Centro de Investigación en Matemáticas

Modelación Estadística de Máximos por Bloques.

Autor:
José Eliud Vilchis Carrera
Directores de tesis:
Eloísa Díaz-Francés Murguía
Joaquín Ortega Sánchez

Agosto 2006

Agradecimientos

Quiero agradecer al Centro de Investigación en Matemáticas, A.C. por haberme brindado la oportunidad de formar parte de él y por el apoyo económico que recibí en parte de mis estudios. También, agradezco a CONACyT y CONCyTEG por la beca que me brindaron para poder llevar a cabo mis estudios.

Agradezco al Dr. Miguel Nakamura Savoy y al Dr. Enrique Raul Villa Diharce por haberme apoyado en la culminación de este trabajo.

Agradezco al Dr. David Sprott y al Dr. Luis Escobar por sus valiosos comentarios referentes a esta tesis.

Muy en especial, quiero agradecer a la Dra. Eloísa Díaz-Francés Murguía y al Dr. Joaquín Ortega Sánchez, por su tiempo, paciencia, dedicación y enseñanzas hacia conmigo.

También muy en especial le agradezco a la Dra. Graciela González Farías, por ser como es.

Finalmente, desde el fondo de mi corazón quiero agradecer a mi familia, amigos y a Ana Laura, por ser el condimento de mi vida.

Índice general

Agradecimientos I

Introducción V

1. Preliminares 1
1.1. Teoría clásica de Valores Extremos . . . 1
1.2. Cercanía entre distribuciones . . . 5
1.3. Herramienta estadística . . . 7
1.3.1. Funciones de logverosimilitud para la DVEG y las DVE . . . 7
1.3.2. Función de verosimilitud perfil . . . 8
1.3.3. Regiones de verosimilitud-confianza . . . 9
1.3.4. Métodos gráficos de diagnóstico . . . 11
1.4. Modelación estadística . . . 12
1.4.1. Proceso de modelación . . . 12
1.4.2. Parámetros umbral . . . 13
1.4.3. Propuesta de modelación del máximo por bloques . . . 14

2. Ejemplos de modelación 17
2.1. Niveles de lluvia en El Temazcal Charo, Michoacán. . . . 17
2.2. Niveles de lluvia en la estación de Bombeo, Michoacán. . . . 22
2.3. Máximos anuales de niveles de mar en Port Pirie, Australia . . . 26
2.4. Niveles de lluvia en Morelia, Michoacán. . . . 30
2.5. Niveles de lluvia en Maiquetía, Venezuela . . . 34
2.6. Niveles de lluvia en Presa Cointzio, Michoacán. . . . 39
2.7. Niveles de acumulación de nieve en Carolina del Norte, E.U.A. . . . 43
2.8. Comentarios finales . . . 49

3. Simulaciones 51
3.1. Introducción . . . 51
3.2. Planteamiento de los escenarios de simulación . . . 51
3.3. Resultados de coberturas . . . 52
3.4. Cociente de intervalos . . . 55
3.5. Comentarios finales . . . 57

4. Conclusiones 59

Introducción

Dentro de la Teoría clásica de Valores Extremos, el teorema de Fisher-Tippet (1928) nos dice que la distribución del máximo reescalado de una sucesión de variables aleatorias independientes e idénticamente distribuidas, bajo ciertos supuestos, en caso de converger, lo hace a una y sólo una de las siguientes familias: Weibull, Fréchet o Gumbel. La Distribución de Valores Extremos Generalizada (1954) engloba a las tres familias anteriores, correspondiendo a cada una de ellas según el valor del parámetro de forma, llamémosle c, de la misma. Posterior a esta reformulación, en el proceso de modelación del máximo de alguna variable de interés se consideraba alguno o los tres submodelos mencionados anteriormente para describir la distribución del máximo, llevando a cabo la selección de un modelo de manera empírica (ver Castillo [4]). En la literatura reciente de Teoría de Valores Extremos (ver Coles [6] y Finkenstädt y Rootz*é*n [9]) se enfatiza modelar el comportamiento del máximo a través de la DVEG basados en que toda la incertidumbre respecto a la distribución límite está contemplada en ella, evitando con esto la selección subjetiva de un modelo.

Por otro lado, la estimación de intervalos de confianza en el ámbito de extremos se realiza comunmente a través de las propiedades asintóticas del estimador de máxima verosimilitud. Es de reciente aplicación y muy poco explorado el uso de la verosimilitud perfil como una alternativa para la inferencia en valores extremos. Esta herramienta resulta sencilla de implementar y útil en la inferencia en presencia de parámetros de estorbo, como resulta ser el caso de la modelación de la distribución del máximo, donde el interés se centra en inferencias de cuantiles grandes.

De lo anterior, la presente tesis tiene por objeto argumentar la importancia de la selección de un submodelo de la DVEG para las inferencias de interés, proponer un proceso de modelación estadística que conduzca a una representación adecuada del fenómeno en cuestión, y comparar el uso de la verosimilitud perfil en la estimación respecto al método de inferencia tradicionalmente usado.

En concreto, en el proceso de modelación se sugieren de inicio uno o dos submodelos a través de la verosimilitud perfil relativa del parámetro de forma de la DVEG. La inferencia basada en verosimilitud se realiza restringiéndose a los submodelos de la DVEG. Finalmente, en el proceso de modelación se consideran modelos que hagan sentido al contexto del problema, y la selección del mejor modelo deberá hacerse al final del proceso de estimación, validación y crítica de sus consecuencias.

Veremos que basados en este proceso de modelación estadística generalmente se logra una buena representación de la distribución del máximo a través de algún submodelo de la DVEG.

Los antecedentes teóricos más relevantes respecto a la modelación del máximo por bloques se exponen en el Capítulo 1. En el mismo capítulo se aborda la cercanía entre las distribuciones Weibull, Fréchet y Gumbel. Así mismo, se presentan las herramientas estadísticas que se utilizarán en el proceso de modelación estadística. Se añaden comentarios existentes en la lite-

ratura respecto a parámetros umbrales, modelación y estimación, así como algunas propuestas relevantes a considerar. Finalmente, se establece la metodología a seguir basada en los objetivos del presente proyecto.

Diversos conjuntos de datos de fuentes variadas referentes a extremos se describen en el Capítulo 2. Para cada conjunto de observaciones se aplica la metodología de modelación estadística del máximo por bloques propuesta en esta tesis. En el Capítulo 3 se presenta un estudio de simulación cuyo propósito es dar sustento a la metodología estadística aquí propuesta. Las conclusiones del trabajo se expresan en el Capítulo 4.

El software utilizado para el desarrollo del presente trabajo se programó en MATLAB 7.0.

Capítulo 1

Preliminares

La presente sección introduce, en primera instancia, la teoría de extremos referente a la modelación de una sucesión de variables aleatorias independientes e idénticamente distribuidas. Luego, se dedica una sección a la cercanía de los modelos Gumbel, Fréchet y Weibull. Después, se describe la teoría estadística de estimación y validación de modelos a utilizar en la presente tesis. Finalmente, se presenta la propuesta de modelación estadística respecto a la teoría de extremos descrita.

1.1. Teoría clásica de Valores Extremos

La teoría clásica de extremos se centra en el estudio de la función de distribución del máximo de una sucesión de variables aleatorias independientes e idénticamente distribuidas (iid). Comenzaremos por considerar una sucesión de variables aleatorias $X_1, X_2, \ldots, X_n$ independientes con función de distribución común $F_X(.)$. De conocer F_X, podríamos calcular la función de distribución exacta de $M_n = max\{X_1, X_2, \ldots, X_n\}$. Esto es

$$F_{M_n}(x) = P(M_n \leq x) = P(X_1 \leq x, \ldots, X_n \leq x) = \prod_{i=1}^{n} P(X_i \leq x) = [F(x)]^n. \tag{1.1}$$

Puesto que desconocemos F, una alternativa es estimar F basado en $X_1, X_2, \ldots, X_n$ y usar la estimación de F para representar a $[F(x)]^n$. Este camino requiere de estimaciones muy precisas de F, en particular de la cola superior, o de lo contrario se pueden obtener discrepancias serias para F_{M_n} debido a la potencia n.

Otra alternativa es buscar una aproximación de F_{M_n} basado en observaciones extremas y aludiendo a la teoría asintótica. Para ello, no basta considerar sólo el comportamiento de F_{M_n}, pues la potencia n hace que la distribución del máximo se degenere. Para ver lo anterior, consideremos $x_1 < x^+$ y $x_2 \geq x^+$, donde $x^+ = sup\{x : F(x) < 1\}$. Entonces se tiene que

$$F_{M_n}(x_1) = P(M_n \leq x_1) = [F(x_1)]^n \longrightarrow 0 \tag{1.2}$$

pues $F(x_1) < 1$, mientras que

$$F_{M_n}(x_2) = P(M_n \leq x_2) = [F(x_2)]^n = 1^n \longrightarrow 1. \tag{1.3}$$

Es decir, la función de distribución del máximo se degenera en x^+ a medida que n crece. Puesto que M_n es una sucesión creciente, la convergencia en probabilidad implica convergencia con

probabilidad 1. Lo anterior no hace atractivo el uso de la distribución exacta F_{M_n}, y se busca una aproximación para F_{M_n} a través de una normalización. El siguiente resultado se encuentra en Embrechts, *et al.* [8], Teorema A1.5, p. 554.

Teorema 1.1 (Convergencia a familias) *Sean G(x) y H(x) dos funciones de distribución propias, ninguna de las cuales está concentrada en un punto. Supongamos que para $n \geq 0$, X_n, U y V son variables aleatorias con funciones de distribución F_n, G y H, respectivamente. Sean $a_n > 0, \alpha_n > 0, b_n \in R, \beta_n \in R$.*

a) Si

$$F_n(a_n x + b_n) \longrightarrow G(x), \qquad F_n(\alpha_n x + \beta_n) \longrightarrow H(x), \tag{1.4}$$

entonces existen $A > 0$ y $B \in R$ tales que, cuando $n \longrightarrow \infty$,

$$\frac{\alpha_n}{a_n} \longrightarrow A > 0, \qquad \frac{\beta_n - b_n}{a_n} \longrightarrow B \tag{1.5}$$

y

$$H(x) = G(Ax + B) \tag{1.6}$$

b) Si (1.5) vale, entonces cualquiera de las relaciones en (1.4) implica la otra y (1.6) vale.

El teorema anterior nos muestra que la distribución a la cual converge una sucesión de variables aleatorias reescaladas no cambia al tomar diferentes sucesiones de constantes de normalización, excepto en los parámetros de localización y escala. Este teorema es relevante para la obtención del siguiente resultado, debido originalmente a R. A. Fisher y L. H. C. Tippet, el cual resuelve la anterior cuestión de aproximar la distribución de F_{M_n} a través de una distribución asintótica. Este resultado se encuentra en Embrechts, *et al.* [8], Teorema 3.2.3, p. 121.

Teorema 1.2 (Fisher y Tippet) *Si existen sucesiones de constantes $\{a_n > 0\}$ y $\{b_n\}$ tales que*

$$Pr\{(M_n - b_n)/a_n \leq z\} \longrightarrow G(z) \tag{1.7}$$

cuando $n \longrightarrow \infty$, donde G es una función de distribución no degenerada, entonces G pertenece a una de las siguientes familias:

$$Gumbel : \Delta(z) = \exp\{-\exp[-(\frac{z-\mu}{\sigma})]\}, \qquad -\infty < z < \infty \tag{1.8}$$

$$Fréchet : Phi(z) = \begin{cases} 0, & z \leq \mu \\ \exp\{-(\frac{z-\mu}{\sigma})^{-\beta}\}, & z > \mu; \end{cases} \tag{1.9}$$

$$Weibull : \Psi(z) = \begin{cases} \exp\{-[-(\frac{z-\mu}{\sigma})^{\beta}]\}, & z < \mu \\ 1, & z \geq \mu, \end{cases} \tag{1.10}$$

para parámetros $\sigma > 0, \mu$ y $\beta > 0$.

Cabe resaltar que en el Teorema 1.2 no se toma en cuenta la forma de F, no se asegura la existencia de las constantes de normalización, ni se dice cual de las tres familias es la familia a la cual convergerá la sucesión de máximos reescalados. El Teorema 1.2 identifica como límite de la distribución del máximo reescalado de la sucesión de variables consideradas inicialmente a una de las 3 familias mencionadas. La familia (1.8) es la familia Gumbel, la cual tiene su soporte definido en todo R; la familia (1.9) es la familia Fréchet, la cual tiene restringido su soporte a ser

no menor al valor del parámetro μ; finalmente la familia (1.10) es la familia Weibull de máximos, cuyo soporte está restringido a ser no mayor a μ. Cada familia, por sus características, representa un comportamiento diferente del máximo: para la distribución Weibull, el valor máximo que puede tomar la variable en cuestión es finito; para las familias Gumbel y Fréchet, la variable en cuestión no tiene cota superior, pero la función de distribución Fréchet tiene cola derecha más pesada en comparación al modelo Gumbel. Las tres familias cuentan con parámetros de localización y escala (μ y σ, respectivamente); las dos últimas cuentan con un parámetro de forma α. Las tres familias, en conjunto, se conocen como las Distribuciones de Valores Extremos (DVE).

Las expresiones de los cuantiles correspondientes a las DVE (1.8), (1.9) y (1.10), están dadas por las siguientes expresiones:

$$Gumbel : Q_p = \mu - \sigma * (\log[-\log(p)]), \tag{1.11}$$

$$Fréchet : Q_p = \mu + \sigma * [(-\log(p))^{\frac{-1}{\beta}}], \tag{1.12}$$

$$Weibull : Q_p = \mu - \sigma * [(-\log(p))^{\frac{1}{\beta}}], \tag{1.13}$$

donde p representa la probabilidad de no exceder Q_p, $p = P[Z \leq Q_p]$.

R. Von Mises en 1954 y A. F. Jenkinson en 1955 propusieron una reparametrización de las familias (1.8), (1.9) y (1.10) en una sola:

$$G(z) = \begin{cases} \exp\{-[1 + c(\frac{z-a}{b})]^{-\frac{1}{c}}\}, & c \neq 0, \quad 1 + c(\frac{z-b}{a}) > 0 \\ \exp\{-\exp[-(\frac{z-a}{b})]\} & c = 0, \quad b > 0, a \in R. \end{cases} \tag{1.14}$$

La función de distribución (1.14) se conoce como la Distribución de Valores Extremos Generalizada (DVEG). La DVEG tiene como parámetros a, b y c, de localización, escala y forma, respectivamente. Los cuantiles para la DVEG están dados por la siguiente expresión:

$$Q_p = \begin{cases} a - \frac{b}{c} * (1 - [(-\log(p))^{-c}]), & c \neq 0, \\ a - b * (\log[-\log(p)]), & c = 0. \end{cases} \tag{1.15}$$

Cuando $c > 0$, se tiene la familia Fréchet; si $c < 0$ se tiene la familia Weibull; finalmente la familia Gumbel representa el caso límite cuando c tiende a cero. La relación paramétrica entre las DVE y la DVEG está dada de la siguiente manera:

$$\begin{array}{llll} c > 0: & \alpha = 1/c & \sigma = b/c & \mu = a - b/c \\ c < 0: & \alpha = -1/c & \sigma = -b/c & \mu = a - b/c \\ c = 0: & & \sigma = b & \mu = a. \end{array} \tag{1.16}$$

Estos resultados permiten aproximar la distribución del máximo de una sucesión de variables aleatorias de la siguiente manera: consideraremos una muestra $x_1, x_2, \ldots, x_w$ la cual dividiremos en n bloques (renglones) de tamaño m

$$\begin{array}{cccc} x_{11} & x_{12} & \ldots & x_{1m} \\ x_{21} & x_{22} & \ldots & x_{2m} \\ & & \vdots & \\ x_{n1} & x_{n2} & \ldots & x_{nm}. \end{array} \tag{1.17}$$

De cada bloque obtenemos

$$x_i^* = max(x_{i1}, x_{i2}, \ldots, x_{im}), \quad i = 1, 2, \ldots, n \tag{1.18}$$

y supondremos que se cumplen las condiciones tales que $F_{x_i^*} \sim$ DVEG. A este proceso se le conoce como *modelación del máximo por bloques.*

La aproximación de la distribución del máximo de una muestra dada a través de la DVEG supone que las variables de las cuales se obtiene el máximo son independientes. Existen resultados en la literatura referentes a la relajación del supuesto de independencia en el Teorema 1.2. Estos resultados establecen que la distribución del máximo reescalado sigue siendo una de las DVE cuando la sucesión de variables aleatorias no son independientes, pero cumplen condiciones tipo mixing. Para mayor detalle de estas condiciones se refiere al lector a Leadbetter, *et al.* [12].

La modelación del máximo de una sucesión de variables aleatorias usando como aproximación la DVEG tiene como ventaja, desde el punto de vista estadístico, que a través de la estimación de c se pueden realizar inferencias acerca de la distribución límite a la cual converge el máximo de la sucesión de variables iid consideradas.

Para los casos en que la variable aleatoria de interés no es el máximo sino el mínimo de una sucesión de variables aleatorias iid. (como por ejemplo, sistemas de falla), existen resultados análogos a los expuestos para el caso de máximos que describen el comportamiento de $m_n = \text{mín}\{X_1, X_2, \ldots, X_n\}$. Estos resultados se basan en la relación existente entre el máximo y el mínimo: si $M_n = \text{máx}\{X_1, X_2, \ldots, X_n\}$ y $m_n = \text{mín}\{Y_1, Y_2, \ldots, Y_n\}$, donde $Y_i = -X_i$ entonces $M_n = -m_n$. Con esto, hay dos formas de proceder para modelar la función de distribución de m_n. La primera, es hacer uso de los resultados asintóticos para aproximar a F_{m_n} como los descritos para F_{M_n}, y hacer el ajuste directamente. La DVEG para mínimos se expresa de la siguiente manera:

$$G(z) = \begin{cases} 1 - \exp\{-[1 - c(\frac{z-\tilde{b}}{a})]^{-\frac{1}{c}}\}, & c \neq 0, \\ 1 - \exp\{-\exp[\frac{z-a}{\tilde{b}}]\} & c = 0, \end{cases} \tag{1.19}$$

(ver Coles [6] p. 53) donde $\tilde{b} = -b$ en la DVEG. Al igual que en el caso de máximos, la DVEG para mínimos engloba a las familias Gumbel, Fréchet y Weibull para mínimos, y es el valor de c quien determina la familia límite.

Una segunda opción es, dada una muestra $\{x_1, x_2, \ldots, x_n\}$ correspondiente a mínimos de variables aleatorias iid, ajustar la DVEG para máximos a $\{y_1, y_2, \ldots, y_n\}$ donde $y_i = -x_i$. Hay que tener precaución en la interpretación de las estimaciones e inferencias al regresar del máximo al mínimo. Las estimaciones vía máxima verosimilitud resultantes para los parámetros de la DVEG de máximos coinciden con los de la DVEG de mínimos, salvo la corrección de signo para $\tilde{b} = -b$.

Como se mencionó anteriormente, el teorema de Fisher y Tippet identifica las distribuciones límite del máximo reescalado de una sucesión de variables aleatorias iid. En la literatura existen resultados concernientes a F, que caracterizan a la función de distribución límite de F_{M_n} bajo ciertas condiciones referentes al comportamiento de $\bar{F} = 1 - F$ cercano a x^+. Más aún, los resultados muestran cómo calcular las constantes de normalización que permiten estabilizar la función de distribución del máximo, logrando con esto la convergencia en distribución a una de las DVE. Este tópico de teoría de extremos se conoce como Caracterización de los Dominios de Atracción. Una exposición muy completa de éste se encuentra en Embrechts, *et al.* [8].

El resumen que hemos presentado pretende dar un panorama general de la teoría clásica de valores extremos, también conocida como modelación por bloques. Un tratado muy completo

del tópico se encuentra en Embrechts, et. al. [8]. Una exposición de la teoría de valores extremos, desde un punto de vista de modelación estadística, se encuentra en Coles, [6]. Para consultar detalles de las demostraciones del resultado principal de Fisher y Tippet, se refiere al lector a Resnick, [18]. Un estudio de valores extremos con aplicaciones en ingeniería se encuentra en Castillo, *et al.* [3].

1.2. Cercanía entre distribuciones

Recordando, la DVEG representa una reformulación la cual engloba a las tres distribuciones posibles para el límite del máximo de una sucesión de variables aleatorias que se suponen iid. Los casos Weibull y Fréchet corresponden a valores negativos y positivos, respectivamente, del parámetro de forma c de la DVEG. La distribución Gumbel se obtiene tomando el límite de la DVEG cuando c tiende a cero. Esto es:

$$\lim_{c \longrightarrow 0} \exp\{-[1 + c(\frac{z-a}{b})]^{-\frac{1}{c}}\} = \exp\{-\exp[-(\frac{z-a}{b})]\}. \tag{1.20}$$

Es decir, existen familias Fréchet y Weibull muy cercanas a una familia Gumbel.

En sentido inverso, una cuestión de interés es si dada una familia Weibull o Fréchet, podemos encontrar una familia Gumbel que esté lo suficientemente cerca de la primera, en términos de su función de distribución. El objeto de esto es para determinar un rango de valores para c alrededor de cero para el cual se tiene una representación adecuada de la distribución (y consecuentemente de los cuantiles) Weibull o Fréchet a través de un modelo Gumbel antes de alcanzar el límite.

Para tal efecto, se realizó un algoritmo computacional el cual encuentra la distribución Gumbel que minimiza una distancia entre la distribución de referencia (Weibull o Fréchet) y la distribución Gumbel. La distancia que se propone para establecer la cercanía de las distribuciones es una distancia de tipo Cramer-Von Mises (ver D'Agostino y Stephens [7]), dada por la siguiente expresión:

$$D(\Phi, \Delta) = \int \frac{[\Delta(z) - \Phi(z)]^2}{(1 - \Phi(z))} d\Phi(z). \tag{1.21}$$

Esta distancia es un promedio de la diferencia al cuadrado de la función de distribución Gumbel respecto a la distribución de referencia, ponderada por los valores de la cola derecha de la distribución de referencia.

Se seleccionaron como distribuciones base las DVE correspondiente a los parámetros de la DVEG $a = 1$, $b = 1$ y $c = 0.1$, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01, -0.01, -0.02, -0.03, -0.04, -0.05, -0.06, -0.07, -0.08, -0.09 y -0.1. En la gráfica superior de la Figura 1.1 se presentan las distancias del modelo Gumbel más cercano al modelo de referencia con base en la distancia propuesta. La distancia disminuye a medida que c se acerca a cero, y es ligeramente asimétrica respecto a $c = 0$. Las discrepancias de los modelos Fréchet considerados respecto a la distribución Gumbel más cercana son ligeramente menores en comparación con los modelos Weibull. Es decir, se encontraron distribuciones Gumbel más cercanas a los modelos Fréchet que a los modelos Weibull.

Por otra parte, en la gráfica inferior de la Figura 1.1 se presentan los parámetros de la Gumbel más cercana con base en la distancia propuesta. Notamos que el parámetro μ de la familia Gumbel más cercana es prácticamente igual al valor de $a = 1$, mientras que el valor de

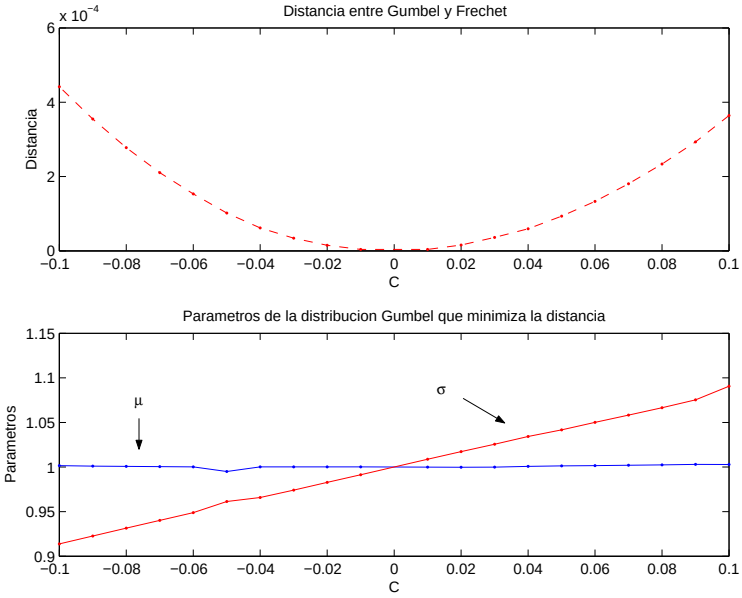


Figura 1.1: Distancia entre distribuciones y parámetros de la distribución Gumbel que minimiza la distancia considerada

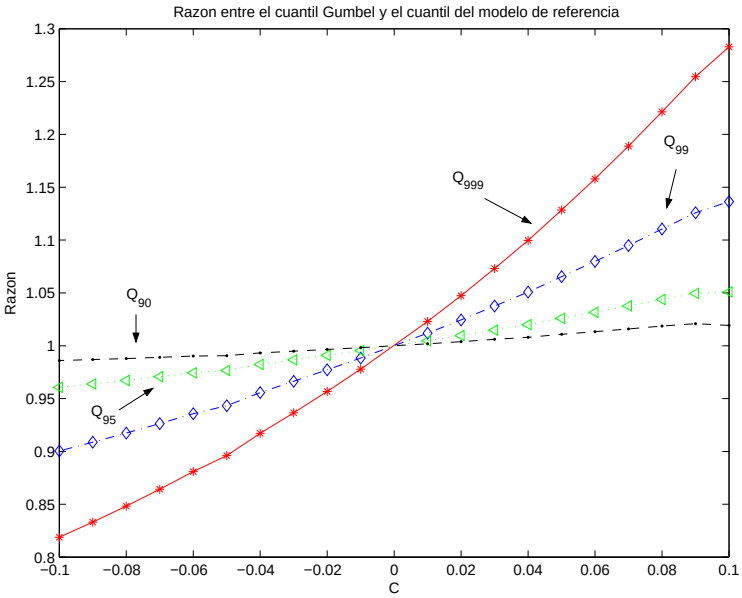


Figura 1.2: Razón de los cuantiles de la Gumbel que minimiza la distancia considerada

σ del modelo Gumbel más cercano vará de acuerdo al valor de c, siendo menor que $b = 1$ para los casos Weibull y mayor que $b = 1$ para los casos Fréchet, teniendo mayores discrepancias de $b = 1$ cuando c está más alejado de $c = 0$.

