponderados por probabilidad tienen un desempeño similar cuando se consideran tamaños de muestra “moderada” ($n = 50, 100$), pero la desviación estándar de los estimadores de momentos ponderados por probabilidad son substancialmente menores que las obtenidas para los estimadores de máxima verosimilitud cuando se consideran muestras “pequeñas” ($n = 15, 25$). Además, aunque los estimadores de momentos ponderados por probabilidad son asintóticamente ineficientes comparados con los estimadores de máxima verosimilitud, esa ineficiencia no se detectó en muestras de tamaño 100 o menor.

El mejor desempeño de los estimadores puntuales de momentos ponderados por probabilidad con relación a los estimadores de máxima verosimilitud cuando se usan muestras pequeñas, es estudiado por Coles y Dixon [3], los cuales proponen que la superioridad en muestras pequeñas de los estimadores de momentos ponderados por probabilidad se debe a la suposición de un espacio paramétrico restringido que corresponde a momentos finitos en la población. Para incorporar información similar en un análisis basado en verosimilitud, Coles y Dixon proponen “un estimador de máxima verosimilitud penalizado, el cual retiene la flexibilidad en modelación y la optimidad en muestras de tamaño grande de los estimadores de máxima verosimilitud, pero mejora en propiedades con muestras pequeñas.” (Coles y Dixon [3], pag. 5) Para verificar sus afirmaciones, Coles y Dixon consideraron un estudio basado en simulaciones. Al igual que en Hosking et al. [9], Coles y Dixon analizan estimadores *puntuales*, sin embargo, en su trabajo también consideran estimadores de *intervalo*. Dichos intervalos fueron calculados, por un lado, usando la distribución asintótica para los estimadores de máxima verosimilitud, y por el otro, usando un análisis Bayesiano bajo el cual el estimador de máxima verosimilitud penalizado que proponen coincide con la moda de la distribución posterior. Los resultados que reporta Coles corresponden sólo a cuantiles del 99%, para muestras de tamaño 25 y con valores del parámetro de forma $\xi = -0.4, -0.2, 0, 0.2, 0.4$ (ver Tabla

2 en Coles y Dixon [3], pag 17).

Al igual que en los trabajos de Hosking et al. [9] y Coles y Dixon, [3], nosotros elaboramos un estudio basado en simulaciones con el cual deseamos evaluar el desempeño de los estimadores obtenidos con los métodos descritos en el capítulo anterior cuando son aplicados con muestras pequeñas. Dado que las propiedades de los estimadores puntuales obtenidos con el método de máxima verosimilitud y de momentos ponderados por probabilidad, respectivamente, fueron analizadas con detalle en el trabajo de Hosking et al., en este trabajo nos concentramos en el estudio de estimadores de intervalo. Ese tipo de estimadores fue considerado por Coles y Dixon [3], sin embargo, en este trabajo consideramos otros estimadores, específicamente, los intervalos de confianza aproximada obtenidos para los estimadores de momentos ponderados por probabilidad, así como intervalos de verosimilitud confianza obtenidos usando la verosimilitud perfil. Además, como explicamos a continuación, en el presente trabajo consideramos un rango más amplio de posibles escenarios que los considerados por Coles y Dixon.

Los detalles específicos de nuestro estudio basado en simulaciones son los siguientes. Consideramos simulaciones de muestras de tamaño $n = 25, 50, 100$ con parámetro de forma

$$\begin{aligned} \xi = & -0.4, -0.35, -0.3, -0.25, -0.2, -0.175, -0.15, -0.125, -0.1, -0.09, -0.08, -0.07, \\ & -0.06, -0.05, -0.04, -0.03, -0.02, -0.01, 0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, \\ & 0.07, 0.08, 0.09, 0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.3, 0.35, 0.4, \end{aligned}$$

y parámetros de localización y escala iguales a $\mu = 0$ y $\sigma = 1$. Para cada combinación de tamaño de muestra y parámetro de forma, generamos 10,000 simulaciones a partir de la DGVE, para cada una de las cuales obtuvimos los estimadores de intervalo para cuantiles del 95% y 99%, respectivamente, siguiendo los tres métodos considerados en este trabajo. En todos los casos calculamos intervalos con un nivel de confianza del 95%.