La Figura 1.2 presenta los cocientes de Q_{90}, Q_{95}, Q_{99} y Q_{999} del modelo Gumbel más cercano respecto a la familia de referencia acorde a los diferentes valores de c. Notamos que para Q_{90} los cocientes están cercanos a 1 para todos los modelos de referencia considerados. Para Q_{95} los cocientes siguen estando cerca de 1, difiriendo en los valores de $c = 0.1$ y $c = -0.1$ en 5 %. Para valores de $|c| \leq 0.04$ el cuantil 99 de la distribución Gumbel más cercana varía hasta en un 5 % respecto a la distribución de referencia; en contraste, para Q_{999} se tienen discrepancias

hasta en un 10 %. Para Q_{999} la distribución Gumbel más cercana correspondiente a los valores de $|c| \leq 0.02$ difiere hasta en un 5 % respecto a la distribución de referencia. Finalmente vemos que los cuantiles de la distribución Gumbel están más cerca cuando se tiene como referencia a un modelo Weibull que a un Fréchet.

Basados en la distancia propuesta, para valores de $|c| \leq 0.02$ se tienen buenas representaciones de cuantiles grandes (incluso hasta Q_{999}) de las familias Fréchet y Weibull a través de la familia Gumbel.

1.3. Herramienta estadística

1.3.1. Funciones de logverosimilitud para la DVEG y las DVE

Bajo los supuestos de iid y que la distribución del máximo es la DVEG, usando la densidad como aproximación a la verosimilitud, la función de logverosimilitud para los parámetros de la DVEG está dada por la siguiente expresión:

$$\log(a, b, c; z_1, \ldots, z_n) = -n\log(b) - (1 + \frac{1}{c})\sum_{i=1}^{n}\log[1 + c(\frac{z_i - a}{b})] - \sum_{i=1}^{n}[1 + c(\frac{z_i - a}{b})]^{\frac{-1}{c}}, \quad (1.22)$$

donde $c \neq 0$, $1 + c(\frac{z_i - b}{a}) > 0$, $i = 1, \ldots, n$.

De las expresiones para la logverosimilitud no podemos encontrar los estimadores de máxima verosimilitud analíticamente, por lo que se requieren algoritmos de maximización para hallar los EMV numéricamente. Además, la DVEG no cumple los requerimientos de regularidad que permiten aproximar la función de distribución de los estimadores de máxima verosimilitud por la distribución normal multivariada.

Existen resultados en la literatura que exploran la aproximación normal a los estimadores de máxima verosimilitud a pesar que éstos no cumplan con las condiciones de regularidad. Para la DVEG, dichos resultados (debidos a Smith [16]) muestran que esta aproximación es válida para valores de c mayores a -0.5, valores para las cuales se dejan de considerar distribuciones de colas muy ligeras, y que en general no se encuentran casos de este tipo en la práctica según Coles [6]. Lo anterior da sustento al método delta para la obtención de intervalos de confianza para los parámetros a, b y c de la DVEG, y a través de una transformación de éstos, para los cuantiles de interés.

También, presentamos las funciones de logverosimilitud para los parámetros de las DVE:

$$Gumbel: \quad \log L(\mu, \sigma; z_1, \ldots, z_n) = -\sum_{i=1}^{n}\exp[-(\frac{z_i - \mu}{\sigma})] - \sum_{i=1}^{n}(\frac{z_i - \mu}{\sigma}) - n * \log(\sigma), \quad (1.23)$$

$$Fréchet: \quad \log L(\mu, \sigma, \beta; z_1, \ldots, z_n) = -\sum_{i=1}^{n}(\frac{z_i - \mu}{\sigma})^{-\beta} + n * \log(\sigma/\beta) - (\beta + 1)\sum_{i=1}^{n}\log(\frac{z_i - \mu}{\sigma}), \quad (1.24)$$

$$Weibull: \quad \log L(\mu, \sigma, \beta; z_1, \ldots, z_n) = -\sum_{i=1}^{n}(\frac{\mu - z_i}{\sigma})^{\beta} + n * \log(\sigma/\beta) + (\beta - 1)\sum_{i=1}^{n}\log(\frac{\mu - z_i}{\sigma}), \quad (1.25)$$

donde $\mu \in R$; $\sigma, \beta > 0$.

De igual forma, se requiere de métodos numéricos para hallar los estimadores de máxima verosimilitud.

1.3.2. Función de verosimilitud perfil

Uno de los problemas comúnmente encontrados en la inferencia estadística se refiere a las inferencias en presencia de parámetros de estorbo. El panorama del problema es el siguiente: se tiene una familia de modelos estadísticos paramétricos, $f(x;\theta)$, donde $\theta \in \Theta$ es un vector de parámetros, en el cual nuestro principal interés se centra en realizar inferencias para un vector θ_0 de componentes de θ, generalmente unidimensional. Los componentes de θ que no están en θ_0, digamos θ_1, los cuales consideraremos que no son de interés, son los llamados parámetros de estorbo. Las expresiones para la función de verosimilitud, en presencia de modelos multiparamétricos, son en general complicadas de manipular, trayendo como consecuencia que no se pueden aislar expresiones que contengan exclusivamente a θ_0, haciendo que las inferencias para θ_0 se vean afectadas por θ_1.

Una forma sencilla y muy útil de tratar con este problema es a través de la verosimilitud perfil. La verosimilitud perfil es una función exclusiva de θ_0 que se obtiene a partir de la verosimilitud completa, para la cual para cada valor fijo de θ_0 se usa como valor de θ_1 el estimador de máxima verosimilitud $\hat{\theta}_1$ restringido a este valor fijo θ_0 (en lo sucesivo escribiremos EMV para denotar el estimador de máxima verosimilitud).

Formalmente, tenemos la siguiente definición (tomadas de Sprott [17] y Pawitan [14]):

Definición 1.1 *Sea $L(\theta; x_1, x_2, \ldots x_n)$ la función de verosimilitud para $\theta = \{\theta_0, \theta_1\}$. Definimos a la verosimilitud perfil de θ_0, $L_p(\theta_0; x_1, x_2, \ldots x_n)$ como*

$$L_p(\theta_0) = L(\theta_0, \hat{\theta}_1(\theta_0); x_1, x_2, \ldots x_n). \tag{1.26}$$

De igual manera, definimos la verosimilitud perfil relativa de θ_0 como:

$$R_p(\theta_0) = \frac{L(\theta_0, \hat{\theta}_1(\theta_0); x_1, x_2, \ldots x_n)}{L(\hat{\theta}; x_1, x_2, \ldots x_n)}, \tag{1.27}$$

donde $\hat{\theta}_1(\theta_0)$ denota el EMV restringido de θ_1 para un valor fijo de θ_0, y $\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1)$ es el Estimador de Máxima Verosimilitud de $\theta = (\theta_0, \theta_1)$.

La información contenida en la muestra respecto a θ_0 se presenta mediante la gráfica de los puntos (θ_0, $R_p(\theta_0)$) (ver gráfica 1.3). La gráfica de verosimilitud perfil relativa de θ_0 indica la plausibilidad de un valor fijo de θ_0, digamos θ_0^0, respecto al valor de θ_0 máximo verosímil. La interpretación de un nivel de plausibilidad es el porcentaje de credibilidad que representa el valor θ_0^0 respecto a $\hat{\theta}_0$ para la muestra obtenida.

Algunas de las características más importantes de la verosimilitud perfil son:

- La estimación de máxima verosimilitud obtenida a partir de la verosimilitud perfil coincide con la estimación de máxima verosimilitud de la verosimilitud completa.

- El estadístico de razón de verosimilitud basado en la verosimilitud perfil $2 * \{\log L_p(\hat{\theta}_0) - \log L_p(\theta_0)\}$ equivale al estadístico de razón de verosimilitud para la hipótesis $\theta = \theta_0$.

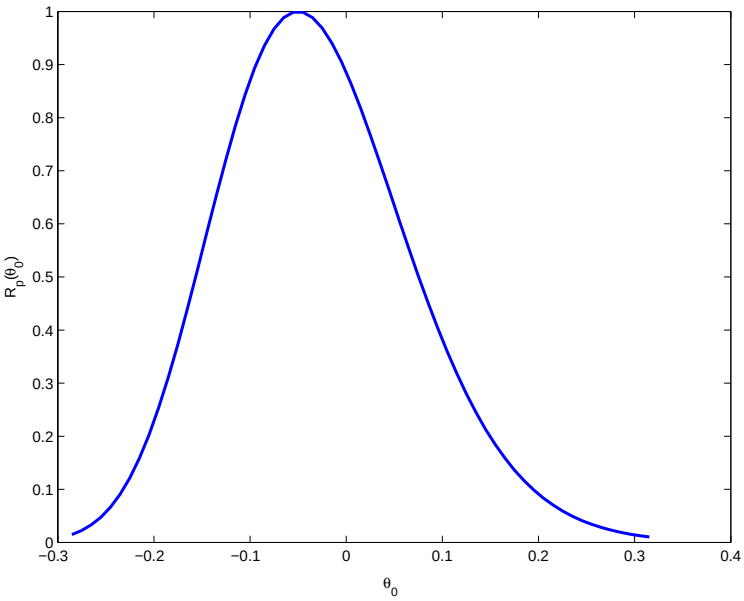


Figura 1.3: Verosimilitud perfil relativa de θ_0

- Una región de verosimilitud basada en la verosimilitud perfil $R^k = \{\theta_0 : L_p(\theta_0) > k\}$ es, en general, una región aproximada de confianza para θ_0. (Ver siguiente sección)

La verosimilitud perfil tiene la ventaja de que es una manera sencilla de tratar a los parámetros de estorbo, además de que es simple calcularla numéricamente. La verosimilitud perfil relativa para θ_0 se obtiene fijando valores de θ_0 en una rejilla y maximizando la verosimilitud sobre los parámetros restantes, además de estandarizar de acuerdo al valor de θ que maximiza la verosimilitud.

Más detalles sobre la verosimilitud perfil, sus propiedades y sus implicaciones se pueden consultar en Sprott [17] y Pawitan [14].

1.3.3. Regiones de verosimilitud-confianza

Bajo el enfoque de verosimilitud, la información contenida en la muestra respecto a los parámetros de interés se presentan por medio de la gráfica de la función de verosimilitud relativa. En el caso de parámetros de estorbo, teniendo bajo consideración la función de verosimilitud perfil, es ésta la función relevante que contiene la información paramétrica de interés.

Una manera de resumir la información contenida en la verosimilitud perfil relativa es a través de regiones de verosimilitud. La siguiente definición es tomada de Sprott [17].

Definición 1.2 *Definimos una región de nivel k de verosimilitud, R^k como:*

$$R^k = \{\theta : L(\theta) \geq k\}. \tag{1.28}$$

En particular, para la verosimilitud perfil se tiene que

$$R^k = \{\theta_0 : L_p(\theta_0) \geq k\}. \tag{1.29}$$

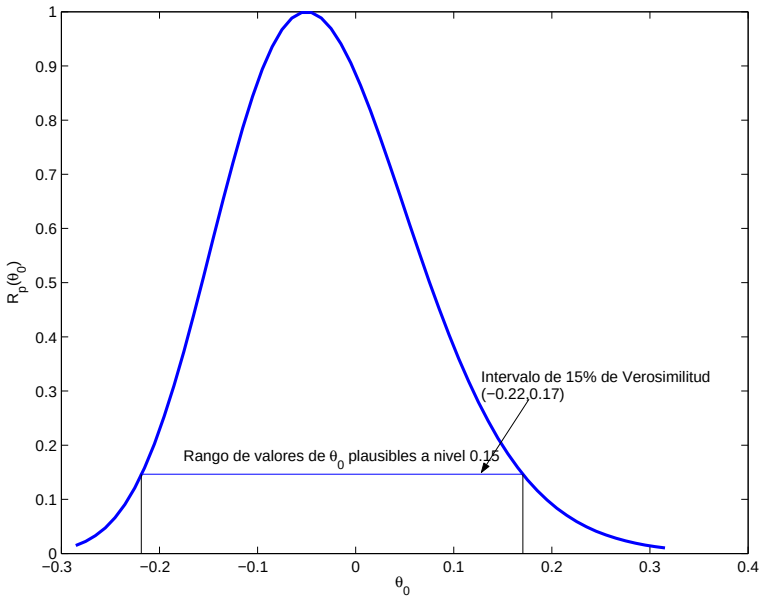


Figura 1.4: Verosimilitud perfil relativa de θ_0

Una región de verosimilitud es un conjunto de puntos de θ_0 para los cuales se tiene un nivel de verosimilitud mayor o igual a k. En otras palabras, las regiones de verosimilitud son una forma de discriminar los valores de θ_0 que resultan plausibles a cierto nivel k de los que no lo son. Los puntos contenidos en una región específica de verosimilitud son preferibles a todos aquellos que no estén contenidos en esta región, ya que tienen plausibilidad más alta. De la figura 1.4 , para un nivel $k = 0.15$ de verosimilitud obtenemos que valores para $\theta_0 \in$ [-0.22, 0.17] son preferibles a aquellos que estén fuera de ese intervalo. Para el caso en que θ_0 es unidimensional, las regiones de verosimilitud son, por lo general, intervalos de verosimilitud.

Una cuestión importante respecto a las regiones de verosimilitud es que la elección del nivel k es arbitraria. El significado del nivel k, en este contexto, se relaciona exclusivamente con un nivel de plausibilidad. Sin embargo, es posible establecer una relación entre el nivel k y un nivel de confianza (y con ello, dar un significado de región de confianza a una región de verosimilitud) a través del siguiente resultado, tomado de Coles [6].

Teorema 1.3 *Sea $X_1, X_2, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas con función de distribución F(x; θ). Sea $\hat{\theta}$ el EMV de $\theta = \{\theta_0, \theta_1\}$. Entonces, bajo condiciones de regularidad, para n suficientemente grande, se tiene que:*

$$D_p(\theta_0) = 2\{l(\hat{\theta}) - l_p(\theta_0)\} \sim \chi^2_s \tag{1.30}$$

donde $l(\hat{\theta})$ y $l_p(\theta_0)$ representan la log-verosimilitud y la log-verosimilitud perfil respectivamente, y s es la dimensión de θ_0

El teorema anterior permite construir regiones de confianza para el parámetro θ_0: si Q_{z_p} es el cuantil $p * 100\,\%$ de la distribución χ^2_s, la región $S^\beta = \{\theta_0 : D_p(\theta_0) \leq z_p\}$ resulta ser una región de confianza del nivel $p*100\,\%$. Puesto que $2\{l(\hat{\theta}) - l_p(\theta_0)\} \leq z_p$ si y sólo si $L_p(\theta_0) \geq \exp(-z_p/2)$, una región de confianza también resulta ser una región de verosimilitud de nivel $\exp(-z_p/2)$. En particular, para el caso en que θ_0 es unidimensional, un nivel de confianza del 95 % implica un nivel de verosimilitud de 0.15.

Para más detalles del significado de las regiones de verosimilitud-confianza, se refiere al lector a Sprott [17].

1.3.4. Métodos gráficos de diagnóstico

En términos de modelación, buscamos representar adecuadamente un fenómeno a través de un modelo estadístico, por medio del cual se realizan las inferencias de interés. Estas inferencias pueden variar de acuerdo al modelo elegido, por lo cual es necesario que el modelo ajuste lo mejor posible, al menos para las observaciones a la mano. A continuación presentamos las formas generales de validar el ajuste de un modelo adoptadas en la presente tesis.

Consideremos una muestra aleatoria (realizaciones independientes e idénticamente distribuidas) ordenada $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, de una población con función de distribución F. La gráfica de probabilidad consiste en la gráfica del conjunto de puntos

$$\{(\hat{F}(x_{(i)}), \frac{i-0.5}{n}) : i = 1, \ldots, n\}, \tag{1.31}$$

donde $\hat{F}(x_{(i)})$ y $\frac{i-0.5}{n}$ representan la probabilidad acumulada de $x_{(i)}$ bajo la función de distribución estimada y la función de distribución empírica, respectivamente. De igual forma, la gráfica de Cuantil-Cuantil (gráfica QQ) consiste en la gráfica del conjunto de puntos

$$\{(\hat{F}^{-1}(\frac{i-0.5}{n}), x_{(i)}) : i = 1, \ldots, n\}, \tag{1.32}$$

donde $\hat{F}^{-1}(\frac{i-0.5}{n})$ y $x_{(i)}$ representan el cuantil $(\frac{i-0.5}{n})$ bajo el modelo estimado y el cuantil empírico, respectivamente.

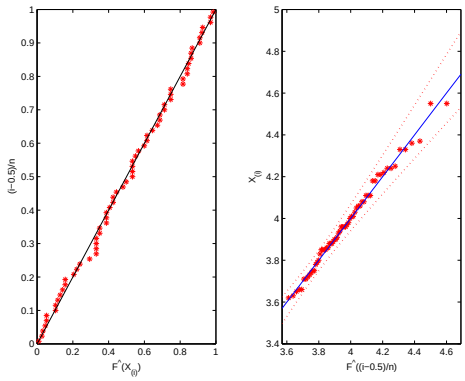


Figura 1.5: Gráfica de papel de probabilidad (izquierda) y gráfica QQ (derecha)

El propósito de ambas gráficas es comparar las probabilidades o los cuantiles del modelo estimado contra las estimaciones empíricas. Si el modelo propuesto representa razonablemente a las observaciones, los puntos graficados deberán estar cerca de la línea identidad (ver Figura 1.5). Para la gráfica QQ es común agregar intervalos de confianza para los cuantiles para considerar la variabilidad de las observaciones en el ajuste.

Por otro lado, para la DVEG se tiene una gráfica particular para la validación de ajuste de un modelo. Retomando la ecuación (1.15), haciendo $y_p = \log(p)$ se tiene que:

$$Q_p = \begin{cases} a - \frac{b}{c} * (1 - y_p^{-c}), & c \neq 0, \\ a - b * (\log(y_p)), & c = 0. \end{cases} \tag{1.33}$$

La gráfica del conjunto de puntos

$$\{\log(y_p), Q_p\} \tag{1.34}$$

tiene por nombre gráfica de niveles de retorno (Figura 1.6). Cada nivel de retorno Q_p (eje vertical) tiene asociado un periodo de retorno 1/(1-p) (eje horizontal). Un periodo de retorno $1/(1-p)$ se interpreta como el número de años que en promedio se requieren para que el nivel de retorno asociado a éste, Q_p, se supere una vez. La gráfica del modelo estimado es lineal para el caso Gumbel, convexa para el caso Weibull y concava para el caso Fréchet. La gráfica aquí descrita se utiliza como presentación del modelo estimado y para validación del mismo, agregando en la gráfica las estimaciones empíricas de los cuantiles e intervalos de confianza para los mismos.

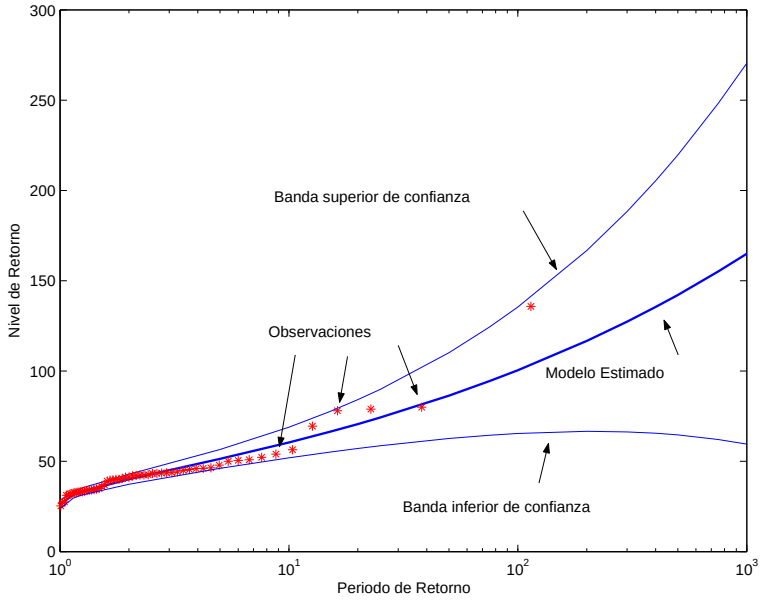


Figura 1.6: Ejemplo de una gráfica de niveles de retorno

1.4. Modelación estadística

1.4.1. Proceso de modelación

Box [1] describe a la modelación estadística como un proceso evolutivo de aprendizaje científico el cual involucra estimación y crítica. En pocas palabras, dicho proceso consiste, en primera instancia, considerar distintos modelos, ya sea empíricos o teóricos, para los cuales se realiza la estimación. Una vez hecho esto es necesario juzgar los modelo con base en el ajuste de los mismos y las consecuencias de sus inferencias, para verificar su congruencia o sugerir modificaciones en caso necesario. Finalmente, se elige el modelo que haga más sentido al contexto del problema.

Este proceso resulta adecuado en el ámbito de extremos debido a que se tienen 3 posibles modelos (Weibull, Fréchet o Gumbel) para representar el comportamiento de una muestra de máximos que se suponen iid. Debido al desconocimiento de la verdadera distribución límite del máximo de las variables aleatorias consideradas, mediante este proceso se pretende sugerir

modelos sensatos para el fenómeno de interés, hacer inferencia bajo estos modelos, validarlos con los datos observados, comparar estos modelos en términos del ajuste a los datos y buena descripción que hagan del fenómeno de interés, para finalmente seleccionar al modelo que mejor explique a los datos y al fenómeno.

1.4.2. Parámetros umbral

Un parámetro umbral es un parámetro caracterizado por delimitar el rango de valores posibles que puede tomar una variable aleatoria. Como ejemplo, en el caso de la función de distribución Fréchet (ecuación 1.9), la variable aleatoria se restringe a ser mayor a μ, mientras que en el caso Weibull (ecuación 1.10), la variable se restringe a ser menor a μ. Este parámetro cobra especial importancia, ya que el hecho de restringir el soporte de la distribución trae como consecuencia problemas tanto desde el punto de vista de modelación como de estimación.

Desde el punto de vista de estimación, un modelo con un parámetro umbral tiene el problema de que no se cumplen las condiciones de regularidad que permiten aproximar la función de distribución de los estimadores de máxima verosimilitud por la distribución normal multivariada. Además, se presentan casos donde la función de verosimilitud basada en la aproximación por medio de la densidad, tienen discontinuidades infinitas cuando el parámetro umbral se acerca a la observación mínima o máxima, trayendo como consecuencia que métodos numéricos de búsqueda adoptados para encontrar los estimadores de máxima verosimilitud fallen en su propósito.

Desde el punto de vista de modelación, un parámetro umbral se introduce en un modelo como artificio matemático para permitir un mejor ajuste del modelo acorde a observaciones (Meeker & Escobar, [13]). Por ejemplo, consideremos el caso Fréchet de dos parámetros

$$\Phi(z) = \begin{cases} 0, & z \leq 0, \\ \exp\{-(\frac{z}{\sigma})^{-\alpha}\} & z > 0, \end{cases} \tag{1.35}$$

donde el soporte de la distribución son los reales positivos. Para una variable aleatoria positiva, es idóneo ajustar un modelo con las características del soporte como el caso Fréchet de dos parámetros; sin embargo, el ajuste de este modelo no siempre es adecuado. Por lo anterior, el considerar la familia Fréchet de tres parámetros (1.9) permite hacer flexible la modelación, y con ello lograr un mejor ajuste.

Puesto que un parámetro umbral restringe el soporte de la distribución, el parámetro umbral adquiere intrínsecamente un significado acorde con el contexto del problema. Dependiendo del contexto, el problema consiste en que el ajuste de un modelo con parámetros umbral en general puede no hacer sentido respecto al significado del mismo. Por ejemplo, refiriéndonos a Coles [6], él modela el comportamiento del mínimo de observaciones referentes a resistencia de fibras a través de la DVEG, encontrando estimaciones negativas para el menor valor que puede tomar el mínimo de resistencia de fibras. Coles expresa la cautela que hay que tener en estas estimaciones, ya que en el contexto, este valor mínimo no puede tomar valores negativos. En acuerdo con Meeker y Escobar [13], en estos casos es importante ajustar un modelo restringiendo el espacio parametral del parámetro umbral de acuerdo al significado de éste en el contexto.

Box y Cox [2] enfatizan diferenciar el carácter lógico de los parámetros y recomiendan realizar inferencias en dos fases: primero estimar los parámetros que tengan sentido de acuerdo al contexto, y con base en estas estimaciones, fijar valores de dichos parámetros para realizar posteriormente las inferencias de interés. Por otro lado, para el caso Weibull (1.10), Lawless [11] recomienda explorar la robusticidad en los cuantiles ante cambios en μ cercanos a $\hat{\mu}$ para

verificar si el modelo reducido fijando $\mu = \hat{\mu}$ ajusta bien a las observaciones. En los casos donde se tenga robusticidad, se podría hacer uso de un modelo más simple y regular. Para el mismo caso Weibull, Meeker y Escobar [13] consideran dos alternativas: ajustar de inicio un modelo con menos parámetros que haga sentido al contexto del problema, o ajustar un modelo restringiendo el espacio parametral a un subconjunto de ellos que haga sentido con el contexto.

En conjunto, las recomendaciones anteriores , aplicadas a las DVE Weibull y Fréchet (1.10) y (1.9), se refieren en primera instancia a ajustar un modelo que considere el significado de los parámetros de acuerdo al contexto del problema. Ajustar la familia Fréchet restringiendo μ a ser no negativo al considerar variables aleatorias positivas tiene como propósito inicial considerar modelos que hagan sentido respecto al significado de μ. Luego, se recomienda explorar la sensibilidad de la inferencia de los cuantiles ante cambios en valores plausibles de μ.

Los puntos aquí comentados, más los problemas de estimación respecto a los parámetros umbral, se retomarán posteriormente.