3.2. Análisis de Resultados

En este trabajo usamos la cobertura como principal indicador del desempeño de los estimadores de intervalo. Sin embargo, comenzamos analizando el porcentaje de casos en que no fue posible obtener alguno(s) de los estimadores de intervalo, situación que ocurrió con mayor frecuencia para el método de máxima verosimilitud y de momentos ponderados por probabilidad, respectivamente. Finalmente, también complementamos la información

proporcionada por la cobertura con un análisis de la longitud de los intervalos obtenidos. Dicho análisis muestra la magnitud en que se incrementa la longitud de los estimadores cuando tienen una mayor cobertura.

3.2.1. Fallas en el Cálculo de Estimadores de Intervalo

Dada una muestra de la DGVE, es posible que no se puedan calcular alguno(s) de los estimadores de intervalo que son de interés en este estudio basado en simulaciones. Lo anterior puede ocurrir por que el cálculo de dichos intervalos requiere el uso de métodos numéricos, o bien, debido a restricciones de tipo teórico. Comentarios específicos para cada método aparecen a continuación.

Al usar el método de máxima verosimilitud, el primer paso para obtener estimadores de intervalo para cuantiles consiste en calcular los estimadores de máxima verosimilitud para los parámetros de la DGVE. Dado que no existen expresiones analíticas en este caso para dichos estimadores, es necesario recurrir a métodos numéricos. Como menciona Hosking [9], este procedimiento no siempre es satisfactorio debido a que la función de log verosimilitud para una muestra $\{x_1, x_2, \dots, x_n\}$ es

$$\log L = -n \log \sigma - (1 + \xi) \sum y_i - \sum \exp -y_i,$$

$$y_i = (1/\xi) \log \{1 + \xi(x_i - \mu)/\sigma\},$$

y $\log L$ puede hacerse arbitrariamente grande al considerar $\xi < 0$ y escogiendo μ y σ de modo que la cota superior $\mu + \sigma/\xi$ de la distribución esté suficientemente cerca al máximo de la muestra. “En la práctica, los estimadores de máxima verosimilitud de los parámetros son obtenidos encontrando un máximo local de $\log L$. Para algunos ejemplos, sin embargo, parece que $\log L$ no tiene un máximo local” (Hosking [9], 255). Para nuestro estudio basado en simulaciones, el porcentaje de muestras para las cuales no fue posible calcular los estimadores de máxima verosimilitud fue, en general, muy pequeño. Sólo aparecieron cuando consideramos tamaños de muestra igual a 25, y en ese caso, con $\xi \geq -0.3$ no rebasaron el 1%, mientras que para $\xi = -0.35$ y $\xi = -0.4$ el porcentaje estuvo entre 1% y 2%. Por otro lado, una vez que calculamos los estimadores de máxima verosimilitud para los parámetros podemos obtener los estimadores de máxima verosimilitud para los cuantiles y usar la normalidad asintótica de dichos estimadores para obtener intervalos de confianza aproximada. Como mencionamos en el capítulo anterior, podemos garantizar que esa distribución asintótica es válida cuando $\xi > -0.5$. Aunque en este trabajo sólo estamos considerando muestras simuladas de la

DGVE con $\xi > -0.4$, es posible que obtengamos estimadores de máxima verosimilitud para el parámetro de forma, $\widehat{\xi}$, menores que -0.5 , en particular al usar muestras pequeñas. En esos casos no calculamos los intervalos de confianza aproximada porque los estimadores puntuales de máxima verosimilitud son usados como valores aproximados de los parámetros reales de la distribución para obtener los parámetros de la distribución asintótica, y por lo tanto, el resultado teórico no nos permite tomar valores para el parámetro de forma menores a -0.5 . Por lo tanto, lo anterior es una restricción en el cálculo de estimadores de intervalo para cuantiles usando el método de máxima verosimilitud. Combinando estos casos con aquellos para los cuales no pudimos encontrar los estimadores puntuales de máxima verosimilitud, obtenemos los porcentajes que aparecen en la Figura 3.1.

Ahora consideramos los errores que pueden aparecer en el cálculo de intervalos de confianza aproximada al usar el método de momentos ponderados por probabilidad. Primero, es posible que los estimadores puntuales de los parámetros de la distribución no sean congruentes con los datos, o como denominamos a esa situación en el capítulo anterior, dichos estimadores podrían ser no factibles, lo cual es más probable que suceda cuando el parámetro de forma es negativo y el tamaño de muestra aumenta. Por otro lado, la distribución asintótica de los estimadores puntuales para los cuantiles es válida sólo cuando $\xi < 0.5$. Por lo tanto, y de modo análogo a lo que ocurre en el método de máxima verosimilitud, no calculamos los estimadores de intervalo para las muestras cuyo estimador de momentos ponderados por probabilidad del parámetro de forma es mayor que 0.5 . Los porcentajes de muestras para las cuales no fue posible calcular los estimadores de intervalo, debido a las dos situaciones descritas, aparecen en la Figura 3.1.

Por último, en el cálculo de los intervalos de verosimilitud-confianza para los cuantiles es necesario recurrir a métodos numéricos de maximización para evaluar la verosimilitud perfil. Sólo para un pequeño porcentaje de los casos considerados en nuestro estudio se obtuvo un error al aplicar dichas rutinas. Para el cuantil del 95 % dichos porcentajes fueron menores al 0.2 % para todas las combinaciones de tamaño de muestra y valor del parámetro de forma, mientras que para los estimadores del cuantil del 99 % los porcentajes fueron menores a 1.0 %, excepto para el caso en que consideramos un tamaño de muestra igual a 25 y $\xi = -0.4$, en el cual dicho porcentaje fue 2.18 %.

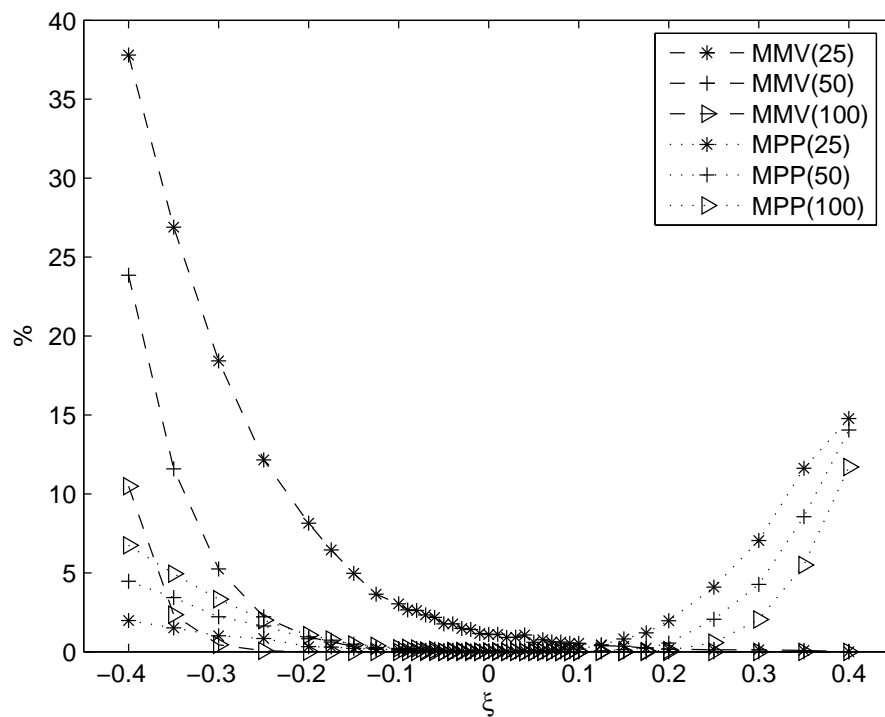


Figura 3.1: Porcentaje de muestras para las cuales no fue posible obtener los estimadores de intervalo para cuantiles usando el método correspondiente. Los porcentajes anteriores son válidos sin importar el nivel del cuantil que estemos estimando. El número entre paréntesis denota tamaño de la muestra. MMV - Método de Máxima Verosimilitud; MPP - Momentos Ponderados por Probabilidad.