1.4.3. Propuesta de modelación del máximo por bloques

Como vimos en la Sección 1.1, la distribución a la cual converge el máximo reescalado es una de las DVE, donde cada una de ellas representa un comportamiento distinto del máximo. La DVEG es simplemente una reformulación que engloba a las 3 DVE. En aplicaciones de la teoría de extremos es común que se adopte la DVEG como modelo para realizar las inferencias de interés debido a que engloba la incertidumbre de una elección de la distribución límite. Las inferencias basadas en verosimilitud a través de la DVEG pueden comprender simultaneamente a las 3 DVE y, con ello, abarcar 3 comportamientos distintos del fenómeno.

Además, para los modelos Weibull y Fréchet, el parámetro de localización μ adquiere automáticamente un significado de acuerdo al contexto del fenómeno como consecuencia de que delimita el soporte de la distribución. Según lo expuesto en la Sección 1.4.2, y en particular cuando de variables aleatorias positivas se trata (para la presente tesis, se consideran variables positivas como niveles de acumulación de lluvia y niveles de acumulación de nieve), es recomendable ajustar en primera instancia un modelo que haga sentido al fenómeno. Para el caso Fréchet, la noción del parámetro μ es clara, por lo cual es recomendable ajustar de primera instancia un modelo restringiendo μ a ser no negativo. Para el caso Weibull, no resulta evidente restringir el espacio parametral de μ de tal forma que haga sentido el fenómeno en cuestión. Además, el modelo Weibull (al igual que el modelo Gumbel) asigna probabilidades positivas a valores negativos de la variable en cuestión por no estar acotado inferiormente su soporte, lo cual tampoco resulta sensato en términos de las variables positivas. Sin embargo, para estos casos se justifica el uso de estos modelos para variables positivas cuando la probabilidad que se le asigne a valores negativos de la variable sea muy pequeña, de tal modo que no afecte al resto de las probabilidades. Al ajustar una familia Weibull y Fréchet, se considerará hacer inferencias para los cuantiles fijando diferentes valores de μ, para verificar la sensibilidad de las inferencias ante cambios en μ.

Tomando en cuenta a Box, dentro del contexto de modelación del máximo por bloques, consideraremos incialmente el ajuste de la DVEG. Puesto que sólo una de las DVE corresponde a la distribución límite del máximo reescalado bajo el supuesto de iid, usaremos la reparametrización DVEG para realizar estimación respecto al parámetro de forma c, y con ello restringirnos a las subclases sugeridas de la estimación de c que hagan sentido al problema en mano. De ahí prosigue el proceso de estimación y crítica, para finalmente seleccionar el modelo que tenga mejor representatividad.

Resumiendo los párrafos anteriores, se tiene la siguiente propuesta de modelación estadística:

- Aprovechar la DVEG para sugerir una o más de las DVE para aproximar la distribución del máximo. Considerar un modelo de acuerdo a los sugeridos que haga sentido al contexto del problema.
- Hacer las inferencias de los parámetros de interés bajo las DVE consideradas.
- Validar los modelos estimados con los datos y analizar sus consecuencias.
- Seleccionar al mejor modelo DVE para la distribución del máximo.

De esta forma, procederemos de la siguiente manera. Consideraremos una muestra de máximos de tamaño n para los cuales asumiremos que se cumplen los supuestos iid para las variables aleatorias en cuestión. Para esta muestra:

1. Ajustaremos la DVEG a los máximos y haremos inferencias para los parámetros a, b y c. En especial nos enfocaremos en la verosimilitud perfil relativa del parámetro c, ya que a través de ella se sugerirán los modelos límite para el máximo reescalado. Valores de c muy cercanos a 0 corresponden a modelos Fréchet y Weibull que prácticamente resultan indistinguibles a un modelo Gumbel. En el caso que el modelo Gumbel resulte plausible, la gráfica de verosimilitud perfil hará plausible a las tres familias. Sin embargo, se sugerirán sólo la familia correspondiente al EMV $\hat{c}$ y a la familia Gumbel, ya que la familia restante tiene como mejor modelo uno que es indistinguible del modelo Gumbel, además de que las estimaciones de sus parámetros resultan ser de magnitud muy grande, carentes de interpretabilidad.

2. Una vez tomados en cuenta los modelos sugeridos, se procederá en acuerdo con Box, a realizar la estimación y comparar el ajuste y las consecuencias de las inferencias del modelo o los modelos sugeridos. En el caso que $\hat{c}$ sea positivo, se considerará inicialmente el ajuste de la familia Fréchet restringiendo el espacio parametral de μ a ser no negativo cuando se trate de datos no negativos, enfatizando la posibilidad del ajuste del modelo Fréchet de dos parámetros. En cuanto a estimación, se grafican la verosimilitud perfil relativa de los parámetros μ, σ, β de la familia Weibull o Fréchet, según sea el caso, y se considerarán valores altamente plausibles para μ para explorar sus consecuencias en las inferencias respecto a los cuantiles. Posteriormente se graficarán las verosimilitudes perfil relativas de Q_{95} y Q_{99}. En cuando a la validación de los modelos, se presentarán gráficas de probabilidad, gráficas QQ y gráficas de niveles de retorno.

3. Finalmente, basados en lo anterior se determinará si el o los modelos considerados son adecuados respecto a lo observado. Para el caso Fréchet, en caso que la familia de dos parámetros no produzca una buena descripción del fenómeno, consideraremos el ajuste de la familia Fréchet de tres parámetros para tener una mejor representatividad del fenómeno, realizando nuevamente la estimación y la validación del modelo.

La presente tesis adopta la función de verosimilitud perfil para las inferencias de interés, criterio poco explorado en el ámbito de extremos. Respecto al problema de las discontinuidades en las densidades, en la presente tesis se restringe μ a estar a lo más a .001 unidades de la observación máxima (o mínima, según sea el caso), logrando con esto que la maximización a

través de métodos numéricos encuentre un máximo local para el cual adoptaremos a los valores correspondientes de los parámetros como los EMV.

Puesto que en aplicaciones de la teoría clásica de extremos es común el uso de la DVEG para realizar inferencias respecto a los cuantiles, se adoptará a la DVEG para establecer las comparaciones pertinentes respecto a la propuesta aquí realizada. Dado que en aplicaciones de teoría de extremos es común el uso de los intervalos obtenidos a través del método delta, se harán las inferencias basados en este método a manera de comparación.

Capítulo 2

Ejemplos de modelación

En el presente capítulo se aplicará la metodología de inferencia estadística presentada en el capítulo anterior a diversos conjuntos de datos obtenidos de variadas fuentes bibliográficas, así como de datos de lluvia obtenidos en el estado de Michoacán a través del Instituto Mexicano de Tecnología del Agua (IMTA). Se presentarán primero dos ejemplos que fueron descritos adecuadamente por un modelo Weibull, posteriormente dos ejemplos modelados a través de la familia Gumbel, y finalmente tres ejemplos para los cuales el modelo Fréchet fue el que mejor explicó a los fenómenos de interés. Para todos los ejemplos se comentará la influencia del parámetro umbral de las familias Weibull o Fréchet en la inferencia sobre cuantiles de interés.

Denotaremos al EMV como el estimador de máxima verosimilitud del parámetro del cual se esté hablando. La DVEG es la distribución de valores extremos generalizada; $\hat{a}, \hat{b}, \hat{c}$ los estimadores de máxima verosimilitud de sus parámetros de localización, escala y forma respectivamente. Los estimadores de máxima verosimilitud de los parámetros de localización, escala y forma de la distribución Fréchet o Weibull, según sea el caso, son $\hat{\mu}, \hat{\sigma}, \hat{\beta}$, respectivamente; $\hat{Q}_{95}, \hat{Q}_{99}$ son los estimadores de máxima verosimilitud de los cuantiles 95 y 99, respectivamente.

Se presentarán estadísticas descriptivas de cada conjunto de datos bajo estudio. Para ellas, n representa el total muestral, $x_{(1)}$ y $x_{(n)}$ representan la observación mínima y máxima respectivamente; $\tilde{Q}_p$ representa el cuantil empírico de probabilidad p; $\bar{X}$ y s representan la media muestral y la desviación estándar muestral, respectivamente.

2.1. Niveles de lluvia en El Temazcal Charo, Michoacán.

La estación meteorológica de El Temascal Charo se encuentra ubicada en la zona noreste de la cuenca de Morelia. Dicha estación cuenta con un periodo de registro de niveles de acumulación diaria de lluvia de 39 años, comprendidos entre 1965 y 2003. Se consideró el máximo anual de éstos (Figura 2.1) para realizar el análisis de valores extremos.

n	$x_{(1)}$	$\tilde{Q}_{25}$	$\tilde{Q}_{50}$	$\bar{X}$	$\tilde{Q}_{75}$	$x_{(n)}$	s
39	6	47.67	64.5	59.15	68.37	93	17.91

Cuadro 2.1: Estadísticas descriptivas

En la Figura 2.1 notamos un decremento ligero de la magnitud de las observaciones a través del tiempo, encontrando un nivel muy bajo en $x_{(1)} = 6$ mm. (la observación más reciente) en

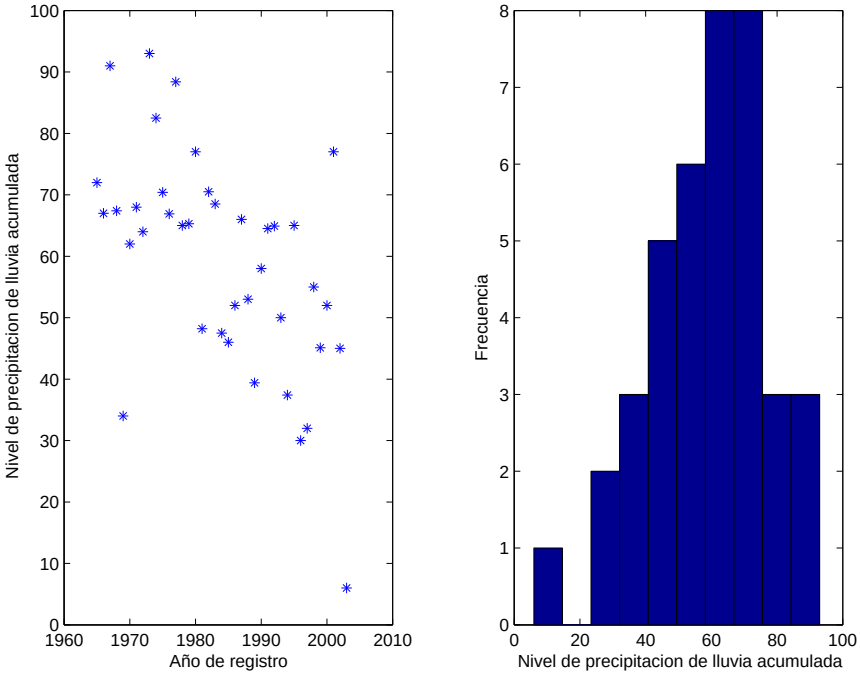


Figura 2.1: Máximos anuales de niveles de lluvia en El Temazcal Charo.

comparación al resto de los datos. Esta observación hace que $\bar{X}$ se recorra a la izquierda de $\tilde{Q}_{50}$. Notamos una ligera asimetría en el histograma de las observaciones.

Modelo	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{Q}_{95}$	$\hat{Q}_{99}$
DVEG	54.07	18.9	-0.42	86.3	92.8

Cuadro 2.2: Estimaciones de máxima verosimilitud

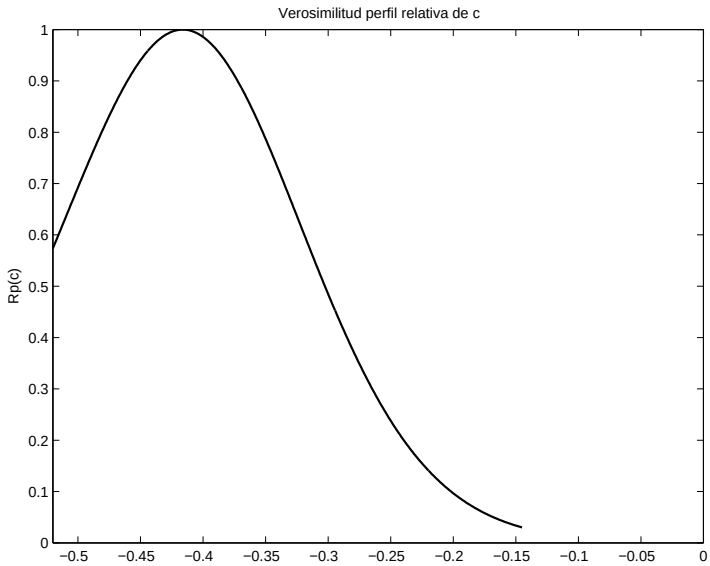


Figura 2.2: Verosimilitud perfil del parámetro de forma de la DVEG

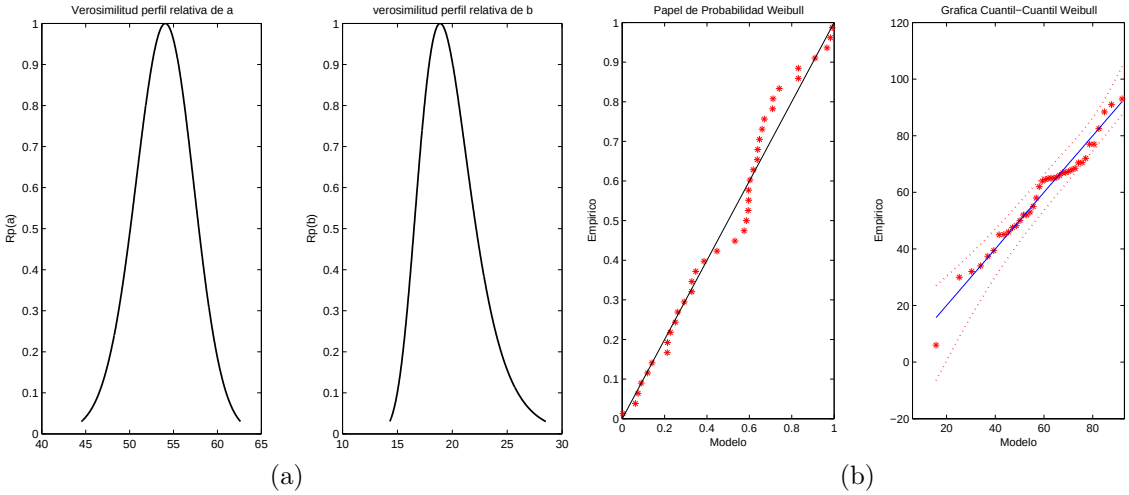


Figura 2.3: (a) Verosimilitudes perfil de a y b del modelo DVEG (b) Gráfica de probabilidad y gráfica Cuantil-Cuantil

Al ajustar la DVEG obtuvimos que el EMV de c fue $\hat{c} = -0.42$; de la Figura 2.2, la gráfica de la verosimilitud perfil relativa de c es ancha y sólo valores negativos de c resultan plausibles (la plausibilidad relativa del modelo Gumbel es 0), sugiriendo únicamente al modelo Weibull. El modelo Weibull correspondiente a $(\hat{a}, \hat{b}, \hat{c})$ según la gráfica QQ (Figura 2.3 (b)) ajusta de manera adecuada a los datos, obteniendo estimaciones que no distan mucho de lo observado.

Modelo	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{Q}_{95}$	$\hat{Q}_{99}$
Weibull	99.49	45.42	2.4	86.3	92.8

Cuadro 2.3: EMV de los parámetros del modelo Weibull

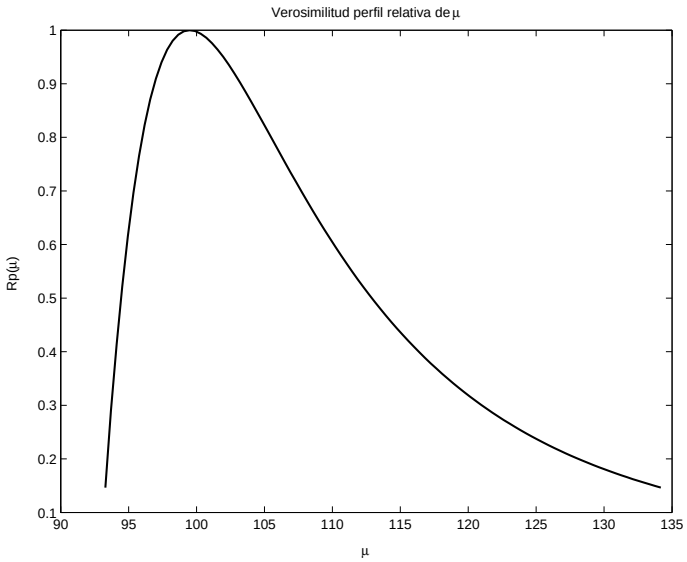
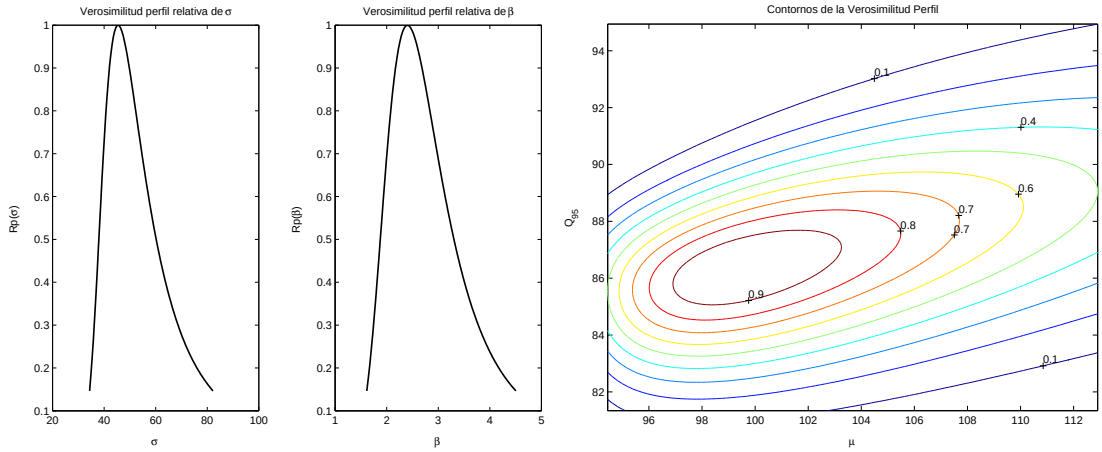


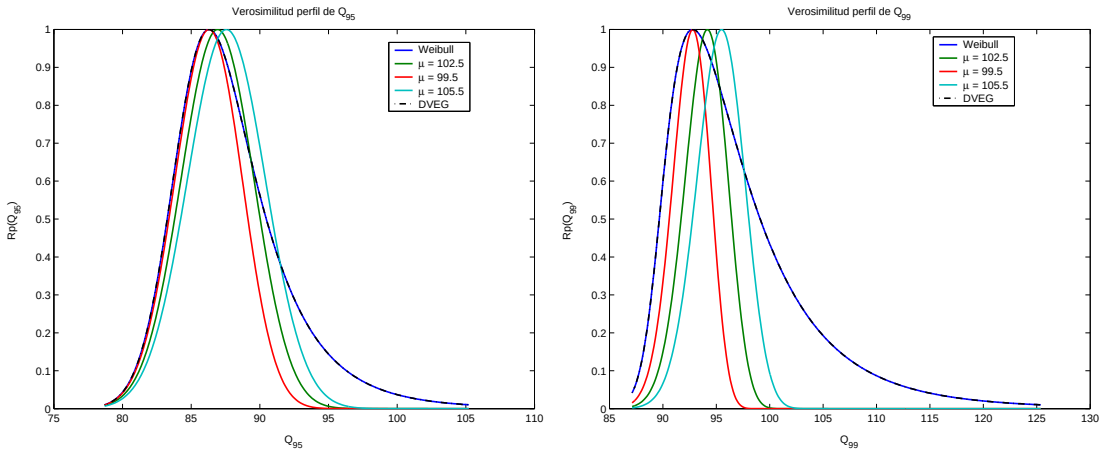
Figura 2.4: Verosimilitud perfil de μ del modelo Weibull



(a) (b)

Figura 2.5: (a) Verosimilitudes perfil de σ y β del modelo Weibull (b) Contornos de verosimilitud perfil de (μ, Q_{95})

La verosimilitud perfil relativa del parámetro μ (Figura 2.4) resulta ser asimétrica respecto a $\hat{\mu}$. En términos de lluvias, la curva de verosimilitud perfil relativa de μ muestra el nivel de plausibilidad del rango de valores tal que ningun registro de lluvia acumulada pueda superar este nivel. El EMV $\hat{\mu} = 99.49$ está cercano a $x_{(n)} = 93$, y considerando niveles de 15 % de verosimilitud encontramos que valores de $\mu = 135$ resultan plausibles. Dado que μ representa el límite superior del soporte de la variable Weibull y nuestro interés se centra en estimar cuantiles grandes, adoptaremos una postura conservadora y consideraremos sólo valores de μ mayores o iguales a $\hat{\mu}$ de plausibilidad no menor a 0.8 para explorar la robusticidad de los cuantiles ante cambios en μ. Por otro lado, las gráficas de verosimilitud perfil relativa para σ y β (Figura 2.5 (a)) son asimétricas a la derecha de $\hat{\sigma}$ y $\hat{\beta}$. Además, en la Figura 2.5 (b) notamos que varía la magnitud de los valores de los extremos del contorno de verosimilitud de nivel 0.1 a medida que μ crece.



(a) Q_{95} (b) Q_{99}

Figura 2.6: Verosimilitudes perfil: (a) Cuantil 95 (b) Cuantil 99

La gráfica de la verosimilitud perfil relativa para Q_{95} (Figura 2.6) sin restricciones sobre μ es asimétrica respecto a $\hat{Q}_{95}$. La asimetría es mayor en la gráfica correspondiente para Q_{99},

abarcando un rango mucho mayor de valores plausibles a la derecha de $\hat{Q}_{99}$ en comparación al rango de valores plausibles a la izquierda de $\hat{Q}_{99}$. La verosimilitud perfil de los cuantiles basados en los valores fijos de μ a la derecha de $\hat{\mu}$, asociados a un nivel mayor de 0.8 de plausibilidad resultan ser, en general, simétricas y más angostas que las verosimilitudes perfiles relativas sin restricción para μ. La variabilidad en las colas en estas curvas es mínima para ambos cuantiles considerados. Para Q_{95} estas curvas están muy cerca de la curva de verosimilitud perfil relativa sin restricciones, la cual contempla la variabilidad en μ, con lo cual podemos decir que para Q_{95} hay robusticidad. En contraste, para Q_{99} la verosimilitud perfil sin restricciones es mucho más ancha que las verosimilitudes perfil bajo los valores propuestos, siendo no robusto. El valor registrado de $X_{(n)} = 93$ mm. tiene un nivel de verosimilitud mayor a 0.2 para Q_{95}, cuantil que ocurre en promedio una vez en 20 años.

Cabe mencionar que la probabilidad máxima acumulada hasta 0 de los modelos Weibull involucrados en el cálculo de la perfil de los cuantiles es de 0.0014, mientras que la misma probabilidad para los modelos Weibull restringidos a $\mu = \hat{\mu}$ es de 0.0065. Es decir, la probabilidad de obtener valores negativos es prácticamente 0.

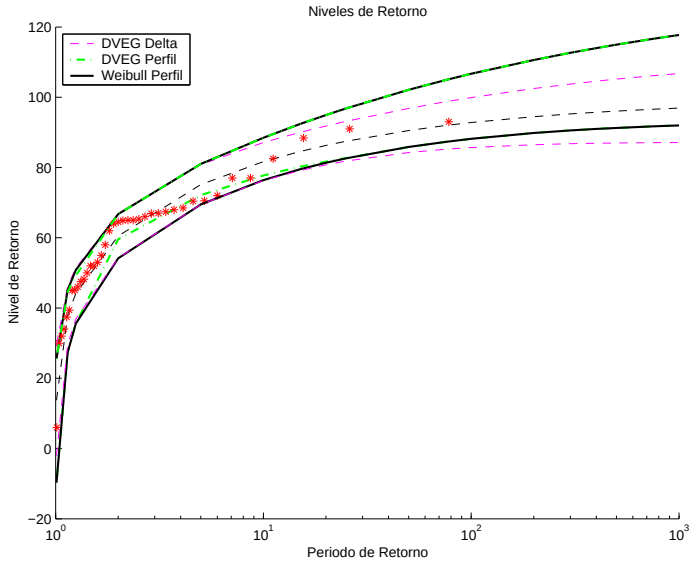


Figura 2.7: Gráfica de niveles de retorno en años

La gráfica de niveles de retorno (Figura 2.7) nos muestra la incertidumbre en la extrapolación de los datos al aumentar la longitud de los intervalos cuando el período de retorno crece. Los intervalos resultantes de la verosimilitud perfil de los cuantiles bajo el modelo Weibull, como ya se había comentado, resultan asimétricos a la derecha del EMV. Estos intervalos captan a los datos. Por otro lado, los intervalos de la verosimilitud de los cuantiles para la DVEG coinciden con los intervalos Weibull comentados para los cuantiles grandes; sin embargo, notamos que hay dos observaciones cercanas al período de retorno de 10 años que no están incluidas en las bandas de verosimilitud para la DVEG. Los intervalos basados en el método delta para los cuantiles de la DVEG también captan a las observaciones. La diferencia con los primeros radica en la asimetría existente en la verosimilitud: las bandas obtenidas con la verosimilitud perfil relativa pueden captar valores más extremos en contraste a las bandas basadas en el método delta para la DVEG.

Basados en que el modelo Weibull sugerido por la verosimilitud perfil relativa de c para representar la distribución de los máximos anuales de acumulación de lluvia en El Temazcal,

Charo tiene un buen ajuste según las gráficas de diagnóstico consideradas, y que las curvas de verosimilitud perfil relativa para los cuantiles proporcionan inferencias adecuadas para los cuantiles, se sugiere el submodelo Weibull como base para las inferencias de interés. Este es un caso para el cual se tiene robusticidad en la inferencia para Q_{95} en contraste a Q_{99} basados en valores de μ de plausibilidad alta.

2.2. Niveles de lluvia en la estación de Bombeo, Michoacán.

El siguiente análisis considera observaciones del nivel de precipitación de lluvia registrado en la estación meteorológica de Bombeo, Michoacán. Dicha estación se sitúa al noreste de la cuenca de la ciudad de Morelia, y cuenta con un registro de 38 años comprendidos entre los años de 1966 y 2003, los cuales presentamos en la Figura 2.8.