3.2.2. Cobertura

Para evaluar el desempeño de los estimadores de intervalo obtenidos bajo cada enfoque calculamos su cobertura, es decir, el porcentaje de veces en que los intervalos simulados contienen el verdadero valor del cuantil. Una aclaración importante es que dicho porcentaje fue calculado con respecto al número total de simulaciones, por lo tanto, en cada escenario los estimadores tienen al menos una reducción de la cobertura igual al porcentaje de casos para los cuales no fue posible calcular las estimaciones. Las coberturas encontradas en nuestro estudio para las estimaciones relativas al cuantil del 95 % y 99 % aparecen en las Figuras 3.2 y 3.3, respectivamente. Es interesante notar que los intervalos de verosimilitud-confianza obtenidos usando la verosimilitud perfil presentan coberturas cercanas al 95 % en casi todos los casos. Por otro lado, los intervalos obtenidos usando el método de máxima verosimilitud presentaron, en general, el peor rendimiento pues las caídas de las coberturas cuando $\xi < 0$ fueron las más pronunciadas, y además, si bien existe un incremento de las coberturas conforme ξ aumenta, siempre presentan valores menores a los que se obtienen usando los intervalos basados en la verosimilitud perfil. Finalmente, las coberturas para los intervalos obtenidos con el método de momentos ponderados por probabilidad son bajas con $\xi > 0$ pero se incrementan hasta alcanzar valores mayores al 90 % cuando ξ es negativo.

3.2.3. Longitud de los Estimadores de Intervalo

Las coberturas encontradas muestran que las estimaciones de intervalo obtenidas usando la verosimilitud perfil tienen el mejor desempeño de entre los métodos considerados en este estudio. En esta sección comparamos la longitud de los intervalos obtenidos con cada método, para lo cual, consideramos los siguientes indicadores.

1. El cociente entre la longitud de los intervalos obtenidos usando la verosimilitud perfil y la longitud de los obtenidos usando el método de máxima verosimilitud.
2. El cociente entre la longitud de los intervalos obtenidos usando el método de momentos ponderados por probabilidad y la longitud de los obtenidos usando el método de máxima verosimilitud.

Resumimos los resultados obtenidos usando los cuartiles de cada serie de cocientes, dichas estadísticas aparecen en las Figuras 3.4 y 3.5. Con el uso de los cuartiles obtenemos información sobre el comportamiento medio de los cocientes, así como de su dispersión. Los cocientes

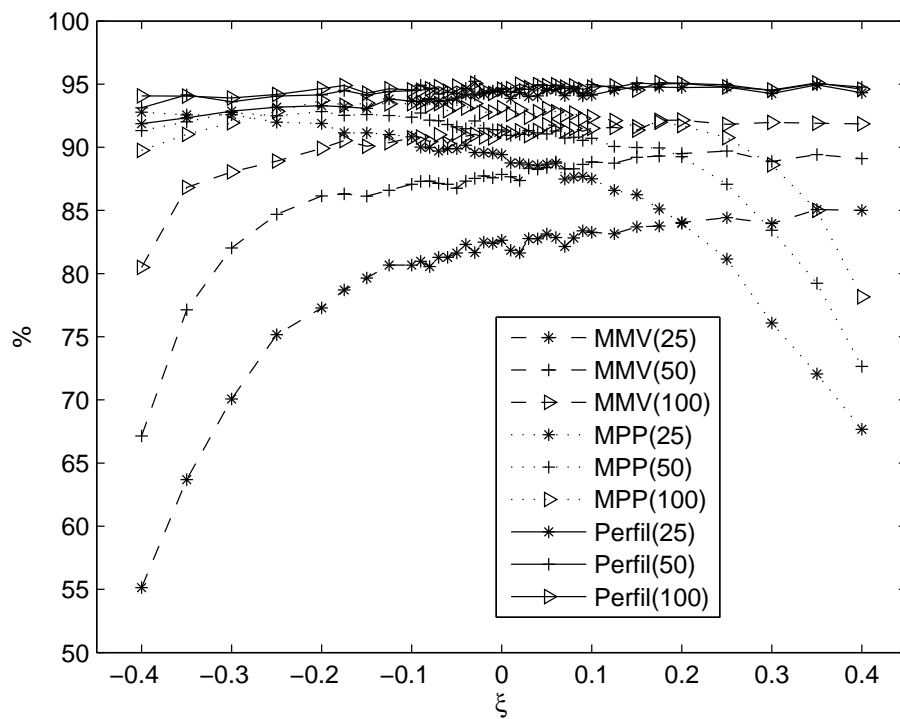


Figura 3.2: Cobertura de los estimadores de intervalo para el cuantil del 95 % con un nivel de confianza del 95 %. Los métodos considerados son: MMV - Método de Máxima Verosimilitud; MPP - Momentos Ponderados por Probabilidad; Perfil - Intervalos de verosimilitud confianza basados en la verosimilitud perfil. El número entre paréntesis denota tamaño de la muestra.