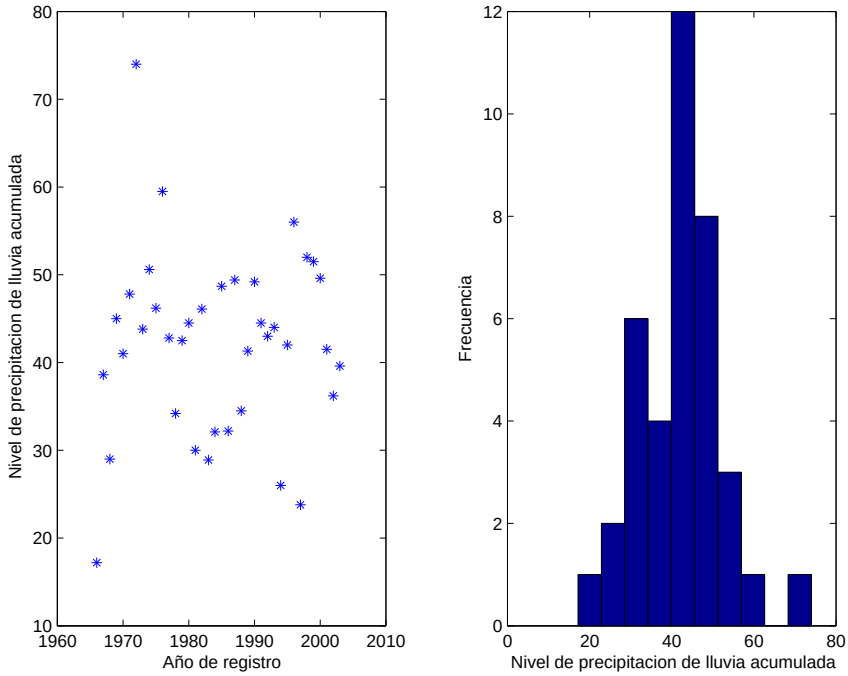


Figura 2.8: Máximos anuales de niveles de lluvia en la estación de Bombeo.

n	$x_{(1)}$	$\tilde{Q}_{25}$	$\tilde{Q}_{50}$	$\bar{X}$	$\tilde{Q}_{75}$	$x_{(n)}$	s
38	17.2	34.5	42.9	42.07	48.7	74	10.59

Cuadro 2.4: Estadísticas descriptivas

Los datos están concentrados alrededor de $\bar{X}$ y $\tilde{Q}_{50}$, estas últimas prácticamente coinciden. Se nota el valor de $x_{(n)} = 74$ que discrepa de los anteriores pero no en magnitud considerable. Esa observación está balanceada por la observación mínima la cual también sobresale del resto por su magnitud respecto a las demás. Lo anterior nos hace esperar una distribución de máximos de colas no pesadas y aproximadamente simétrica.

Modelo	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{Q}_{95}$	$\hat{Q}_{99}$
DVEG	37.99	10.28	-0.2	61.12	69.11

Cuadro 2.5: Estimaciones de máxima verosimilitud

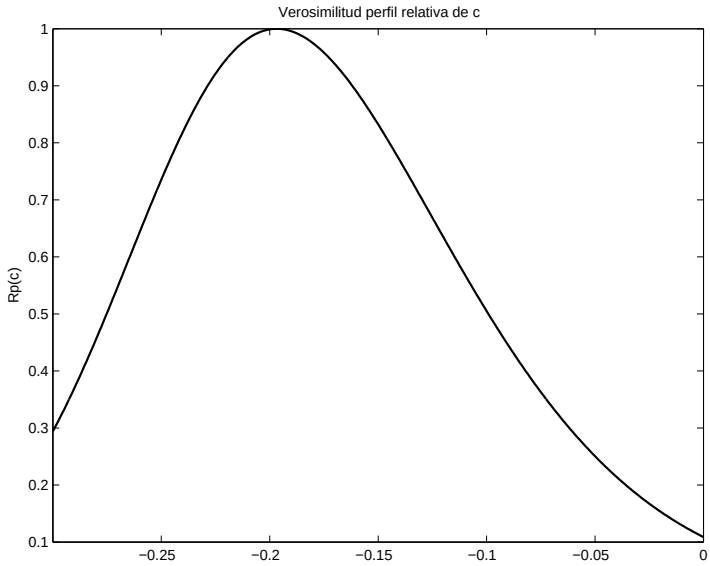
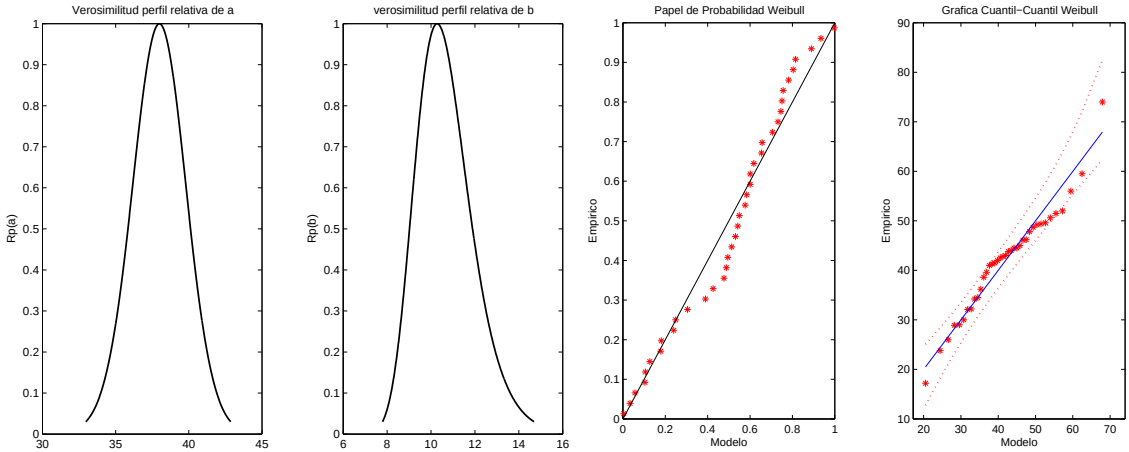


Figura 2.9: Verosimilitud perfil de c del modelo DVEG



(a) (b)

Figura 2.10: (a) Verosimilitudes perfil de a y b del modelo DVEG (b) Gráfica de probabilidad y gráfica Cuantil-Cuantil

La verosimilitud perfil de c (Figura 2.9) se sitúa en el rango negativo de valores de c, y asigna plausibilidades muy bajas a valores de c no negativos. Las inferencias obtenidas para c son $\hat{c} = -0.2$ (no tan lejano a $c = 0$ como en el caso anterior) y el valor $c = 0$ tiene plausibilidad de 0.12, dando poco sustento para el modelo Gumbel y sugiriendo únicamente a la familia Weibull. El ajuste del modelo Weibull correspondiente a $(\hat{a}, \hat{b}, \hat{c})$ según la gráfica QQ (Figura 2.10 (b)) es bueno respecto a los valores centrales, mientras que para las observaciones mínima (y máxima) el modelo sobreestima (subestima) lo observado ya que la estimación del modelo queda por debajo (encima) de la línea identidad. Considerando el ajuste del modelo

Modelo	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{Q}_{95}$	$\hat{Q}_{99}$
Weibull	90.24	52.25	5.08	61.12	69.11

Cuadro 2.6: EMV de los parámetros del modelo Weibull

con las bandas de verosimilitud (ver Figura 2.14), notamos que las bandas captan a $x_{(1)}$ y $x_{(n)}$.

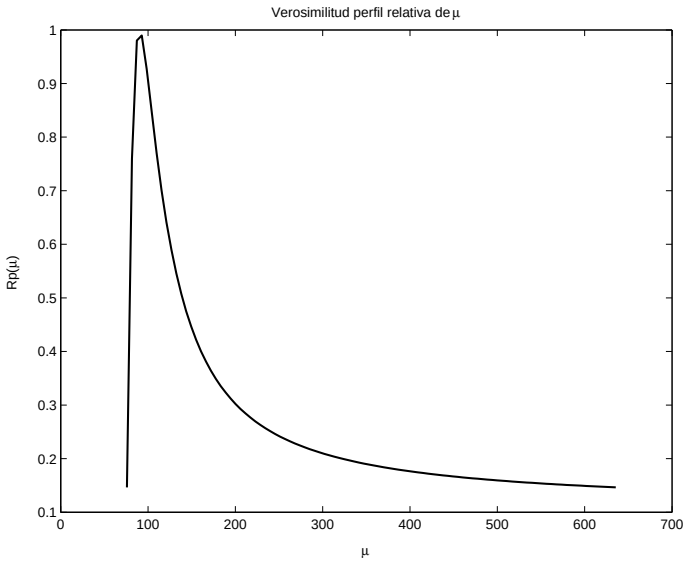
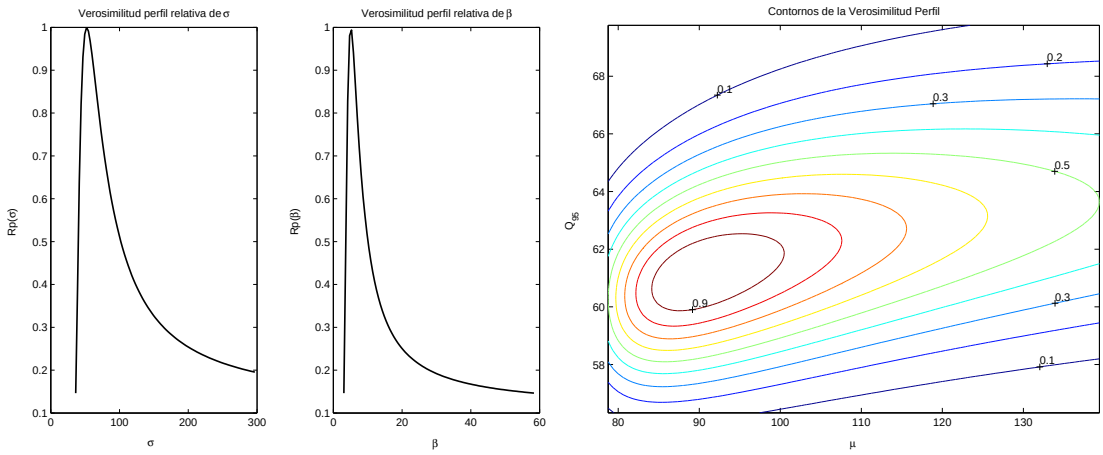


Figura 2.11: Verosimilitud perfil de μ del modelo Weibull

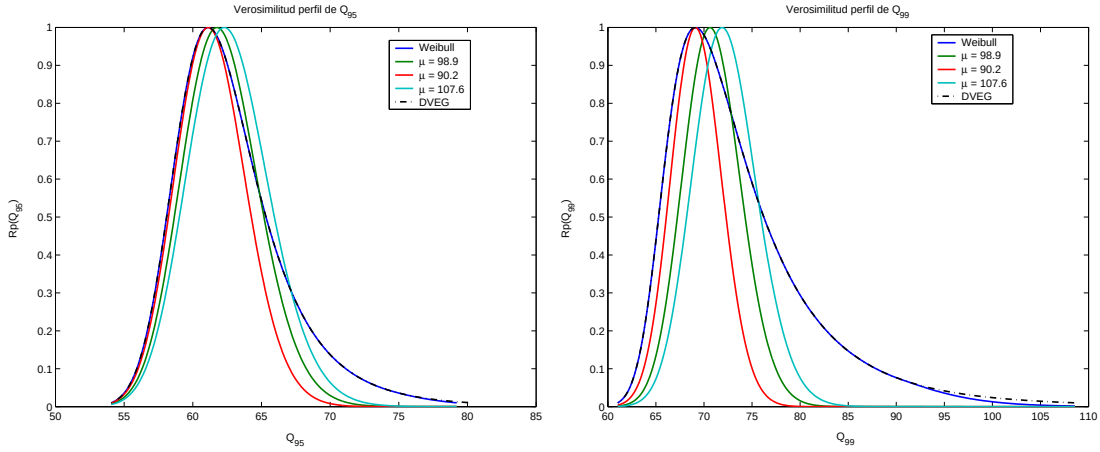


(a) (b)

Figura 2.12: (a) Verosimilitudes perfil de σ y β del modelo Weibull (b) Contornos de verosimilitud perfil de (μ, Q_{95})

La verosimilitud perfil para μ (Figura 2.11) resulta ser muy asimétrica en su cola derecha respecto a $\hat{\mu}$. El EMV es $\hat{\mu} = 90.24$, y la curva es muy angosta para valores cercanos a $\hat{\mu}$ mientras que valores de magnitud mucho mayor al máximo observado (niveles de lluvia de 600 mm.) tienen asignada una plausibilidad de 0.15. Por otro lado, en los contornos de verosimilitud para (μ, Q_{95}) vemos que las variaciones en los extremos del contorno de 10 % de verosimilitud

son mínimas al considerar diferentes valores fijos de μ de entre los graficados, los cuales están asociados a una plausibilidad mayor a 0.5.



(a) (b)

Figura 2.13: Verosimilitudes perfil: (a) Cuantil 95 (b) Cuantil 99

La curva de verosimilitud perfil relativa para Q_{95} (Figura 2.13) basada en el modelo Weibull sin restricciones para μ resulta ser ligeramente asimétrica a la derecha de $\hat{Q}_{95}$. Para Q_{99} la asimetría de la verosimilitud perfil relativa correspondiente se hace más evidente. El nivel observado de $x_{(n)} = 74$ tiene plausibilidad menor de 0.1 para Q_{95}; en contraste, dicho registro tiene plausibilidad mayor a 0.7 para Q_{99}, asociado a que ocurran una vez en 100 años.

Las curvas de verosimilitud perfil relativa para los cuantiles basados en los valores de μ mostrados en la Figura (2.13) (de plausibilidad mayor a 0.8), son de forma similar, difiriendo en 3 mm. de lluvia en la cola derecha de las curvas para Q_{95}, por lo que para este ejemplo se tiene robusticidad. Sin embargo, para Q_{99} no se tiene robusticidad de los cuantiles frente a cambios en μ, y la verosimilitud perfil sin restricciones toma en cuenta este hecho. Finalmente, la verosimilitud perfil relativa basada en el modelo DVEG coincide completamente con la curva basada en el modelo Weibull para ambos cuantiles.

Cabe mencionar que la probabilidad máxima acumulada hasta 0 de los modelos Weibull involucrados en el cálculo de la perfil de los cuantiles es del ordel de las millonésimas, mientras que la misma probabilidad para los modelos Weibull restringidos a $\mu = \hat{\mu}$ es del orden de las cienmilésimas. Valores negativos de niveles de lluvia son prácticamente improbables.

De la Figura 2.14, notamos que la asimetría de los intervalos de verosimilitud para los cuantiles hace que se cubra a la observación máxima; los intervalos basados en el método delta también cubren las observaciones mayores. Ambos intervalos sobreestiman ligeramente una observación correspondiente a un período de retorno cercano a 10 años. Notamos que en extrapolaciones grandes la apertura de los intervalos aumenta considerablemente.

En conclusión, el modelo Weibull representa adecuadamente el comportamiento del máximo anual de niveles de acumulación de lluvias en la estación de Bombeo. En este ejemplo se tiene robusticidad en la inferencia para Q_{95} frente a cambios en el parámetro μ, correspondiendo a niveles de verosimilitud de 0.8. Al igual que el ejemplo anterior, se recomienda el uso de la distribución Weibull de tres parámetros.

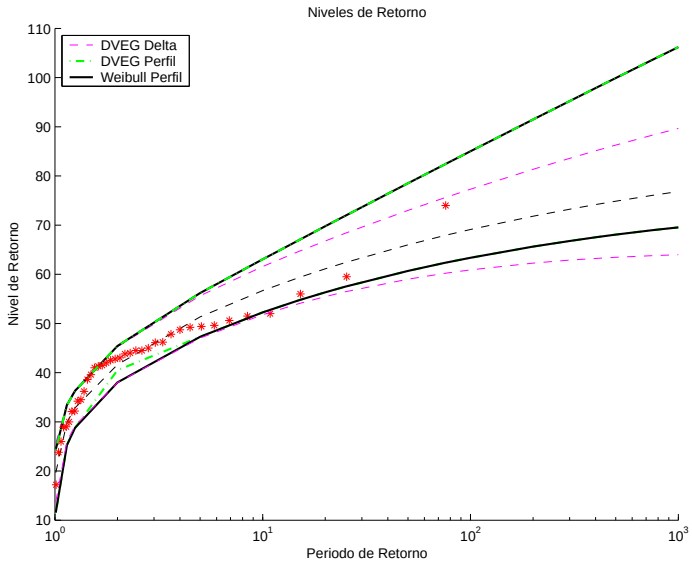


Figura 2.14: Gráfica de niveles de retorno en años

2.3. Máximos anuales de niveles de mar en Port Pirie, Australia

El siguiente conjunto de datos, presentado y analizado por Stuart Coles [6] , corresponde a mediciones de máximos anuales de niveles del mar en Port Pirie, Australia. Dichas mediciones abarcan el periodo de 1923 a 1987. El objetivo del estudio es, basado en estos datos, modelar adecuadamente los máximos niveles de mar que podrían ocurrir en Port Pirie.

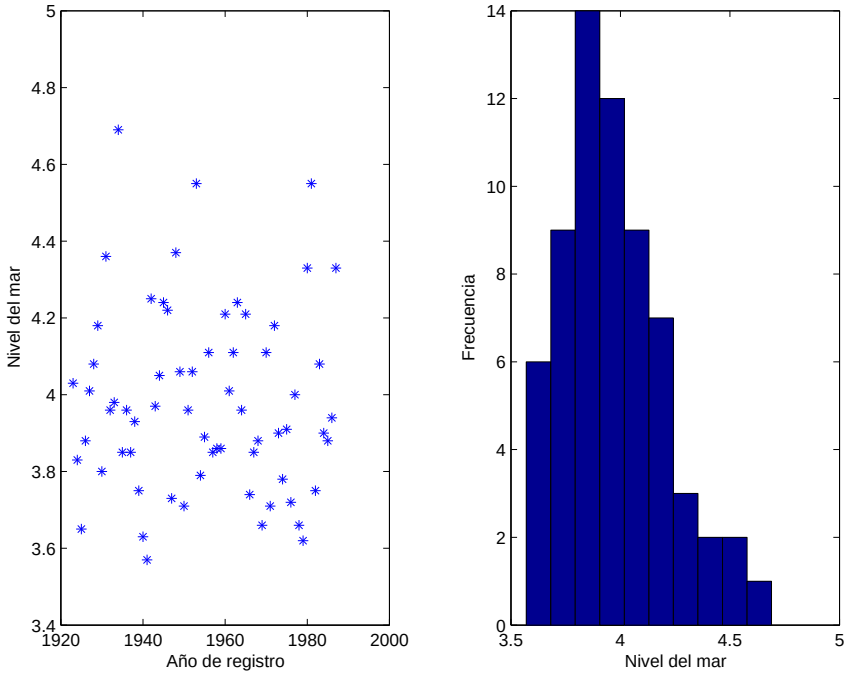


Figura 2.15: Máximos anuales de niveles de mar en Port Pirie, Australia.

n	$x_{(1)}$	$\tilde{Q}_{25}$	$\tilde{Q}_{50}$	$\bar{X}$	$\tilde{Q}_{75}$	$x_{(n)}$	s
65	3.57	3.82	3.96	3.98	4.11	4.69	0.24

Cuadro 2.7: Estadísticas descriptivas

De la Figura 2.15, Stuart Coles comenta que no se percibe un patrón que rija el comportamiento de los datos en el periodo observado, por lo cual es razonable modelar las mediciones como observaciones independientes de la DVEG. Al respecto, el autor recalca la importancia de este supuesto en la extrapolación, ya que cualquier inferencia obtenida del análisis estará sujeta a la validez de dichas condiciones.

Modelo	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{Q}_{95}$	$\hat{Q}_{99}$
DVEG	3.87	19.8	-0.05	4.42	4.69
Modelo	$\hat{\mu}$	$\hat{\sigma}$		$\hat{Q}_{95}$	$\hat{Q}_{99}$
Gumbel	3.87	19.49		4.45	4.77

Cuadro 2.8: Estimaciones de máxima verosimilitud

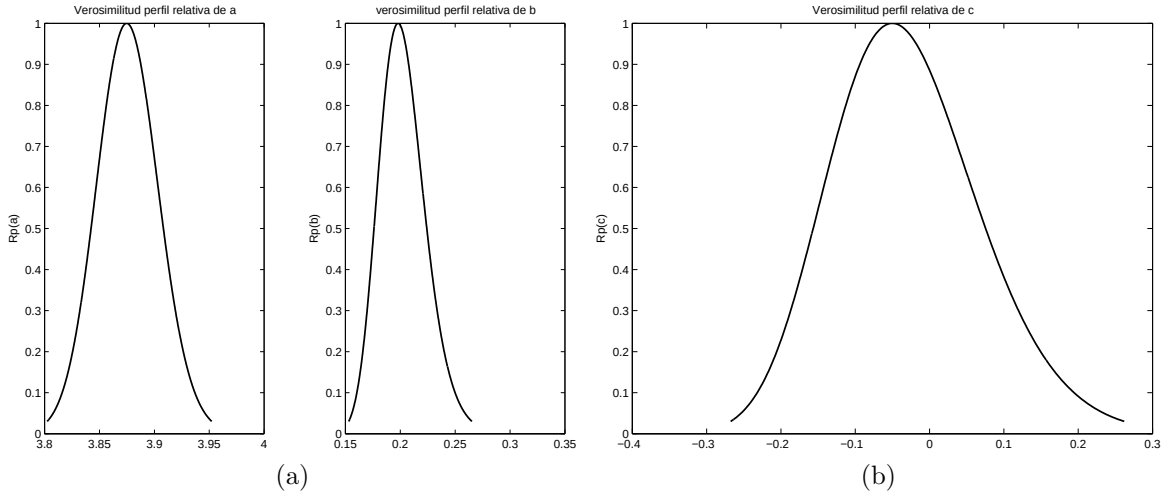
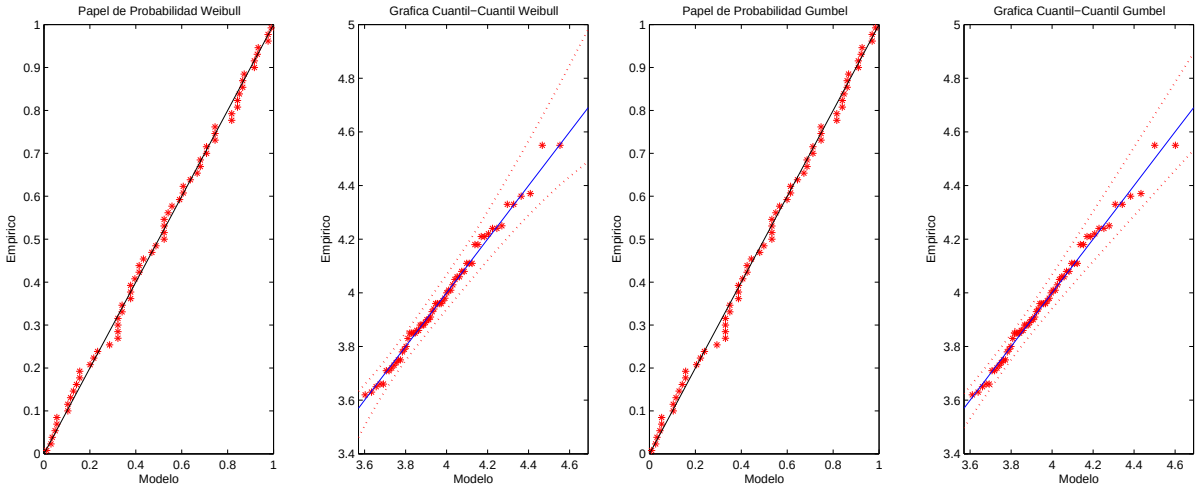


Figura 2.16: Verosimilitudes perfil relativa del modelo DVEG: (a) a y b (b) c

La verosimilitud perfil relativa de c (Figura 2.16) es aproximadamente simétrica, teniendo como EMV a $\hat{c}$=-0.05 y asignando plausibilidad de 0.88 al valor $c = 0$, sugiriendo que se consideren los submodelos Weibull y Gumbel. De las gráficas de diagnóstico (Figura 2.17) vemos que el ajuste para ambos modelos es bueno ya que los puntos graficados están cerca de la línea identidad; en la gráfica QQ, la estimación a través del modelo Gumbel es más cercana a los puntos de mayor magnitud en comparación al modelo Weibull.