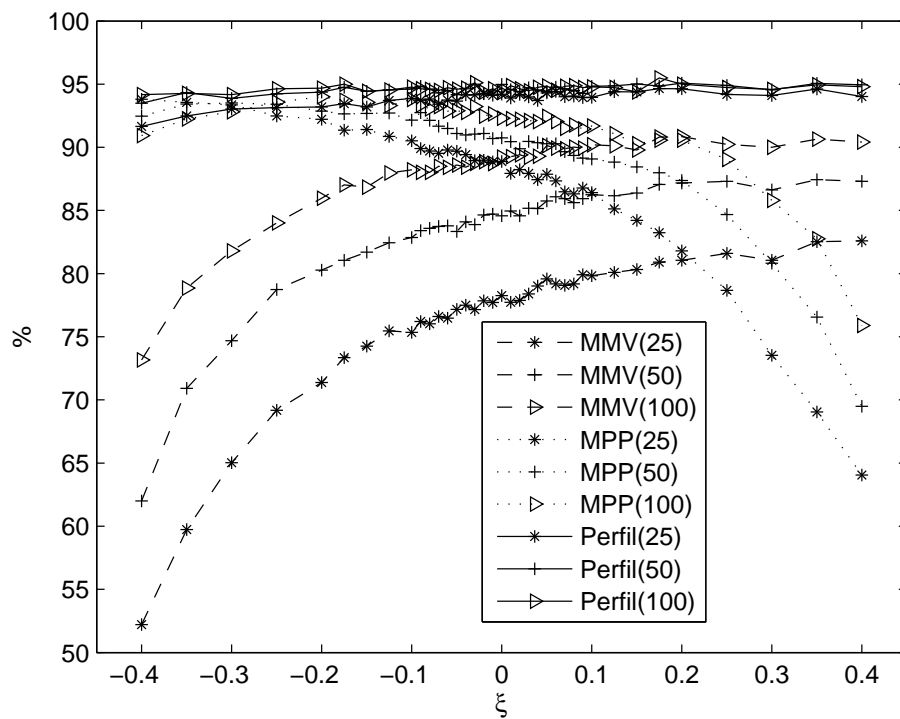


Figura 3.3: Cobertura de los estimadores de intervalo para el cuantil del 99 % con un nivel de confianza del 95 %. Los métodos considerados son: MMV - Método de Máxima Verosimilitud; MPP - Momentos Ponderados por Probabilidad; Perfil - Intervalos de verosimilitud confianza basados en la verosimilitud perfil. El número entre paréntesis denota tamaño de la muestra.