Modelo	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{Q}_{95}$	$\hat{Q}_{99}$
Weibull	7.83	3.95	19.95	4.42	4.69

Cuadro 2.9: EMV de los parámetros del modelo Weibull



(a) DVEG (b) Gumbel

Figura 2.17: Gráfica de probabilidad y gráfica Cuantil-Cuantil

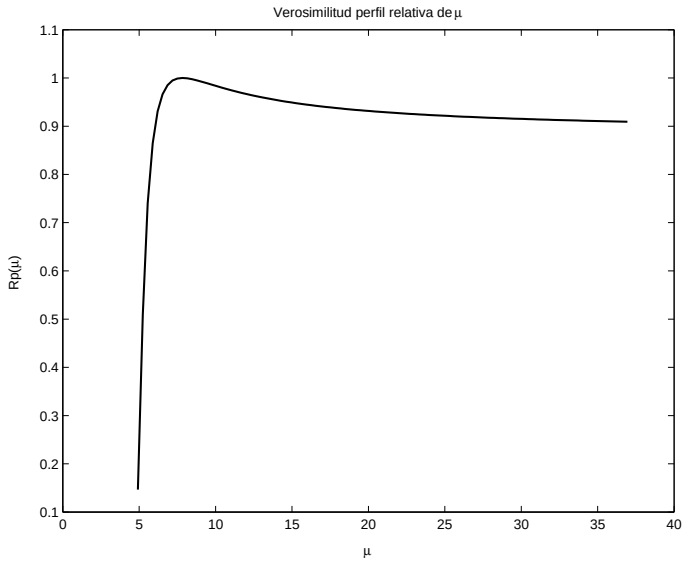
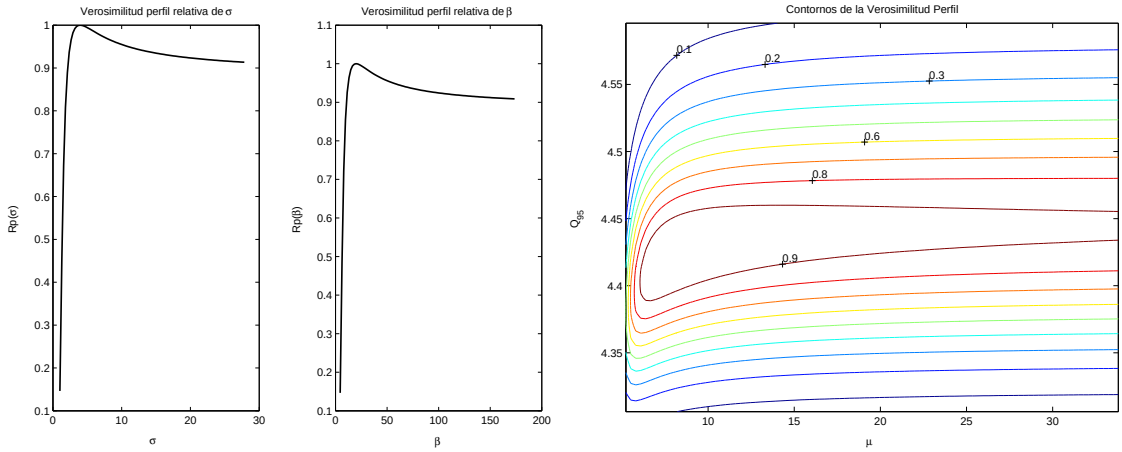


Figura 2.18: Verosimilitud perfil de μ del modelo Weibull

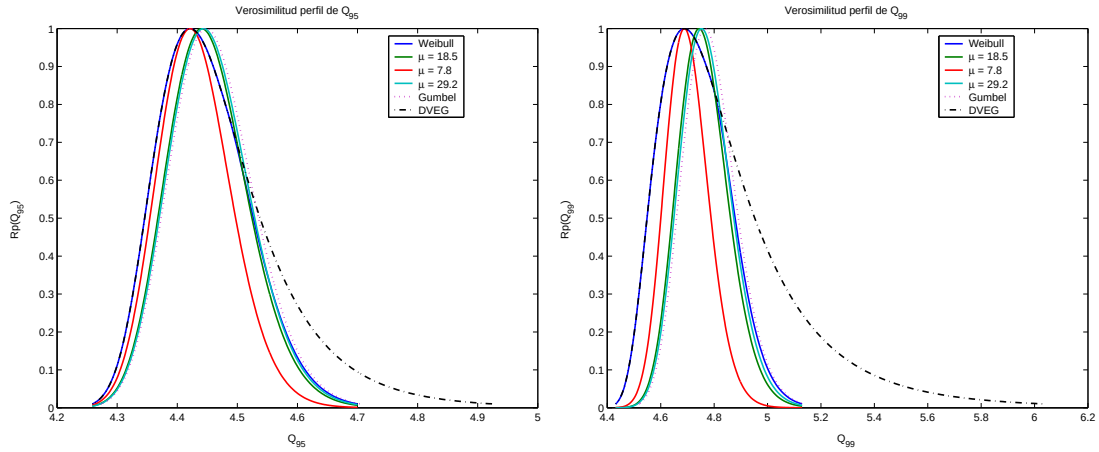
El valor de $\hat{\mu}$ indica que el mínimo valor del nivel de mar que no se superará jamás es de 7.83 mts. Sin embargo, la forma aplanada de la verosimilitud perfil relativa de μ (Figura 2.18) muestra que el rango de valores plausibles para μ se extiende considerablemente a la derecha de $\hat{\mu}$; más aún, la plausibilidad de los mismos se mantiene muy elevada, con lo que prácticamente cualquier valor a la derecha de $\hat{\mu}$ es fuertemente sustentado por los datos. En términos del nivel de mar, la forma aplanada de la verosimilitud perfil relativa para μ indica que cualquier valor mayor a $\hat{\mu}$ puede fungir como el mínimo valor tal que el nivel del mar no lo superará jamás. En términos de modelación, plausibilidades altas para valores grandes de μ apuntan a que el modelo Gumbel resulte plausible. Por otro lado, las gráficas de verosimilitud perfil relativa para σ y β resultan aplanadas, aludiendo también a la modelación Gumbel. En la gráfica de contornos correspondiente a (μ , Q_{95}) (Figura 2.19 (b)) notamos muy poca variabilidad tanto en anchura como en los valores de los extremos de los intervalos para Q_{95} al considerar diferentes valores



(a) (b)

Figura 2.19: (a) Verosimilitudes perfil de σ y β del modelo Weibull (b) Contornos de verosimilitud perfil de (μ, Q_{95})

de μ.



(a) (b)

Figura 2.20: Verosimilitudes perfil: (a) Cuantil 95 (b) Cuantil 99

Las verosimilitudes perfil de Q_{95} y Q_{99} (Figura 2.20) muestran un gran traslape de las curvas para los modelos Gumbel y Weibull, haciendo que las inferencias sean muy similares basados en cualquiera de éstos. Considerando los valores de μ descritos en la Figura 2.20 se tienen curvas de verosimilitud para Q_{95} las cuales son semejantes en forma y se traslapan mucho incluso con la verosimilitud perfil Weibull, por lo que se tiene una situación de robusticidad. Variaciones de una décima de unidad y menos traslape se nota en las mismas curvas para Q_{99}. Para Q_{95}, la función de verosimilitud perfil asigna plausibilidad muy baja a valores como el registro de 4.69 mts, mientras que el mismo valor resulta altamente plausible para Q_{99}, asociado a un período de retorno de 100 años. Las gráficas de verosimilitud perfil relativa para los cuantiles basados en la DVEG son asimétricas y de mayor amplitud en la cola derecha en comparación a las curvas restantes.

Para los modelos Gumbel asociados en el cálculo de la verosimilitud perfil relativa de los cuantiles, encontramos que la probabilidad de obtener valores negativos es 0.

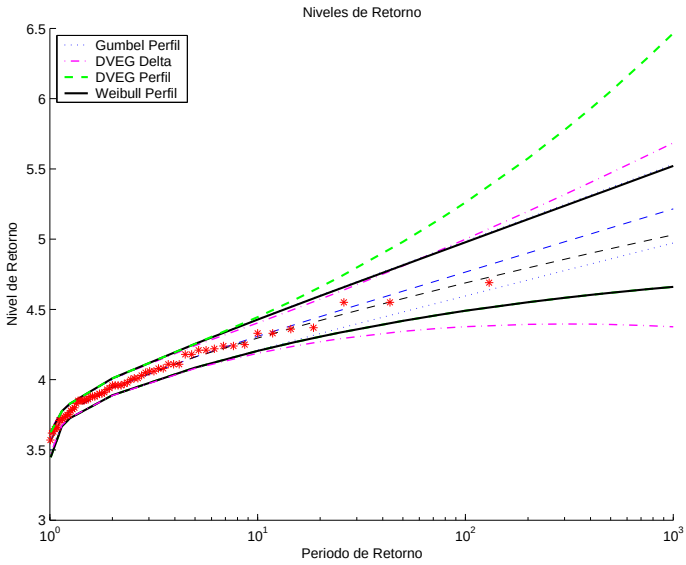


Figura 2.21: Gráfica de niveles de retorno en años

De los niveles de retorno (Figura 2.21) notamos una mayor amplitud de los intervalos de verosimilitud para la DVEG, en relación a los otros modelos. El extremo superior de los intervalos de la DVEG basado en el método delta, y de los modelos Weibull y Gumbel basados en la perfil, resultan muy cercanos. Todos los modelos bajo los diferentes intervalos considerados captan a las observaciones. El modelo Gumbel proporciona los intervalos más angostos.

Retomando, los modelos Gumbel y Weibull sugeridos a través de la verosimilitud perfil de c ajustan adecuadamente según las gráficas de diagnóstico. El modelo Weibull proporciona inferencias prácticamente iguales al modelo Gumbel, y este último resulta un modelo con menos parámetros y más conservador en cuanto a la estimación de cuantiles grandes. Inclusive, las inferencias de los parámetros del modelo Weibull hacen alusión al modelo Gumbel. Con base en el análisis estadístico expuesto, se sugiere el modelo Gumbel para representar la distribución del máximo de nivel de mar para Port Pirie.

En contraste, basado en la gráfica de niveles de retorno y las bandas obtenidas a través del método delta para la DVEG, a pesar que el modelo Gumbel sea justificado por los datos y que las inferencias obtenidas bajo el mismo sean mejores en términos de precisión, Coles se inclina por modelar con la DVEG debido a la incertidumbre en la elección del modelo y a la mayor amplitud de los intervalos obtenidos por el método delta para captar valores más extremos. El anterior argumento no contempla la importancia de la elección de un submodelo de la DVEG, ya que es sólo uno el que está asociado con la distribución del máximo.

2.4. Niveles de lluvia en Morelia, Michoacán.

En esta sección aplicaremos la propuesta inferencial hecha en esta tesis para modelar un conjunto de datos de precipitaciones de lluvia en la estación de Morelia, Michoacán. La estación meteorológica de Morelia se encuentra ubicada al noreste de la cuenca de la ciudad de Morelia, y al igual que la estación Presa Cointzio, resulta una de las más representativas del fenómeno de lluvias en toda la cuenca, según un estudio hidrológico realizado en Michoacán por el IMTA (ver [10]). El período de registro acumulado en esta estación es de 40 años entre 1947 y 1986.

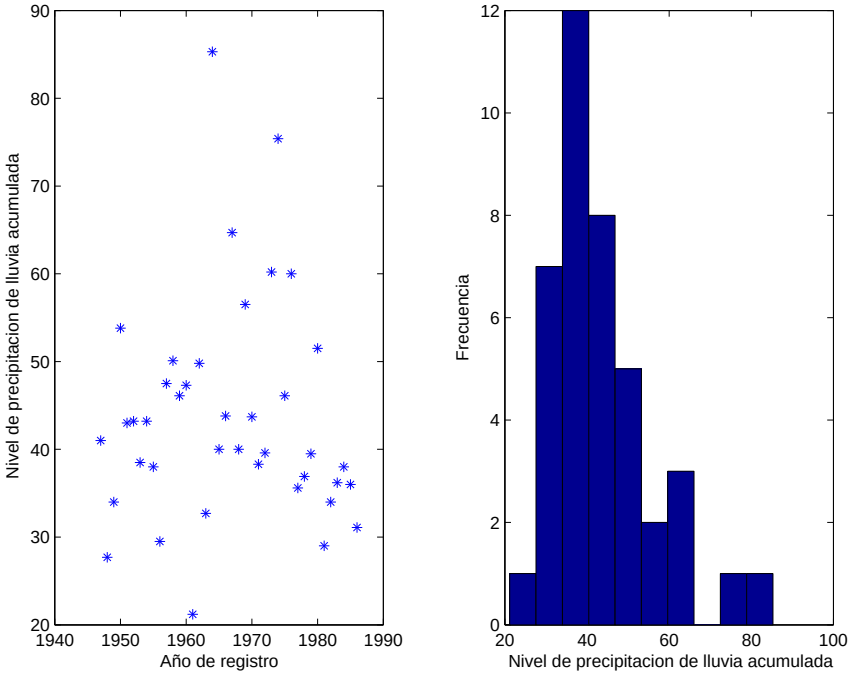


Figura 2.22: Máximos anuales de niveles de lluvia en Morelia.

n	$x_{(1)}$	$\tilde{Q}_{25}$	$\tilde{Q}_{50}$	$\bar{X}$	$\tilde{Q}_{75}$	$x_{(n)}$	s
40	21.2	36.1	40.5	43.7	48.65	85.3	12.63

Cuadro 2.10: Estadísticas descriptivas

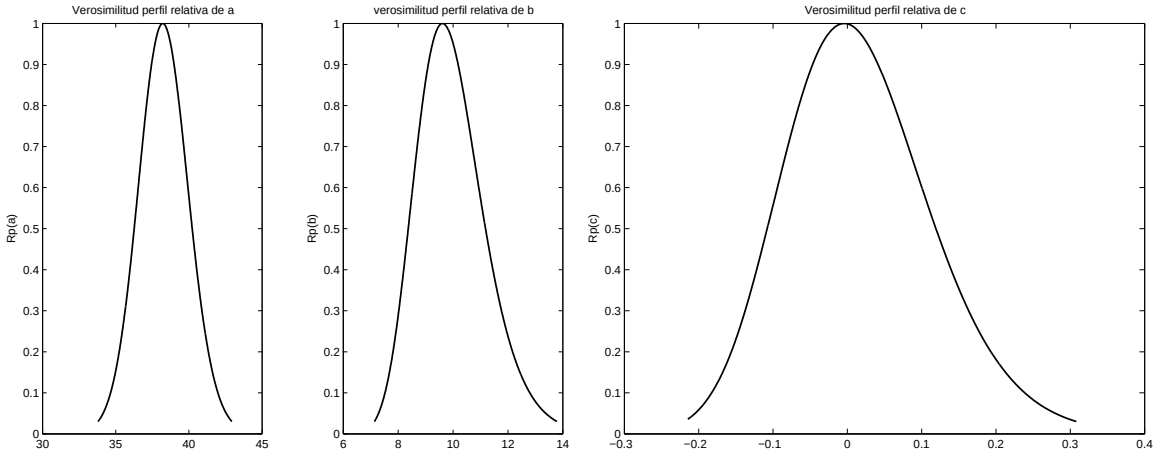
La Figura 2.22 muestra los registros de máximos anuales de acumulación de lluvia en Morelia. De ella vemos que la mayoría de los datos se concentra alrededor de sus medidas centrales, con una ligera asimetría. Sobresale el valor $x_{(n)} = 85.3$ del resto de los datos.

Modelo	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{Q}_{95}$	$\hat{Q}_{99}$
DVEG	38.21	9.62	-0.0004	66.6	82.04
Modelo	$\hat{\mu}$	$\hat{\sigma}$		$\hat{Q}_{95}$	$\hat{Q}_{99}$
Gumbel	38.19	9.61		66.73	82.4

Cuadro 2.11: Estimaciones de máxima verosimilitud

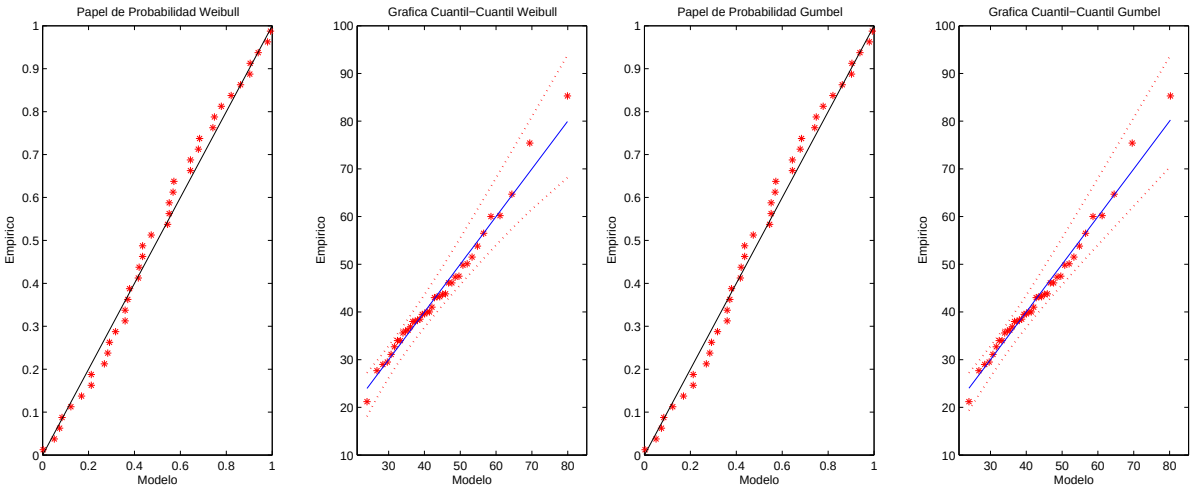
El estimador puntual de c en la DVEG es prácticamente 0, lo cual sugiere fuertemente al modelo Gumbel. Consideraremos el ajuste de la familia Weibull asociada al valor de $\hat{c}$ para fines de comparación de modelos. De la Figura 2.24 notamos que el ajuste de los modelos Gumbel y Weibull resultante de $(\hat{a}, \hat{b}, \hat{c})$ es prácticamente indistinguible. En la gráfica QQ los puntos quedan cerca de la línea identidad salvo los registros más grandes, los cuales no distan mucho de la línea identidad.

El EMV del parámetro μ es $\hat{\mu} = 2361.58$, el cual resulta extremadamente grande en términos de las unidades del fenómeno. La verosimilitud perfil relativa de μ (Figura 2.25) es completamente plana, haciendo plausibles a valores muy grandes para el parámetro. Nuevamente, en



(a) (b)

Figura 2.23: Verosimilitudes perfil relativa del modelo DVEG: (a) c (b) a y b



(a) DVEG (b) Gumbel

Figura 2.24: Gráfica de probabilidad y gráfica Cuantil-Cuantil

términos de lluvias la inferencia respecto a μ no es informativa; en contraste, en términos de modelación, esto nos sugiere considerar el modelo Gumbel. Por otro lado, el comportamiento de la verosimilitud perfil relativa para σ y β (Figura 2.26 (a)) es similar al de μ. Además, en los contornos de verosimilitud perfil para (μ, Q_{95}) (Figura 2.26 (b)) podemos apreciar que variaciones respecto a μ conducen a los mismos intervalos para Q_{95}, siendo muy robusto este caso.

Las curvas de verosimilitud perfil relativa de Q_{95} y Q_{99} (Figura 2.27) resultan muy robustas ante los distintos valores de μ mostrados en la gráfica, siendo indistinguibles en el lado derecho al EMV para todas las familias consideradas a excepción del modelo DVEG. Este último tiene mayor amplitud en la cola derecha, asignando plausibilidad mayor a cero a un rango más amplio de valores en comparación a los modelos Weibull y Gumbel. Basados en el modelo Gumbel, un registro como el de $x_{(n)} = 85.3$ no resulta plausible respecto a la verosimilitud perfil de Q_{95}, sin embargo resulta altamente plausible para Q_{99} asociado a un periodo de retorno de 100 años.

Cabe resaltar que la probabilidad máxima de obtener valores negativos de entre todos los

Modelo	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{Q_{95}}$	$\hat{Q_{99}}$
Weibull	2361.58	2323.37	241.58	66.6	82.04

Cuadro 2.12: EMV de los parámetros del modelo Weibull

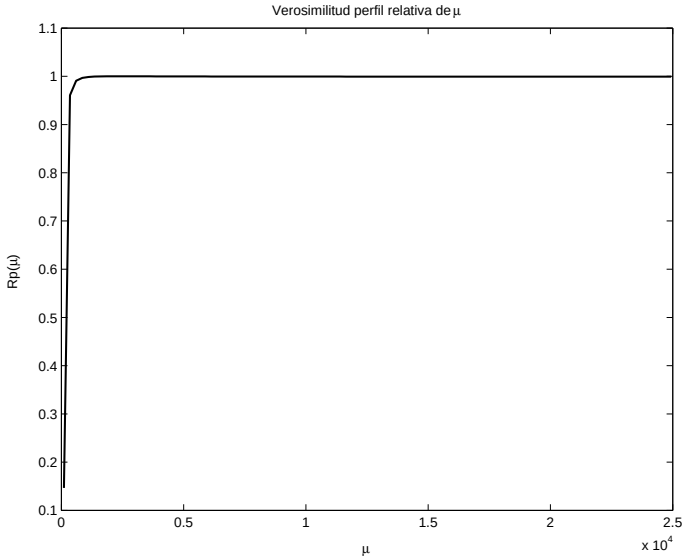
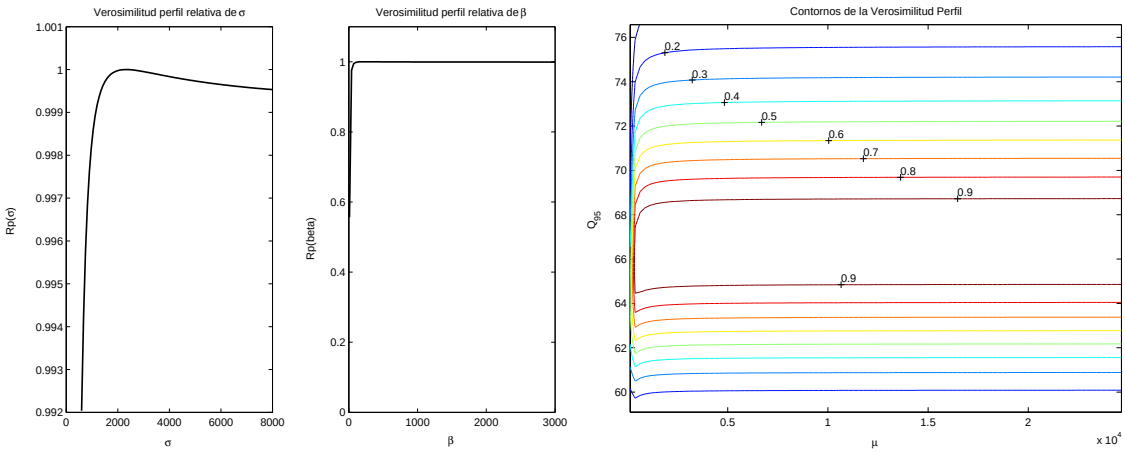


Figura 2.25: Verosimilitud perfil de μ del modelo Weibull



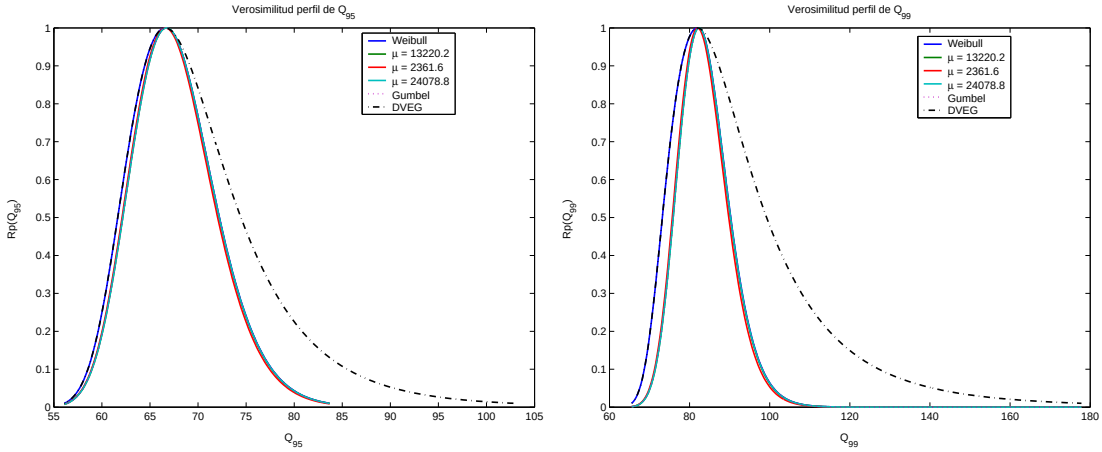
(a) (b)

Figura 2.26: (a) Verosimilitudes perfil de σ y β del modelo Weibull (b) Contornos de verosimilitud verfil de (μ, Q_{95})

modelos Gumbel considerados en el calculo de la verosimilitud perfil relativa de los cuantiles es de 0.006.

Prácticamente es indistinguible el ajuste del modelo Gumbel y el Weibull según la Figura 2.28. Incluso para valores muy grandes de extrapolación en los cuantiles, como lo es Q_{999} asociado a un periodo de retorno de 1000 años, las inferencias son las mismas, con lo que el modelo Weibull y el Gumbel son prácticamente iguales. Ambos modelos captan todas las observaciones.

En especial cabe resaltar la ganancia respecto a la anchura de los intervalos que se obtiene



(a) (b)

Figura 2.27: Verosimilitudes perfil: (a) Cuantil 95 (b) Cuantil 99

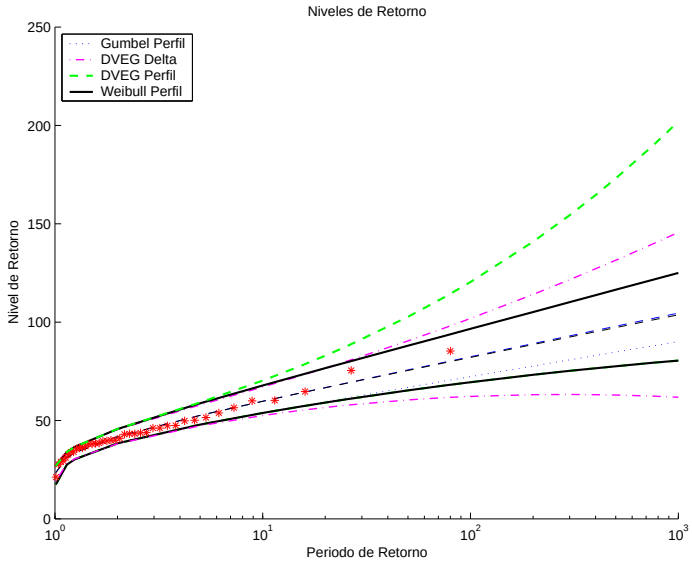


Figura 2.28: Gráfica de niveles de retorno en años

al elegir una subclase de modelos en vez de considerar la DVEG en este caso.

Al igual que Port Pirie, este es un caso donde la magnitud gigantesca de los parámetros del modelo estimado Weibull y sus perfiles son indicio que el modelo Gumbel está muy cercano a éste. Además, el estimador puntual de c es prácticamente 0 y se tiene un ajuste bueno del modelo Gumbel respecto a lo observado. De todo lo anterior, desde el punto de vista estadístico se sugiere el uso del modelo Gumbel para representar al máximo de acumulación de lluvias en la estación meteorológica de Morelia.