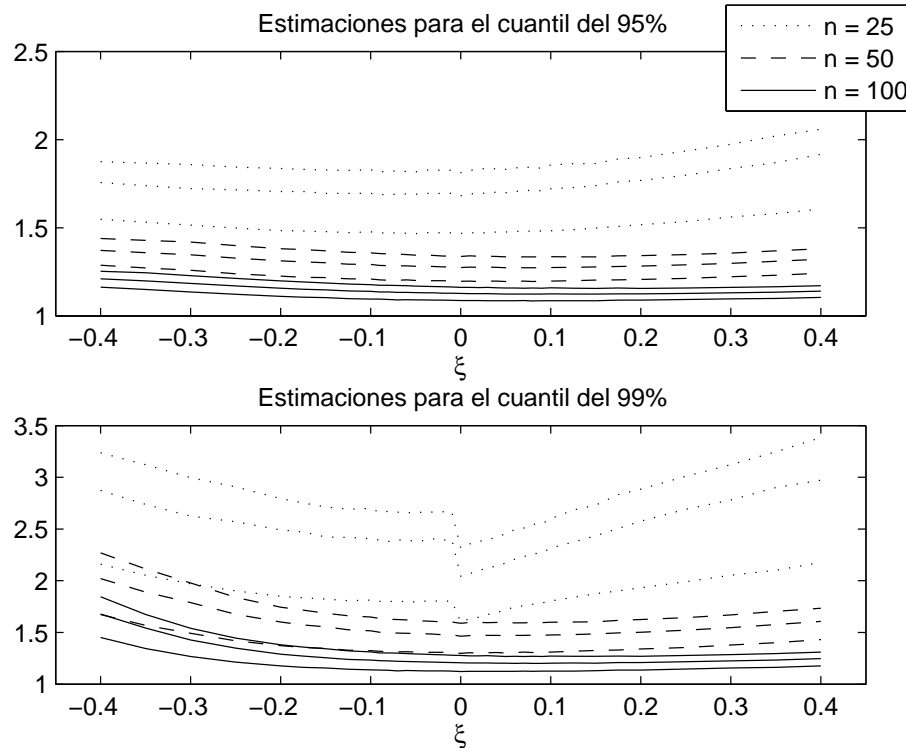


Figura 3.4: Cuartiles del cociente entre la longitud de los estimadores de intervalo obtenidos usando la verosimilitud perfil y la longitud de los obtenidos usando el método de máxima verosimilitud. Para cada valor de n , las tres curvas asociadas se refieren a los cuartiles Q_1 , Q_2 y Q_3 , respectivamente. No tienen etiquetas pues $Q_1 \leq Q_2 \leq Q_3$.

anteriores no están calculados para cada una de las 10,000 muestras simuladas pues, como ya mencionamos, para cada método existe un porcentaje de casos en que no es posible obtener las estimaciones correspondientes. Esta situación debe ser tomada en cuenta al considerar los resultados pues no se refleja de manera explícita.

Los resultados que aparecen en la Figura 3.4 muestran que la longitud de los intervalos de verosimilitud confianza obtenidos con la verosimilitud perfil tienden a ser más grandes que aquellos obtenidos usando el método de máxima verosimilitud. El comportamiento de las curvas sugiere que dicha diferencia aumenta cuando incrementamos el nivel de los cuantiles y/o disminuimos el tamaño de muestra, además, también encontramos que la variabilidad en la diferencia entre las longitudes decrece cuando aumentamos el tamaño de la muestra.

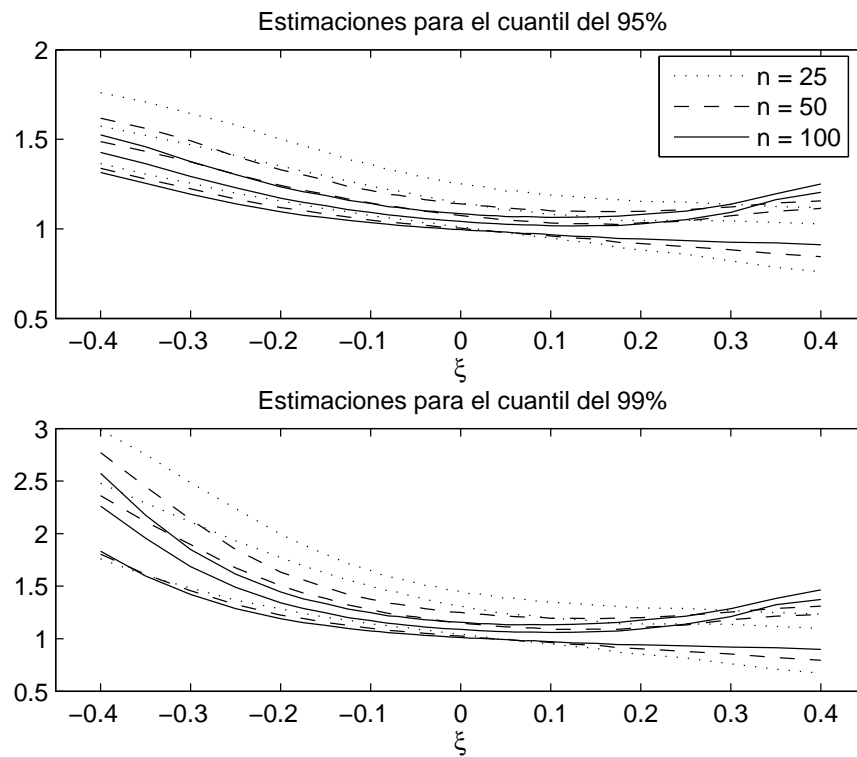


Figura 3.5: Cuartiles del cociente entre la longitud de los estimadores de intervalo obtenidos usando el método de momentos ponderados por probabilidad y la longitud de los obtenidos usando el método de máxima verosimilitud. Para cada valor de n , las tres curvas asociadas se refieren a los cuartiles Q_1 , Q_2 y Q_3 , respectivamente. No tienen etiquetas pues $Q_1 \leq Q_2 \leq Q_3$.