2.5. Niveles de lluvia en Maiquetía, Venezuela

El presente análisis considera los máximos anuales de registros de niveles diarios de lluvia obtenidos en el Aeropuerto Internacional de Maiquetía, Venezuela. Los datos representan el máximo anual del nivel de lluvia diaria acumulada obtenidos desde 1951 a 1999 (Figura 2.29).

El registro del año de 1999 es de especial interés, ya que debido a su magnitud extrema (410 milímetros, más del doble de cualquier registro anterior), trajo consigo consecuencias devastadoras en la zona.

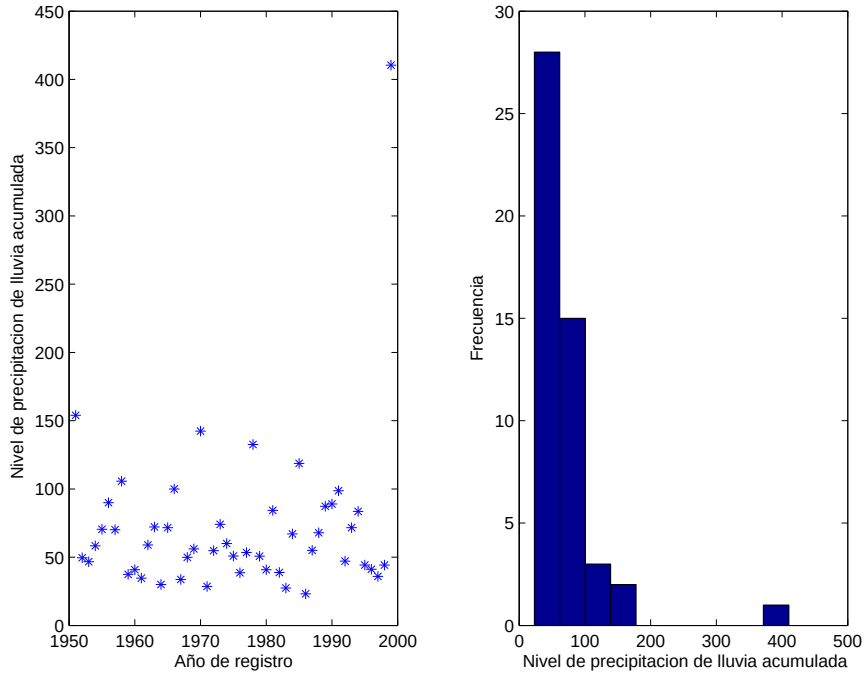


Figura 2.29: Máximos anuales de niveles de lluvia en Maiquetía.

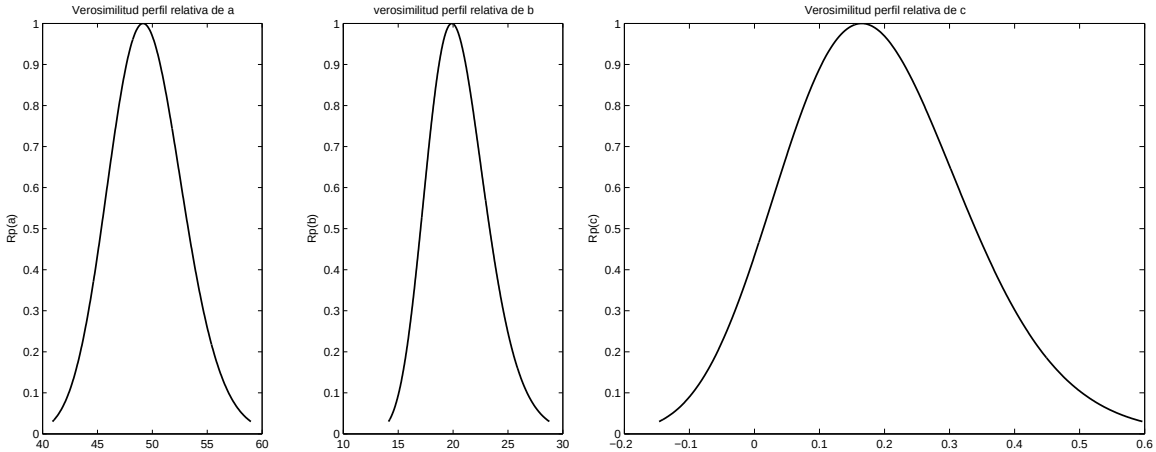
n	$x_{(1)}$	$\tilde{Q}_{25}$	$\tilde{Q}_{50}$	$\bar{X}$	$\tilde{Q}_{75}$	$x_{(n)}$	s
48	23.2	41	55.55	64.21	78.85	154	30.34

Cuadro 2.13: Estadísticas descriptivas

Coles [9] realizó un análisis del fenómeno en cuestión, enfatizando la importancia de un estudio cauteloso basado en Teoría de Extremos para obtener inferencias confiables respecto a valores como el registrado. Por lo anterior, el objetivo del autor es considerar diferentes modelos de teoría de extremos para proveer inferencias adecuadas con relación al valor de 410 mm. basados en los registros anteriores. Aquí se explorará las conclusiones a las que se puede llegar al modelar, de acuerdo a la propuesta de la tesis, el conjunto de datos en cuestión sin la observación de 1999, con la finalidad de valorar si el modelo elegido hubiera proporcionado información relevante de este valor.

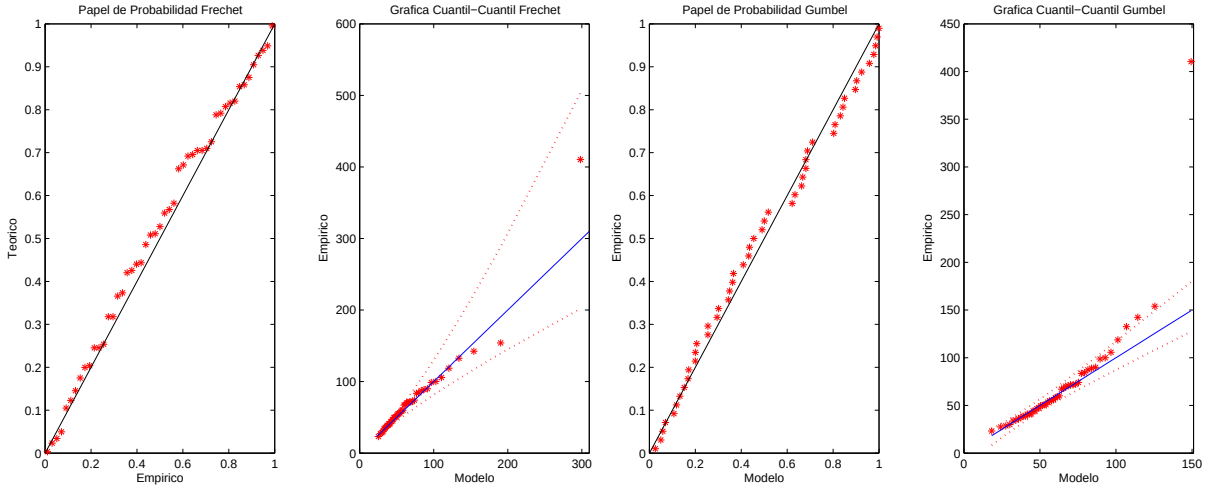
Modelo	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{Q}_{95}$	$\hat{Q}_{99}$
DVEG	49.16	19.89	0.16	125.39	186.1
Modelo	$\hat{\mu}$	$\hat{\sigma}$		$\hat{Q}_{95}$	$\hat{Q}_{99}$
Gumbel	50.99	21.49		114.82	149.84

Cuadro 2.14: Estimaciones de máxima verosimilitud



(a) (b)

Figura 2.30: Verosimilitudes perfil relativa del modelo DVEG: (a) a y b (b) c



(a) Fréchet restringido (b) Gumbel

Figura 2.31: Gráfica de probabilidad y gráfica Cuantil-Cuantil

De la tabla (2.13) notamos que la mayoría de los datos se concentra a la izquierda de $\bar{X}$, esta última afectada por $x_{(n)}$ y valores cercanos que reflejan una cola pesada. El ajuste de la DVEG (Figura 2.30) nos da un EMV de c positivo, sugiriendo a la familia Fréchet como posible modelo para representar a los datos. Por la naturaleza positiva de los datos, ajustaremos la familia Fréchet restringiendo el espacio de μ a valores no negativos, obteniendo con esto que el modelo no asigne probabilidades positivas a valores que son imposibles de observar, además de considerar a una familia Fréchet de dos parámetros como posible modelo más simple para representar a las observaciones. Por otro lado, de la verosimilitud perfil relativa de c, obtenemos que $R_p(c = 0) = 0.43$, sugiriendo también considerar al modelo Gumbel para la muestra en cuestión.

La gráfica QQ para el modelo Gumbel (Figura 2.31) muestra un ajuste inadecuado para las cuatro observaciones más grandes. El modelo Fréchet con $\mu \geq 0$ en la gráfica QQ resulta muy cercano a los datos en la parte central, y para $X = 410$ mm. el valor puntual estimado resulta mucho más cercano que el estimado a través del modelo Gumbel. La gráfica de verosimilitud

Modelo	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{Q_{95}}$	$\hat{Q_{99}}$
Fréchet	0	46.81	2.47	155.55	300.66

Cuadro 2.15: EMV del modelo Fréchet con $\mu \geq 0$

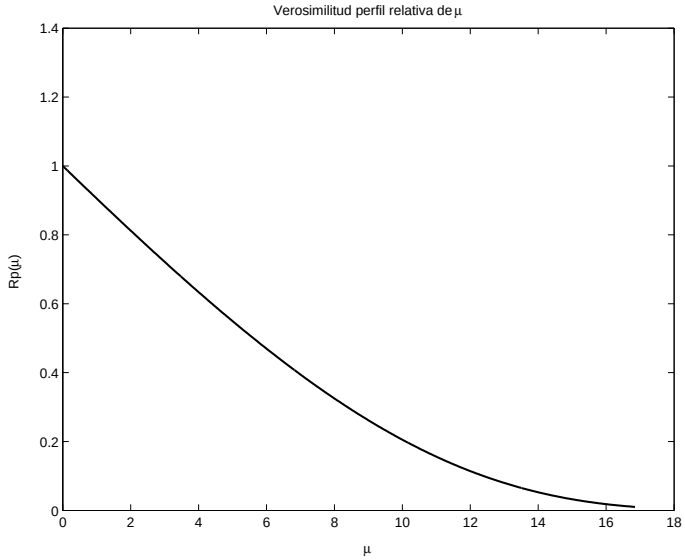
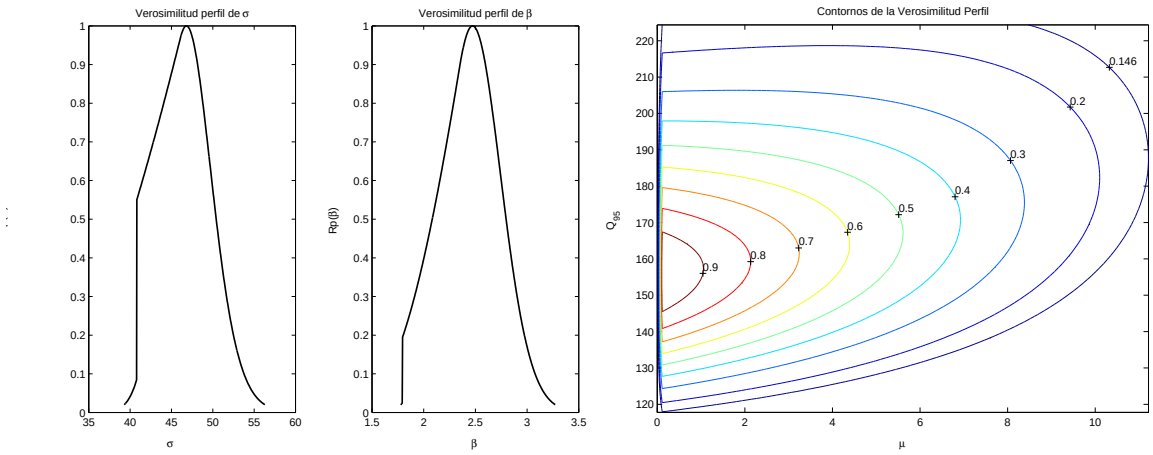


Figura 2.32: Verosimilitud perfil de μ del modelo Fréchet con $\mu \geq 0$

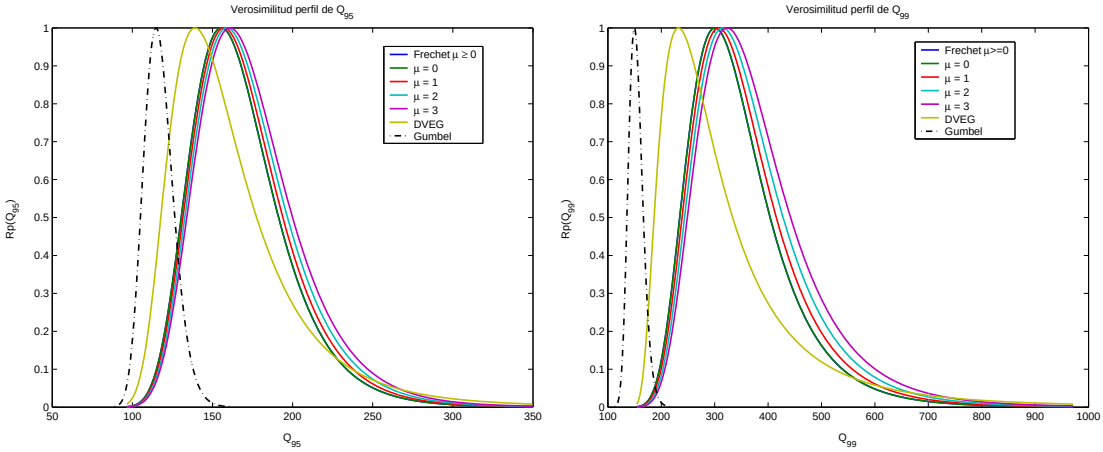


(a) (b)

Figura 2.33: (a) Verosimilitudes perfil de σ y β del modelo Fréchet con $\mu \geq 0$ (b) Contornos de verosimilitud perfil de (μ, Q_{95}) para el modelo Fréchet con $\mu \geq 0$

perfil relativa para μ alcanza su máximo en el extremo izquierdo del espacio restringido, $\mu = 0$, valor correspondiente al modelo Fréchet de dos parámetros. De la Figura (2.33 (b)) notamos que las variaciones en los extremos de los contornos se hacen muy notorias al aumentar μ. Para valores de μ menores a 3, los extremos no varían mucho considerando el contorno de plausibilidad 0.2 .

De la Figura 2.34, notamos que las curvas de verosimilitud perfil relativa para los cuantiles basadas en valores fijos de $\mu = 1, 2$ y 3 de nivel mayor a 0.7 de plausibilidad, así como $\mu = 0$, resultan muy similares. Todas son asimétricas a la derecha del EMV de Q_{95} y de Q_{99}, y de



(a) Q_{95} (b) Q_{99}

Figura 2.34: Verosimilitudes perfil relativa: (a) Cuantil 95 (b) Cuantil 99

anchura tal que asignan plausibilidad cercana a 0.5 para valores extremos como el registro de $X = 410$ mm. La diferencia entre ellas es que se desplazan ligeramente a la derecha cuando μ crece, dando prácticamente las mismas inferencias para los cuantiles. La verosimilitud perfil relativa para los cuantiles basada en el modelo Gumbel resulta ser simétrica y muy angosta, situándose a la izquierda del resto de las curvas y asignando plausibilidad cero a valores del fenómeno como el registro de 410 mm. observado en 1999, e incluso valores de acumulación de lluvia mucho menores a éste. Por otro lado, la verosimilitud perfil relativa para los cuantiles basada en la DVEG es asimétrica y de mayor anchura que la verosimilitud perfil relativa Gumbel. Esta perfil capta fenómenos de magnitud extrema como $X = 410$ mm. a un nivel de plausibilidad mucho menor respecto al modelo Fréchet de dos parámetros. Finalmente, la verosimilitud perfil relativa del modelo Fréchet restringido a $\mu \geq 0$ coincide en su totalidad con la misma curva fijando $\mu = 0$. Bajo este último modelo, se hacen plausibles niveles de acumulación de lluvias más extremas que el valor $X = 410$ mm. y asigna plausibilidad mayor que los modelos Gumbel y DVEG, proporcionando información adecuada respecto a lo observado.

En la gráfica de niveles de retorno (Figura 2.35) notamos que la banda superior de verosimilitud perfil para el modelo Gumbel aumenta ligeramente conforme el periodo de retorno crece. Coles menciona que se necesita un periodo de retorno de millones de años para que el modelo Gumbel registre fenómenos extremos como el registro de 1999. En contraste, las bandas del modelo Fréchet restringido a $\mu \geq 0$ obtenidas de la verosimilitud perfil relativa hacen plausibles a tal registro desde un periodo de retorno de 100 años. Esto es, que el registro de 1999 es plausible que suceda en promedio una vez en cien años. Las bandas de verosimilitud y del método delta para la DVEG necesitan períodos de retorno de más de 100 años para captar al registro de $X = 410$ mm.

En resumen, vemos que las inferencias obtenidas a través de la verosimilitud perfil bajo el modelo Fréchet de dos parámetros proporcionan información adecuada respecto al valor de 410 mm., además de que el ajuste del modelo es bueno según las diferentes gráficas de diagnóstico. Por otro lado, el modelo Gumbel resulta pobre para realizar inferencias. Este es un caso en el que el modelo Fréchet restringido a valores fijos de μ a nivel de plausibilidad de 0.7, proporcionan inferencias que resultan robustas incluso para Q_{99}. En cuanto a la selección del modelo, a pesar que el criterio de plausibilidad de 0.15 hace plausibles a los modelos Fréchet y Gumbel según la perfil relativa de c, el ajuste de ellos difiere considerablemente, proporcionando inferencias

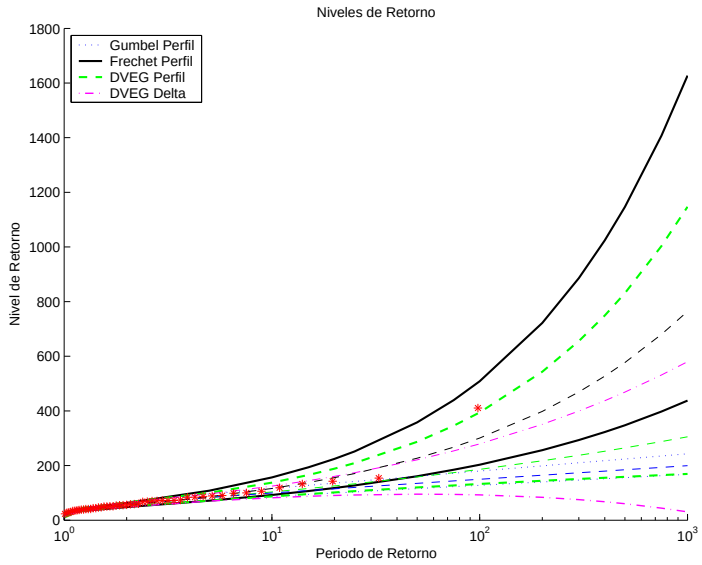


Figura 2.35: Gráfica de niveles de retorno en años

muy distintas.

En relación al modelo Gumbel, Coles comenta que aunque una prueba formal no descarte la restricción al modelo Gumbel, es mejor explotar la falta de conocimiento de la subfamilia involucrada en la DVEG a través de c. Las conclusiones del autor son que muy diversos resultados se obtienen bajo los modelos considerados a medida que se extrapola a un nivel del cuantil mayor; en particular, se requiere de un nivel de retorno de 499 años para dar inferencias adecuadas respecto al registro del año 1999. En contraste con los resultados de Coles obtenidos bajo el uso del método delta, considerar el modelo Fréchet de dos parámetros para las inferencias respecto a cuantiles grandes resulta ser adecuado según lo visto anteriormente a través de la presente propuesta. En conclusión, inferencias respecto a los niveles de acumulación de lluvias en Maiquetía sugerimos se basen en el modelo Fréchet de dos parámetros.

2.6. Niveles de lluvia en Presa Cointzio, Michoacán.

En la presente sección analizaremos un conjunto de datos de máximos anuales de precipitaciones de lluvia acumulada diaria en la estación de Cointzio, Michoacán. La estación meteorológica Presa Cointzio se localiza en el centro de la cuenca de la ciudad de Morelia, y se caracteriza por ser una de las más representativas respecto al fenómeno de lluvias en toda la cuenca, según un estudio hidrológico llevado a cabo por el IMTA (ver [10]). La estación Presa Cointzio cuenta con un periodo de registro de 58 años comprendido entre 1940 y 2002.

n	$x_{(1)}$	$\tilde{Q}_{25}$	$\tilde{Q}_{50}$	$\bar{X}$	$\tilde{Q}_{75}$	$x_{(n)}$	s
58	25.3	33.95	41.2	44.07	46.03	135.8	17.04

Cuadro 2.16: Estadísticas descriptivas

En un análisis previo, Pérez Miranda [15] comenta la existencia de dos valores de magnitudes muy diferentes al resto de las observaciones: un valor muy chico, de 3.7 mm., y un valor de 135.8 mm. Pérez Miranda opta por hacer el análisis de las observaciones sin considerar el valor

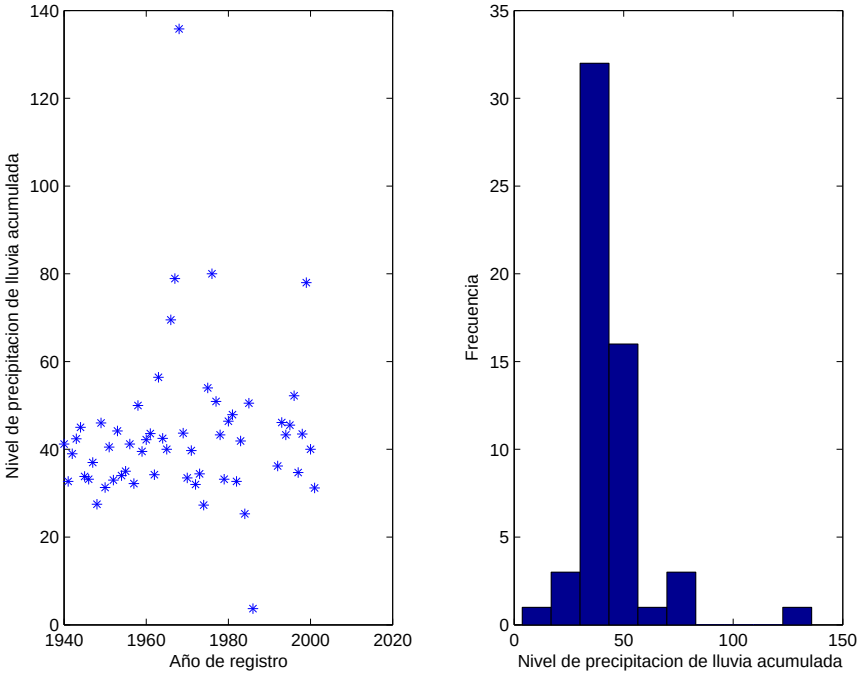


Figura 2.36: Máximos anuales de niveles de lluvia en Cointzio.

mínimo observado, y considerando el valor más grande, esto debido a que se sospecha de un posible error en la medición de la observación mínima. Siguiendo la recomendación de Pérez Miranda, realizaremos el análisis del conjunto de datos sin la observación mínima observada (en la Figura 2.36 se excluye al valor mencionado).

Modelo	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{Q}_{95}$	$\hat{Q}_{99}$
DVEG	37.02	8.1	.21	70.64	100.44

Cuadro 2.17: Estimaciones de máxima verosimilitud

En las estadísticas descriptivas y el histograma notamos la asimetría de los datos con una cola pesada a la derecha debido a $x_{(n)} = 135.8$. El ajuste de la DVEG da como resultado $\hat{c} = 0.21$. La estimación de c es positiva, mientras que la verosimilitud perfil relativa de c (Figura 2.37) está concentrada en valores positivos de c, sugiriendo únicamente una distribución Fréchet. Del mismo modo que en el ejemplo anterior, ajustaremos una distribución Fréchet restringida a $\mu \geq 0$ para garantizar que no se tenga probabilidad positiva para valores imposibles de registrarse, logrando con esto que el modelo cumpla con los requerimientos físicos del fenómeno y explorando la posibilidad del ajuste de un modelo de dos parámetros. El ajuste de la mejor familia Fréchet restringida es adecuado según la gráfica QQ para todos los datos excepto $x_{(n)}$ (Figura 2.38). Las bandas de verosimilitud cubren dicha observación. El modelo Gumbel resulta implausible a un nivel de 0.15.

La gráfica de verosimilitud perfil relativa de μ (Figura 2.39) alcanza su máximo en el borde del espacio restringido, $\mu = 0$, valor correspondiente al modelo Fréchet de dos parámetros. Valores de μ comprendidos hasta $\mu = 17$ mm. resultan plausibles a un nivel de 15 % de verosimilitud. Notamos en la gráfica de contornos de verosimilitud perfil relativa (Figura 2.40 (b)) que

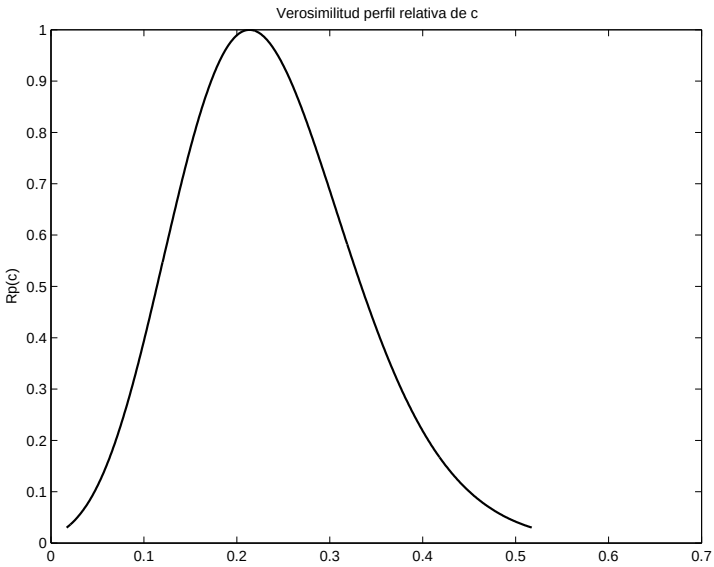
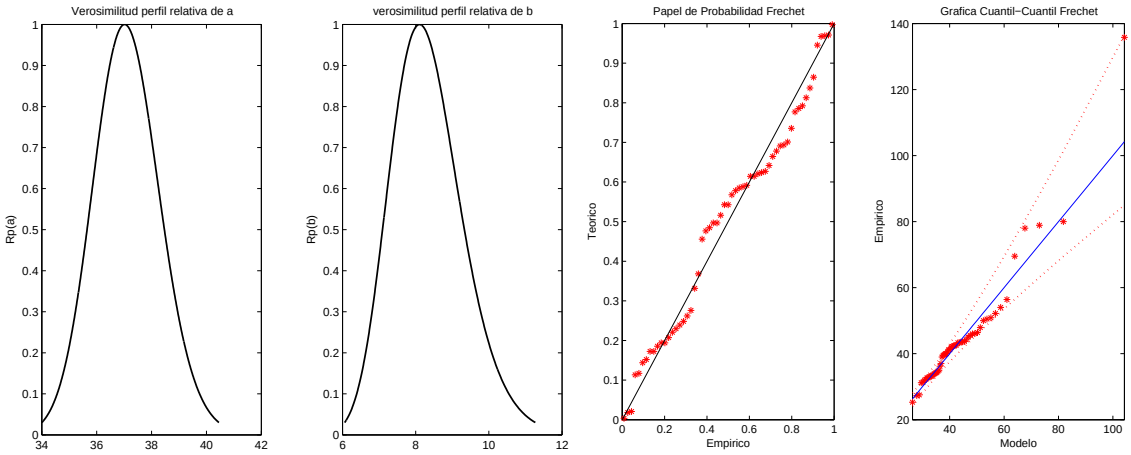


Figura 2.37: Verosimilitud perfil relativa de c del modelo DVEG



(a) (b)

Figura 2.38: (a) Verosimilitudes perfil relativa de a y b del modelo DVEG (b) Gráfica de probabilidad y gráfica Cuantil-Cuantil

Modelo	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{Q_{95}}$	$\hat{Q_{99}}$
Fréchet	0	36.99	4.57	70.87	101.26

Cuadro 2.18: EMV de los parámetros del modelo Fréchet con $\mu \geq 0$

cortes paralelos al eje de Q_{95} producen intervalos cuyos extremos varían poco al considerar diferentes valores de μ menores a 8.

En la verosimilitud perfil relativa para Q_{95} y Q_{99} (Figura 2.41) notamos que las gráficas basadas en valores de $\mu = 0; 2; 5; 8$ asociados a niveles de plausibilidad mayores a 0.9 son simétricas y se traslapan en gran parte de ellas. Las variaciones de estas curvas al considerar los diferentes valores de μ son mínimas, por lo cual tenemos nuevamente un caso donde las inferencias de los cuantiles son robustas ante cambios en valores altamente plausibles de μ. Por otro lado, el modelo Fréchet restringido a $\mu \geq 0$ coincide en su totalidad con el modelo basado

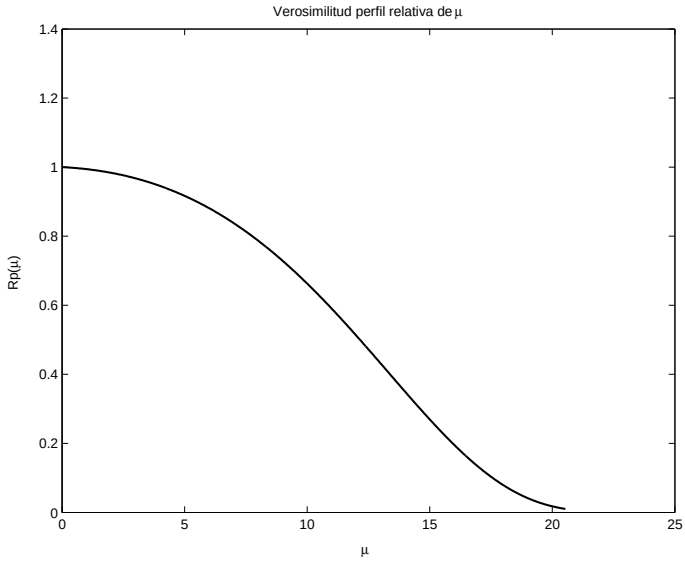
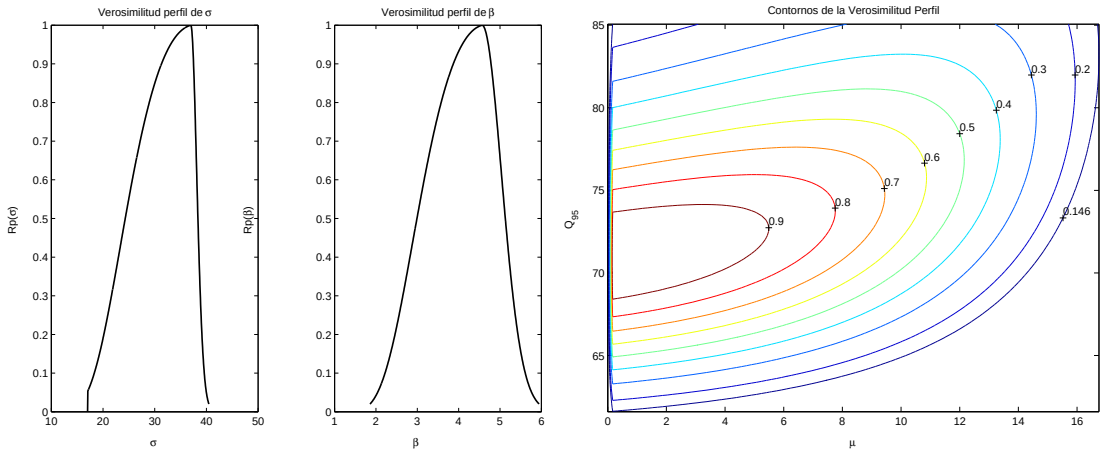


Figura 2.39: Verosimilitud perfil relativa de μ del modelo Fréchet con $\mu \geq 0$



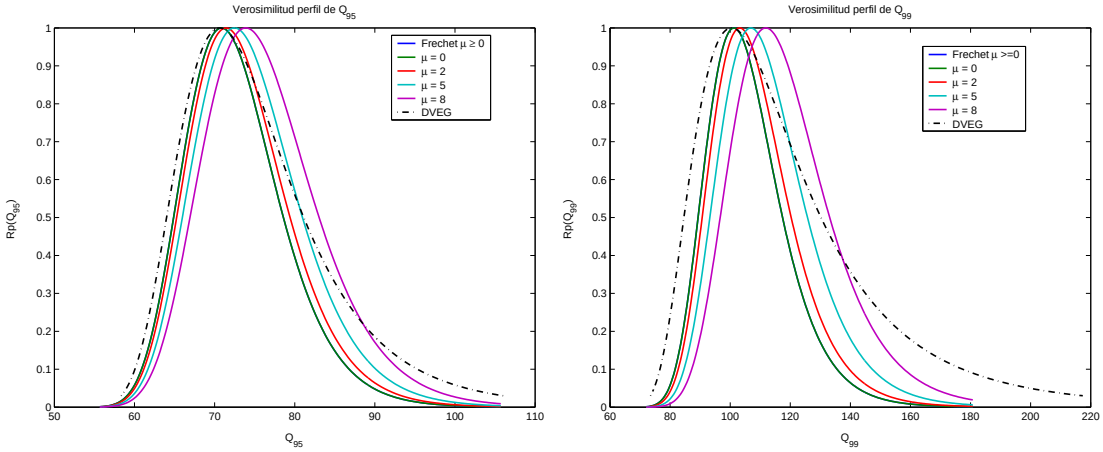
(a) (b)

Figura 2.40: (a) Verosimilitudes perfil relativa de σ y β del modelo Fréchet con $\mu \geq 0$ (b) Contornos de verosimilitud perfil relativa de (μ, Q_{95}) para el modelo Fréchet con $\mu \geq 0$

en $\mu = 0$ (modelo Fréchet de dos parámetros), mientras que la DVEG resulta de mayor anchura que todas las anteriores. Para el valor $x_{(n)} = 135.8$, se tiene que la verosimilitud perfil relativa de Q_{99} basada en la DVEG asigna plausibilidad mucho mayor que la misma curva basada en el modelo Fréchet de dos parámetros.

De la gráfica de niveles de retorno (Figura 2.42) notamos que los intervalos para los cuantiles basados en la verosimilitud perfil relativa para el modelo Fréchet de dos parámetros captan el valor de $x_{(n)} = 135.8$ mm. para un período de retorno de 100 años. Por otro lado, la asimetría hacia valores altos de los cuantiles para las bandas de verosimilitud en la DVEG hace que se capte el registro de $X_{(n)} = 135.8$ mm. para períodos de retorno menores a 100 años; las bandas basadas en el método delta también captan a las observaciones.

Este es un caso donde al ajustar un modelo Fréchet restringiendo μ a ser no negativo se observa que el modelo Fréchet de dos parámetros resulta adecuado para representar el fenómeno



(a) Q_{95} (b) Q_{99}

Figura 2.41: Verosimilitudes perfil: (a) Cuantil 95 (b) Cuantil 99

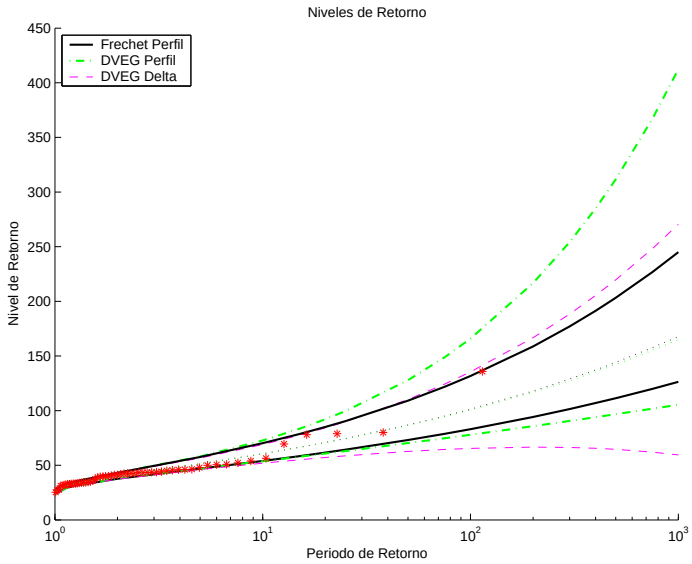


Figura 2.42: Gráfica de niveles de retorno en años

de lluvias. Para este modelo, se tiene robusticidad para la inferencia en los cuantiles ante cambios en μ. Puesto que el modelo Fréchet de dos parámetros proporciona inferencias adecuadas respecto a lo observado, ajusta bien a la mayoría de las observaciones y es más sencillo, la selección del modelo Fréchet de dos parámetros se recomienda para representar la distribución del máximo de nivel de acumulación de lluvias en Cointzio.

2.7. Niveles de acumulación de nieve en Carolina del Norte, E.U.A.

Smith [9] presenta un conjunto de observaciones referente a la acumulación de caida de nieve durante el mes de enero en el aeropuerto Raleigh-Durham ubicado en el estado de Carolina del Norte, E.U.A. El problema suscitado que hace relevante la aplicación de la Teoría de Valores

Extremos es el registro de un valor excepcional (20.3 pulgadas) que condujo a consecuencias graves en la zona. Lo que se pretende explorar es si es posible obtener información relevante para un valor como éste, a través de la herramienta estadística y basados en registros anteriores del fenómeno. La Figura 2.43 muestra los máximos de cada més de enero registrados durante los años de 1948 a 1998. Cabe resaltar el tamaño de muestra pequeño de 29 máximos.

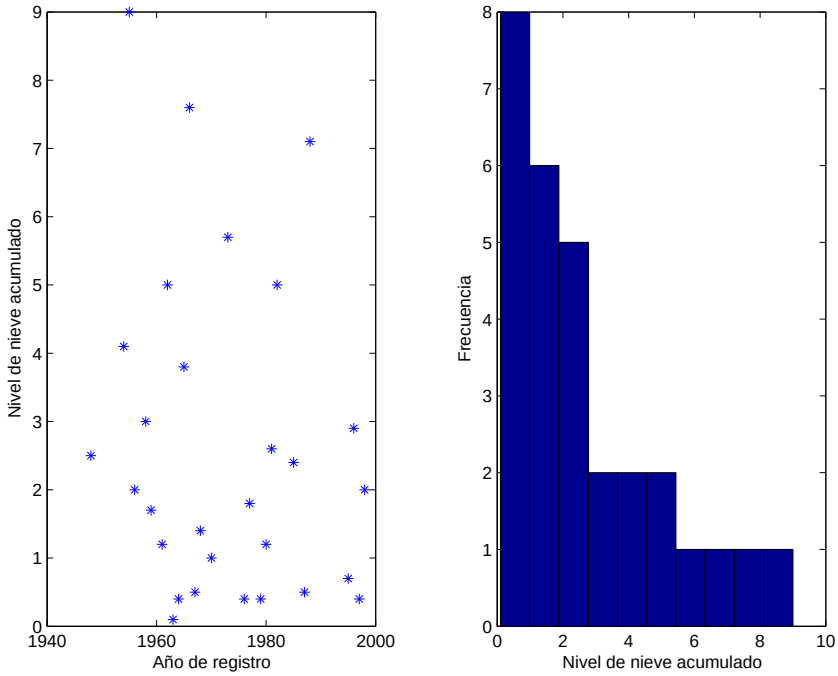


Figura 2.43: Máximos de niveles de acumulación de nieve durante el mes de enero.

n	$x_{(1)}$	$\tilde{Q}_{25}$	$\tilde{Q}_{50}$	$\bar{X}$	$\tilde{Q}_{75}$	$x_{(n)}$	s
29	0.1	0.65	2	2.63	3.88	9	2.39

Cuadro 2.19: Estadísticas descriptivas

Modelo	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{Q}_{95}$	$\hat{Q}_{99}$
DVEG	1.24	1.19	0.5	9.39	22.66

Cuadro 2.20: Estimaciones de máxima verosimilitud

La verosimilitud perfil relativa para c (Figura 2.44) resulta ancha y se concentra en valores positivos de c. El EMV $\hat{c}$ es 0.5, y la plausibilidad de $c = 0$ es 0.098, dando poco sustento al modelo Gumbel y, por consecuencia, sugiriendo únicamente al modelo Fréchet. El modelo Fréchet restringido a $\mu \geq 0$ no ajusta adecuadamente según la gráfica QQ (Figura 2.45), sobreestimando considerablemente a lo observado.

La verosimilitud perfil de μ (Figura 2.46) alcanza su máximo en μ=0, y sólo valores muy cercanos a μ=0 resultan plausibles, haciendo alusión al modelo Fréchet de dos parámetros. Por otro lado, las gráficas de verosimilitud perfil relativa para σ y β (Figura 2.47 (a)) son simétricas

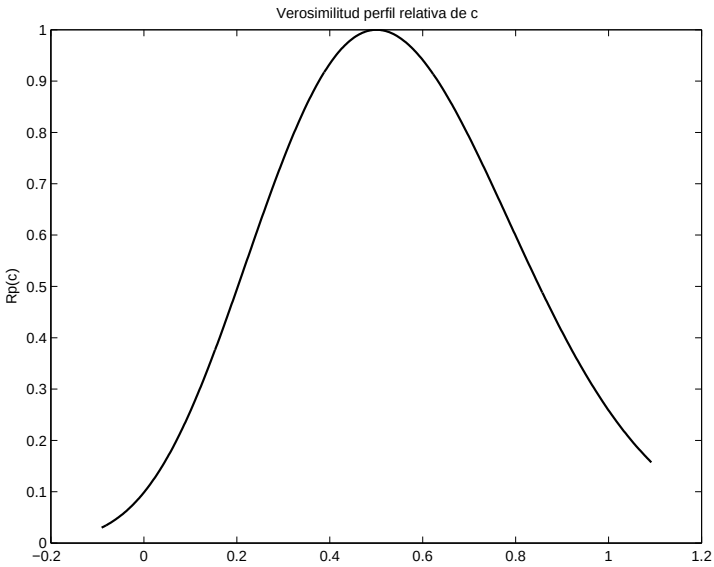
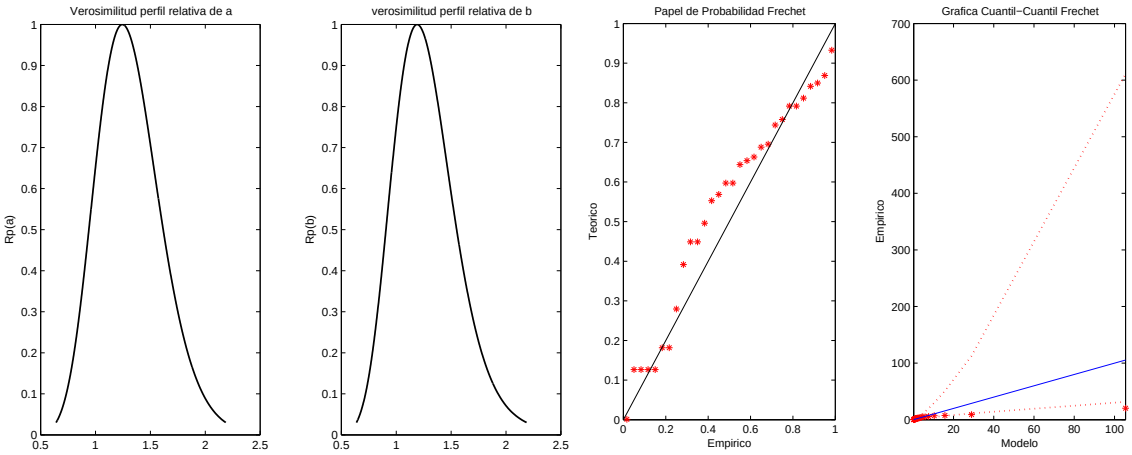


Figura 2.44: Verosimilitud perfil relativa de c del modelo DVEG



(a) (b)

Figura 2.45: (a) Verosimilitudes perfil de a y b del modelo DVEG (b) Gráfica de probabilidad y gráfica Cuantil-Cuantil

Modelo	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{Q_{95}}$	$\hat{Q_{99}}$
Fréchet	0	36.99	4.57	70.87	101.26

Cuadro 2.21: EMV del modelo Fréchet con $\mu \geq 0$

y muy angostas, resultando informativas. También, la gráfica de contornos de verosimilitud relativa para (μ, Q_{95}) (Figura 2.47 (b)) muestran que cambios de μ producen intervalos para Q_{95} cuyos extremos varían mucho, por lo que no hay indicios de robusticidad para este cuantil.

Las gráficas de verosimilitud perfil relativa para Q_{95} y Q_{99} (Figura 2.48) muestran que el modelo Fréchet restringido a $\mu \geq 0$ (que coincide con el modelo Fréchet de dos parámetros) y los modelos bajo los distintos valores de μ considerados, captan adecuadamente el valor de $x = 20.3$ pulgadas, cuya plausibilidad es mayor a 0.4. Sin embargo, las curvas resultan ser extremadamente asimétricas a la derecha del EMV de los cuantiles, dando plausibilidad a

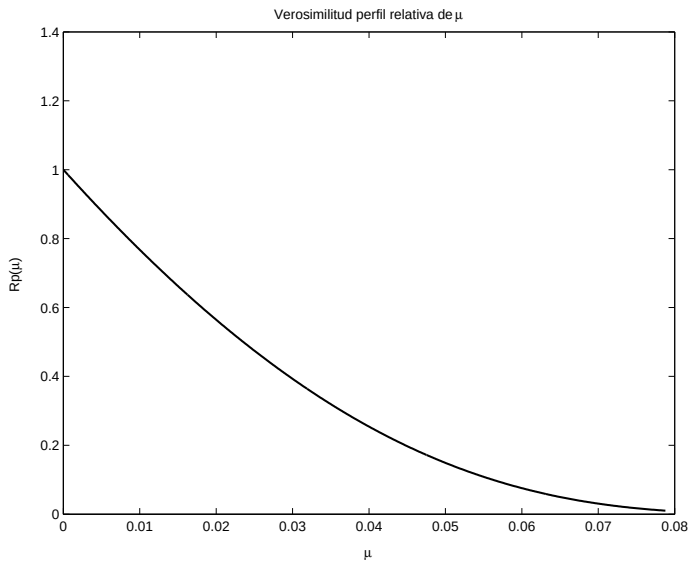
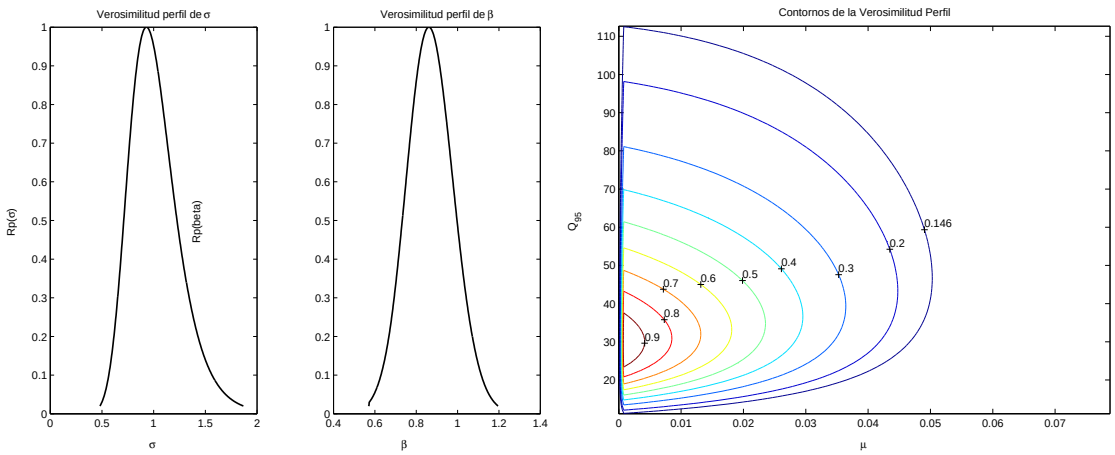


Figura 2.46: Verosimilitud perfil de μ del modelo Fréchet con $\mu \geq 0$



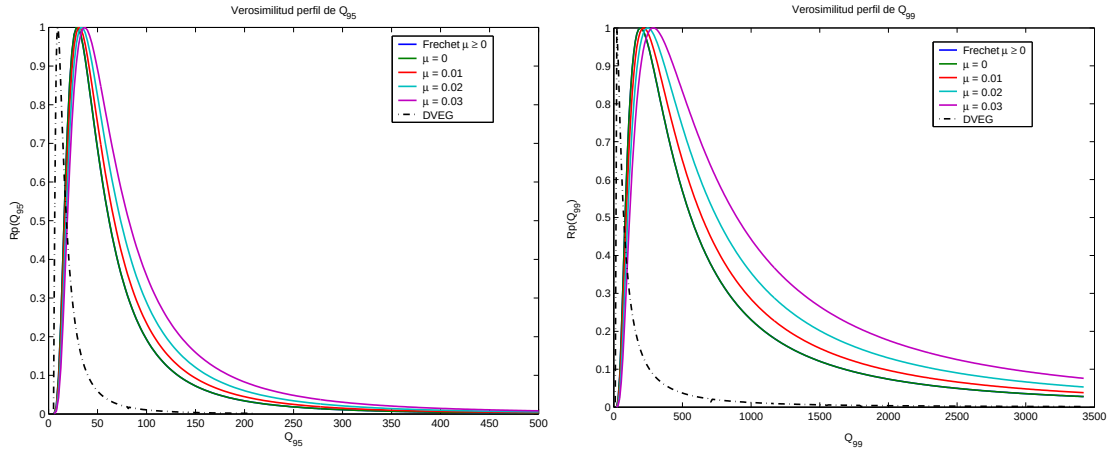
(a) (b)

Figura 2.47: (a) Verosimilitudes perfil de σ y β del modelo Fréchet con $\mu \geq 0$ (b) Contornos de Verosimilitud Perfil de (μ, Q_{95}) para el modelo Fréchet con $\mu \geq 0$

niveles de acumulación de nieve que resultan absurdos.

En la gráfica de niveles de retorno (Figura 2.49) notamos lo extremo que resulta la asimetría de los intervalos de verosimilitud para el modelo Fréchet de dos parámetros en contraste con la DVEG, abarcando niveles de acumulación de nieve de más de 16000 pulgadas para un periodo de retorno de 1000 años. Estas inferencias inadecuadas para el nivel de acumulación de nieve no hacen atractivo el uso del modelo Fréchet de dos parámetros, y procederemos a ajustar el modelo Fréchet de tres parámetros para realizar el análisis de los datos de nieve.

La verosimilitud perfil relativa de μ (Figura 2.50 (a)) alcanza su máximo en $\mu = -1.13$, y es muy asimétrica a la izquierda de $\hat{\mu}$. Esta curva es muy angosta para valores de μ entre -5 y 0; en contraste, la cola izquierda de la verosimilitud perfil relativa se extiende en un nivel de 15 % de verosimilitud y abarca valores mucho menores a -5. Valores de μ no negativos tienen plausibilidad prácticamente 0. Las verosimilitudes para σ y β resultan muy asimétricas a la



(a) Q_{95} (b) Q_{99}

Figura 2.48: Verosimilitudes perfil: (a) Cuantil 95 (b) Cuantil 99

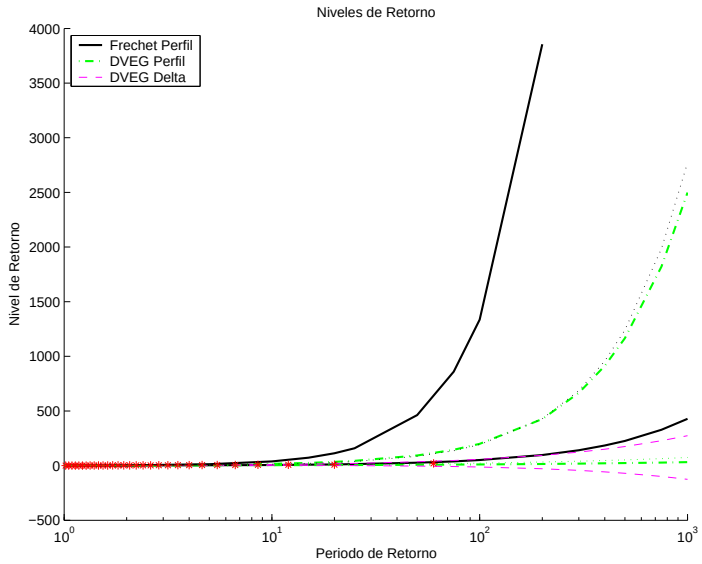


Figura 2.49: Gráfica de niveles de retorno anual respecto los meses de enero.