Por otro lado, la Figura 3.5 sugiere que para $\xi < 0$, la longitud de los intervalos obtenidos usando el método de momentos ponderados por probabilidad tienden a ser más grandes que aquellos obtenidos usando el método de máxima verosimilitud, además de que dicha diferencia aumenta cuando consideramos cuantiles altos y/o el valor de ξ decrece. Un aspecto interesante es que incrementos del tamaño de la muestra no tienen un efecto tan evidente como sí sucedió con el otro cociente.

Conclusiones

La distribución de máximos por bloques puede ser aproximada por la distribución generalizada de valores extremos (DGVE) que consta de un parámetro de localización, otro de escala, y uno más de forma, los cuales denotaremos μ , σ y ξ , respectivamente. Usando la DGVE, elaboramos un estudio basado en simulaciones para evaluar el desempeño, en muestras pequeñas, de tres técnicas estadísticas para calcular estimadores de intervalo para cuantiles. Dichas técnicas son: el método de máxima verosimilitud, el método de momentos ponderados por probabilidad y los intervalos de verosimilitud confianza que se obtienen a partir de la verosimilitud perfil. En el estudio consideramos 37 valores diferentes en el intervalo $(-0.4, 0.4)$ para el parámetro de forma ξ , cuantiles del 95 % y 99 %, muestras de tamaño $n = 25, 50, 100$, y un nivel de confianza para los intervalos del 95 %. Para cada escenario simulamos 10,000 muestras de la DGVE.

Los argumentos asintóticos utilizados en el método de máxima verosimilitud y el método de momentos ponderados por probabilidad imponen restricciones sobre el rango de valores posibles que puede tomar el parámetro de forma. En el primer caso debe cumplirse que $\xi > -0.5$, mientras que en el segundo $\xi < 0.5$. Las restricciones anteriores implican que no siempre es posible calcular intervalos de confianza aproximada cuando se usan dichos métodos, lo anterior ocurre porque los estimadores puntuales obtenidos en cada método, son usados para aproximar los parámetros reales, y en consecuencia, en algunos casos no cumplieron las restricciones correspondientes, lo anterior ocurre a pesar de que el parámetro real que fue usado para obtener las simulaciones sí lo hace. Los resultados de nuestro estudio muestran que el porcentaje de casos para los cuales no es posible calcular los intervalos debido a la situación recién descrita pueden ser muy altos, en particular, cuando se usa el método de máxima verosimilitud con $\xi = -0.4$ y $n = 25$, dicho porcentaje puede llegar a valores cercanos al 40 %. Por otro lado, para los estimadores de momentos ponderados por probabilidad, dicho porcentaje crece conforme aumenta ξ y/o disminuye el tamaño de muestra, llegando a ser igual a 15 %.

La cobertura fue usada como principal indicador del desempeño de los estimadores de intervalo incluidos en este estudio. Un aspecto sobresaliente es que en todos los escenarios los intervalos de verosimilitud confianza obtenidos con la verosimilitud perfil tuvieron coberturas mayores al 90 %, y en la mayoría de los casos, dicho valor fue muy cercano al 95 %. Estos resultados superan los obtenidos por los otros dos métodos en los cuales el tamaño de la muestra, el nivel del cuantil considerado, así como el valor de ξ , influyeron en el valor de las coberturas. Específicamente, para los estimadores obtenidos con el método de máxima verosimilitud encontramos una caída en la cobertura cuando ξ es negativo, por otro lado, para los estimadores de momentos ponderados por probabilidad, también hay una caída de la cobertura pero esta ocurre con valores positivos de ξ .

El buen desempeño de las coberturas de los intervalos de verosimilitud-confianza está acompañado por una mayor longitud cuando los comparamos con los intervalos obtenidos usando los otros dos métodos. En particular, con relación a los intervalos obtenidos usando el método de máxima verosimilitud, los intervalos de verosimilitud-confianza fueron en la mayoría de los casos más grandes y dicha diferencia es mayor cuando los calculos se refieren a cuantiles más altos y/o son consideradas muestras pequeñas.

Con los resultados encontrados en este trabajo no pretendemos determinar cuál es el “mejor” estimador de intervalo para cuantiles en el contexto de análisis de máximos por bloques. En cambio, los resultados deben ser vistos como información que puede usarse en la práctica para seleccionar alguno de los estimadores, donde dicha selección dependerá del contexto específico del problema en cuestión así como de los intereses del cliente.

La contribución de este trabajo a la literatura existente que se refiere el análisis del desempeño de estimadores, consiste en haber estudiado estimadores de intervalo considerando un número amplio de escenarios. Aunque las herramientas teóricas fueron introducidas por Hosking et al. [9], en dicho trabajo sólo fue evaluado el desempeño con muestras pequeñas de los estimadores puntuales de máxima verosimilitud y de momentos ponderados por probabilidad, respectivamente. Por otro lado, el desempeño de estimadores de intervalo sí es considerado por Coles y Dixon [3], aunque los resultados reportados son para un rango reducido de casos y sólo incluyen la cobertura de las estimaciones.

Finalmente, algunos temas de interes para futuras investigaciones son los siguientes. Podemos evitar usar simulaciones de la DGVE, y en cambio, considerar simulaciones de distribuciones específicas para las cuales el dominio de atracción es conocido, de esa forma, los resultados serían más cercanos a lo que ocurre en la realidad. Los resultados de un trabajo de ese tipo darían información sobre la velocidad de convergencia para distribuciones que

pertenecen a diferentes dominios de atracción. Otro tema de interés consiste en comparar estimadores de intervalo para la distribución generalizada de Pareto. Dicha distribución es usada en el análisis de eventos extremos, aunque en ese caso no es considerado análisis de máximos por bloques, sino uno de picos sobre excesos (peaks over thresholds). Singh y Maqsood [17] y Moharram et al. [13] evalúan el desempeño de estimadores para la distribución generalizada de Pareto, pero en ambos casos sólo estimadores puntuales son considerados. Otro posible tema de investigación se refiere a considerar las propiedades de los estimadores para muestras pequeñas desde una perspectiva teórica. Un trabajo muy reciente que estudia los estimadores de momentos ponderados por probabilidad desde ese enfoque fue presentado por Furrer y Naveau [6].

Bibliografía

- [1] Castillo, E., Hadi, A.S., Balakrishnan, N. y Sarabia, J.M.(2005). *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley, New York.
- [2] Coles, S. (2001). *An Introduction to Statistical Modelling of Extreme Values*. Springer, New York.
- [3] Coles, S. y Dixon, M. (1999). Likelihood-Based Inference for Extreme Values Models *Extremes*, **2** : **1**, pp.5-23.
- [4] Dupuis, D. J. (1996). Estimating the Probability of Obtaining Nonfeasible Parameter Estimates of the Generalized Extreme-Value Distribution. *J. Statist. Comput. Sim.*, **56**, pp.23-38.
- [5] Embrechts, P., Klüpelberg, C. y Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, New York.
- [6] Furrer, R., Naveau, P. (2007). Probability Weighted Moments Properties for Small Samples. *Statistics Probability Letters*, **77**, pp.190-195.
- [7] Greenwood J.A., Landwehr, J. M., Matalas, N. C., y Wallis, J. R. (1979). Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form. *Water Resources Research*, **15**, pp.1049-1054.
- [8] Hogg, R.V. y Klugman, S. A. (1984). *Loss Distributions*. Wiley, New York.
- [9] Hosking, J. R. M., Wallis, J.R., y Wood, E. F. (1985). Estimation of the Generalized Extreme Value Distribution by the Method of Probability Weighted Moments. *Technometrics*, **27**, pp.251-261.

- [10] Johnson, N.L. y Kotz, S. (1970). *Distributions in Statistics. Continuous Univariate Distributions I*. Wiley, New York.
- [11] Johnson, N.L. y Kotz, S. (1970). *Distributions in Statistics. Continuous Univariate Distributions II*. Wiley, New York.
- [12] Landwehr, J. M., Matalas, N. C., y Wallis, J. R. (1979). Probability Weighted Moments Compared with Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles. *Water Resources Research*, **15**, pp.1055-1064.
- [13] Moharram, S.H., Gosain, A. K., y Kapoor, P. N. (1993). A Comparative Study for the Estimators of the Generalized Pareto Distribution. *Journal of Hidrology*, **150**, pp.169-185.
- [14] Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.
- [15] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- [16] Shao, J. (1999). *Mathematical Statistics*. Springer, New York.
- [17] Singh, V. P. y Maqsood Ahmad. (2003). A Comparative Evaluation of the Three-Parameter Generalized Pareto Distribution. *J. Statist. Comput. Sim.*, **74 : 2**, pp. 91-106.
- [18] Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, **42**, pp. 67-90.