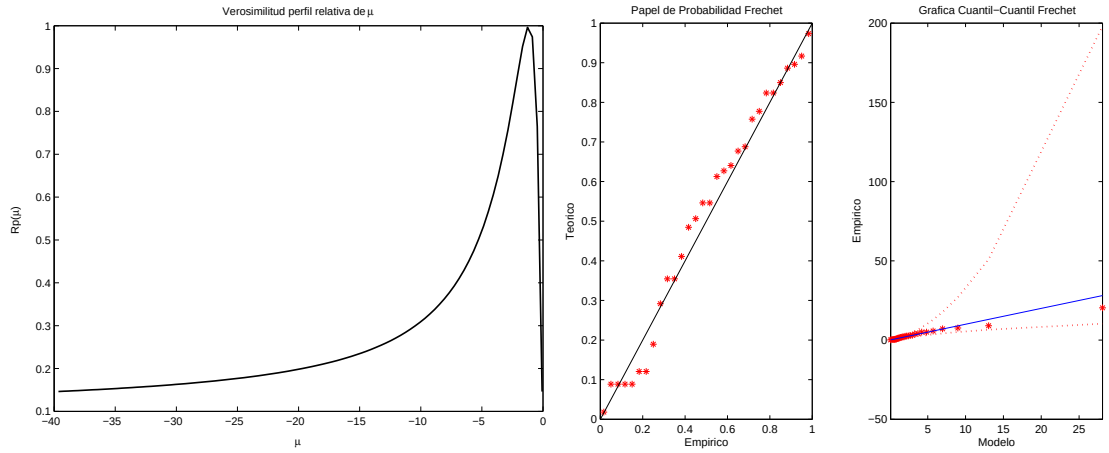
derecha del EMV correspondiente. Por otro lado, la gráfica de contornos de verosimilitud perfil relativa para (μ, Q_{95}) (Figura 2.51) muestran que cambios de μ proporcionan intervalos para Q_{95} que varían considerablemente tanto en longitud como en los valores de los extremos.

La gráfica de verosimilitud perfil relativa para Q_{95} (Figura 2.52) basada en el modelo Fréchet de tres parámetros da plausibilidad mayor a 0.4 al valor extremo de $x = 20.3$ pulgadas. La gráfica resulta asimétrica a la derecha de $\hat{Q}_{95}$, y asigna plausibilidad cercana a 0 a valores de Q_{95} de magnitudes no tan extensas respecto a lo observado. La gráfica correspondiente a Q_{99} tiene una forma similar a la de Q_{95}. La extrapolación a Q_{99} nos lleva a una curva de verosimilitud perfil relativa de Q_{99} que asigna plausibilidad baja a valores muy grandes de acumulación de nieve respecto a los datos históricos. Las curvas de verosimilitud perfil relativa para Q_{95} y Q_{99} basadas en los diferentes valores de μ propuestos son asimétricas y varían considerablemente, por lo que es claramente no robusto.

Para la familia Fréchet general, la probabilidad máxima que se tiene para valores negativos

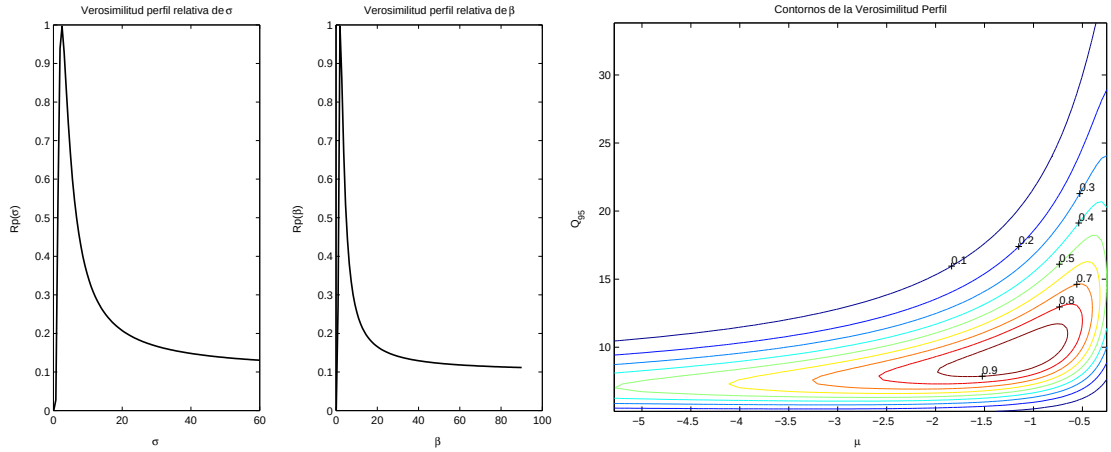
Modelo	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{Q}_{95}$	$\hat{Q}_{99}$
Frechét	-1.13	2.38	2	9.39	22.66

Cuadro 2.22: EMV del modelo Fréchet



(a) (b)

Figura 2.50: (a) Verosimilitud perfil de μ del modelo Fréchet (b) Gráfica de probabilidad y gráfica QQ para el modelo Fréchet

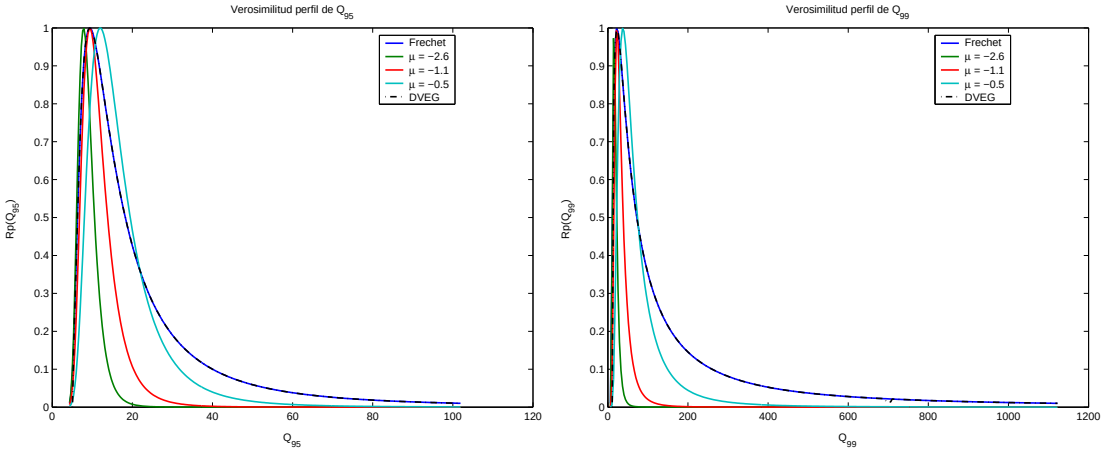


(a) (b)

Figura 2.51: (a) Verosimilitudes perfil de σ y β del modelo Fréchet (b) Contornos de Verosimilitud Perfil de (μ, Q_{95})

de la variable considerando todos los modelos Fréchet en el cálculo de la verosimilitud perfil relativa de los cuantiles es de 0.029, mucho mayor que las probabilidades para valores negativos en los ejemplos anteriores.

La gráfica de niveles de retorno (Figura 2.53) muestra la incertidumbre en la extrapolación a niveles de retorno altos. Las bandas de verosimilitud para el modelo Fréchet de tres parámetros (que coinciden con las bandas de verosimilitud del modelo DVEG) captan a las observaciones. Para niveles de retorno de hasta 20 años se tienen intervalos de verosimilitud que resultan apropiados, ya que abarcan valores de los cuantiles que no distan mucho de lo observado. Para niveles de retorno mayores o iguales a 100 años las bandas se extienden a valores de magnitud



(a) Q_{95} (b) Q_{99}

Figura 2.52: Verosimilitudes perfil: (a) Cuantil 95 (b) Cuantil 99

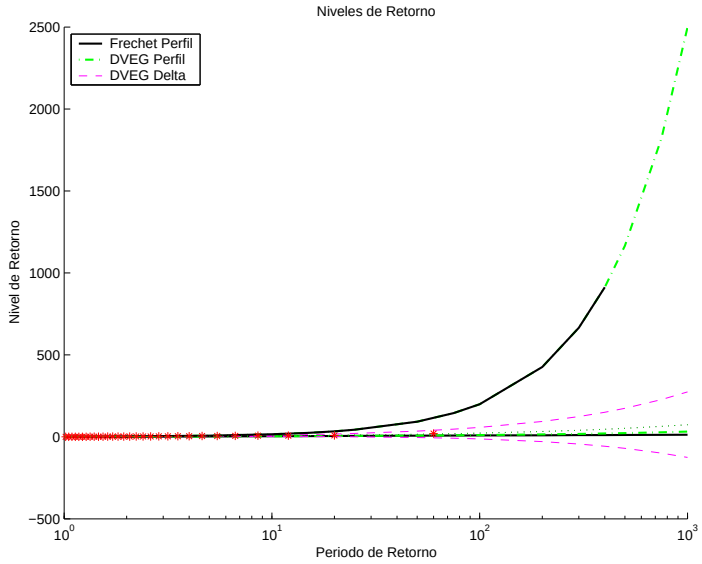


Figura 2.53: Gráfica de niveles de retorno anual respecto al més de enero

mucho mayor que resultan, aunque no tan extremos como en el caso Fréchet de dos parámetros, no adecuados en términos de la acumulación de nieve.

Para el caso de niveles de acumulación de nieve, el modelo Fréchet de dos parámetros sobreestima considerablemente las observaciones más grandes, y la extrapolación a niveles de retorno superiores a 100 años no hace sentido con las observaciones. El modelo Fréchet de tres parámetros proporciona estimaciones más adecuadas respecto a los datos y para el valor de $x = 20.3$ pulgadas. Las inferencias para Q_{95} hacen razonable al suceso extremo del año de 1999. El tamaño de muestra de 29 datos hace que la extrapolación a Q_{99} resulte poco confiable.

2.8. Comentarios finales

De la aplicación de la metodología propuesta a los 6 primeros casos aquí expuestos, encontramos que la verosimilitud perfil relativa proporciona estimaciones congruentes para los

parámetros de interés considerados. Esta misma herramienta proporciona información útil para sugerir submodelos.

Por otro lado, notamos en general la asimetría en la cola derecha de las gráficas de verosimilitud perfil relativa para los cuantiles, abarcando con ello valores que no contemplan los intervalos basados en el método delta. En ocasiones, esta diferencia hace que los intervalos de verosimilitud incluyan a los datos registrados, lo cual da sustento al uso de este método de estimación.

En el caso Fréchet, la restricción del espacio parametral para la estimación puede conducir a modelos más sencillos que describen el comportamiento del máximo satisfactoriamente. Estas restricciones mediante el uso del modelo general DVEG no resultan simples de implementar debido a la combinación de sus parámetros en la delimitación de su soporte.

El último ejemplo analizado resalta la precaución que se debe tener en la extrapolación a cuantiles muy grandes cuando se tiene una muestra pequeña. Aún cuando se tengan muestras de tamaño razonable, puede darse el caso de que el ajuste de cualquiera de las DVE no sea adecuado. Esto sería indicio de que algunas de las suposiciones en la modelación del máximo por bloques puede no hacer sentido.

Finalmente, las variaciones de las inferencias de los cuantiles ante cambios en el parámetro umbral μ se contempla en forma automática en la verosimilitud perfil relativa de los cuantiles Weibull o Fréchet generales.

Capítulo 3

Simulaciones

3.1. Introducción

El nivel de confianza asociado a los intervalos de verosimilitud y del método delta se obtiene mediante la distribución aproximada de la razón de verosimilitud (1.30) y la distribución asintótica normal de los estimadores de máxima verosimilitud. Estas distribuciones asintóticas se basan en condiciones de regularidad, una de las cuales indica que el soporte de la distribución no debe ser función de sus parámetros. Para la DVEG, así como los submodelos Weibull y Fréchet, dicha condición no se cumple, por lo que es necesario verificar empíricamente las coberturas de los intervalos obtenidos a través de métodos asintóticos.

Además, en el capítulo anterior notamos que las gráficas de verosimilitud perfil relativa de los cuantiles Q_{95} y Q_{99} tanto para la DVEG como para la subclase Weibull o Fréchet de tres parámetros (sin restriccion de estos) correspondiente a $\hat{c}$ coincidían cuando la verosimilitud perfil relativa del parámetro de forma c de la DVEG hacía plausible exclusivamente a a la subclase considerada. Cuando la verosimilitud perfil relativa de c hacía plausible a las tres subclases, las gráficas de verosimilitud perfil relativa para Q_{95} y Q_{99} resultaban más amplias para la DVEG en comparación con la subclase correspondiente a $\hat{c}$.

En esta sección se presenta un estudio basado en simulaciónes el cual tiene por objeto verificar si la cobertura empírica de los intervalos de verosimilitud propuestos y del método delta tradicionalmente usados corresponden a la cobertura aproximada según la teoría. Como segundo objetivo, se compararán las longitudes de los intervalos de verosimilitud de la DVEG contra los obtenidos para la subclase Weibull o Fréchet.

3.2. Planteamiento de los escenarios de simulación

Los escenarios de simulación están conformados de la siguiente manera: se simularon 10000 muestras con tamaños de muestra de máximos de 50, 100 y 500. La elección del tamaño se debió al hecho de querer representar tamaños de muestra encontrados en diversas aplicaciones de valores extremos, y también explorar el efecto del aumento de muestra en la inferencia. Las muestras fueron simuladas de la DVEG con parámetros $a = 1$, $b = 1$, y $c = 1$, 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.01, 0, -0.01, -0.05, -0.1, -0.3 y -0.5. El motivo de la elección de los valores de c es para considerar casos donde claramente se tiene un submodelo Fréchet o Weibull, además de explorar el efecto para los casos donde c se aproxima a cero, que en teoría, es donde se está más cerca del modelo Gumbel. La elección de los parámetros a y b se debe a simplicidad del trato

con ellos.

Para cada muestra, se calculan los intervalos de 15 % de verosimilitud para Q_{95} y Q_{99} a través del modelo correspondiente a la estimación puntual $\hat{c}$ por ser el modelo que más fuertemente se sugiere y considerar una selección automática del submodelo, y se calculan los intervalos de 15 % de verosimilitud y los intervalos a través del método delta de 95 % de confianza para Q_{95} y Q_{99} basados en la DVEG.

3.3. Resultados de coberturas

A continuación presentamos los resultados obtenidos de las simulaciones descritas referente a coberturas.

En las gráficas que se muestran en la Figura 3.1 el eje horizontal se encuentran graficados los valores de c a partir de los cuales se generaron las simulaciones; el eje vertical representa el número de muestras cuyos intervalos cubrieron el verdadero valor del cuantil. Los asteriscos representan el número de intervalos de verosimilitud perfil basados en el submodelo Weibull o Fréchet que cubrieron el verdadero valor del cuantil; los diamantes y los triángulos representan el número de intervalos de verosimilitud perfil y método delta, respectivamente, basados en la DVEG que cubrieron el verdadero valor del cuantil. La Figura 3.1 muestra las gráficas estructuradas en 2 columnas y tres renglones. La columna de la izquierda hace referencia a las coberturas para Q_{95}, mientras que la columna de la derecha corresponde a Q_{99}; los renglones representan las coberturas de los intervalos clasificados según su tamaño de muestra: el de arriba corresponde a 50 máximos, el central a 100 y el de abajo a 500 máximos.

El número adecuado de intervalos que cubren el verdadero valor del cuantil en cuestión es de 9500. En general, notamos que los intervalos de verosimilitud tienen mejor cobertura que los intervalos basados en el método delta. A medida que aumenta el tamaño de muestra, los últimos mejoran sus coberturas resultando muy similares a los primeros. Las mayores discrepancias en las coberturas para los diferentes intervalos se observan para las muestras pequeñas de 50 máximos.

Cuando $c = -0.3$, las coberturas de los intervalos basados en el método delta para n=50 se ubican por debajo del 90 %. Recordemos que en estos casos, Smith [16] encontró que los estimadores máximo verosimil existen, aunque la aproximación normal asintótica para estos parámetros no es adecuada.

La Tabla 3.2 presenta el número de intervalos que cubrieron al verdadero valor del cuantil, así como el número de intervalos que lo subestimaron (columnas marcadas con signo +) y lo sobreestimaron (columnas marcadas con signo -). De esta tabla observamos que los intervalos basados en el método delta para la DVEG casi en su totalidad subestiman los verdaderos valores de Q_{95} y Q_{99}. En contraste, para la DVEG como para el submodelo elegido Weibull o Fréchet se tienen intervalos de verosimilitud que sobreestiman en mayor proporción a los primeros. Resaltan las simulaciones generadas a partir de valores de $c = -0.01, 0, 0.001, 0.01, 0.05$ y 0.1 (valores de c cercanos a $c = 0$ asociados con modelos cercanos a un Gumbel) con tamaño de muestra de 50 máximos, ya que las coberturas de los intervalos de verosimilitud del submodelo elegido, a pesar de ser razonables, son menores a las coberturas de verosimilitud perfil de la DVEG. En los casos $|c| \leq 0.1$ con tamaño de muestra $n = 50$, presentamos diagramas de caja (Figuras 3.3 y 3.4) que representan la variabilidad de las estimaciones de c correspondientes a los intervalos que subestimaron el verdadero valor de los cuantiles propuestos.

Notamos que para los intervalos que subestimaron al verdadero valor de Q_{95} y Q_{99} en los

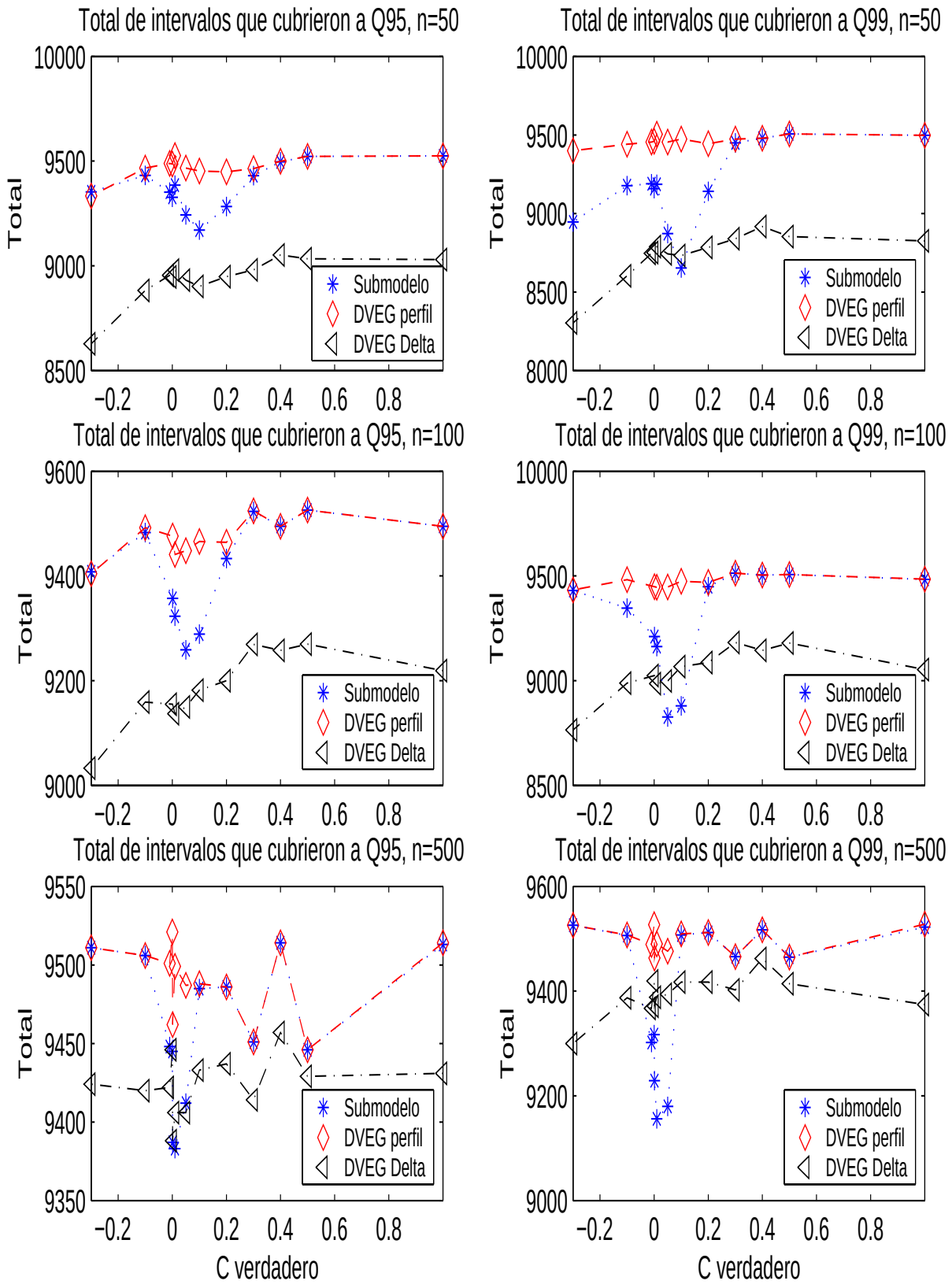


Figura 3.1: Coberturas empíricas

	Modelo	Submodelo Q95 Perfil			DVEG Q95 Perfil			DVEG Q95 Delta			Submodelo Q99 Perfil			DVEG Q99 Perfil			DVEG Q99 Delta		
C Verdadero	n	+	Cobertura	-	+	Cobertura	-	+	Cobertura	-	+	Cobertura	-	+	Cobertura	-	+	Cobertura	-
1	50	251	9525	224	251	9525	224	0	9029	971	284	9498	218	284	9498	218	0	8825	1175
1	100	262	9495	243	262	9495	243	0	9219	781	290	9484	226	290	9485	225	0	9053	947
1	500	246	9513	241	246	9514	240	61	9431	508	248	9522	230	247	9528	225	25	9374	601
0.5	50	212	9522	266	212	9522	266	0	9034	966	231	9506	263	231	9506	263	0	8853	1147
0.5	100	228	9526	246	228	9526	246	0	9270	730	255	9507	238	255	9507	238	0	9180	820
0.5	500	274	9446	280	274	9446	280	100	9429	471	279	9465	256	279	9465	256	60	9414	526
0.4	50	213	9496	291	213	9499	288	0	9052	948	240	9478	282	240	9479	281	0	8917	1083
0.4	100	231	9495	274	231	9495	274	5	9258	737	244	9504	252	244	9504	252	0	9144	856
0.4	500	226	9514	260	226	9514	260	85	9457	458	232	9517	251	232	9517	251	44	9462	494
0.3	50	191	9431	378	191	9464	345	0	8981	1019	221	9450	329	221	9473	306	0	8839	1161
0.3	100	209	9523	268	209	9524	267	5	9269	726	233	9513	254	233	9513	254	0	9182	818
0.3	500	256	9451	293	256	9451	293	114	9414	472	274	9466	260	274	9466	260	73	9402	525
0.2	50	198	9283	519	203	9449	348	0	8948	1052	198	9141	661	236	9444	320	0	8783	1217
0.2	100	226	9433	341	226	9464	310	7	9200	793	235	9448	317	235	9469	296	0	9087	913
0.2	500	244	9486	270	244	9486	270	119	9437	444	241	9512	247	241	9512	247	72	9417	511
0.1	50	179	9171	650	176	9452	372	2	8902	1096	185	8653	1162	183	9474	343	0	8732	1268
0.1	100	215	9289	496	214	9466	320	14	9182	804	219	8879	902	219	9474	307	0	9067	933
0.1	500	239	9485	276	239	9488	273	121	9433	446	230	9507	263	230	9509	261	74	9417	509
0.05	50	190	9243	567	188	9466	346	6	8931	1063	205	8872	923	197	9455	348	0	8743	1257
0.05	100	218	9259	523	212	9448	340	17	9150	833	209	8826	965	203	9447	350	3	8999	998
0.05	500	230	9412	358	229	9487	284	102	9406	492	233	9180	587	232	9476	292	65	9394	541
0.01	50	155	9385	460	150	9526	324	11	8979	1010	193	9186	621	160	9504	336	0	8789	1211
0.01	100	215	9323	462	207	9441	352	20	9137	843	228	9163	609	192	9446	362	3	8986	1011
0.01	500	240	9383	377	221	9499	280	104	9406	490	251	9156	593	225	9490	285	63	9388	549
0.001	100	209	9357	434	190	9476	334	30	9155	815	257	9210	533	204	9449	347	1	9024	975
0.001	500	254	9387	359	236	9462	302	129	9388	483	321	9229	450	257	9463	280	74	9371	555
0	50	170	9327	503	163	9486	351	8	8947	1045	219	9157	624	180	9459	361	0	8751	1249
0	500	220	9445	335	203	9521	276	105	9446	449	280	9317	403	215	9527	258	71	9419	510
-0.01	50	168	9352	480	163	9488	349	8	8954	1038	230	9189	581	180	9455	365	0	8745	1255
-0.01	500	232	9448	320	201	9501	296	101	9422	477	368	9302	330	227	9489	265	68	9367	565
-0.1	50	156	9432	412	151	9466	383	13	8882	1105	268	9177	555	161	9440	399	0	8602	1398
-0.1	100	177	9483	340	174	9492	334	26	9159	815	307	9346	347	176	9482	342	2	8988	1010
-0.1	500	194	9506	300	194	9506	300	108	9420	472	205	9506	289	205	9507	288	55	9387	558
-0.3	50	144	9353	503	165	9334	501	70	8628	1302	100	8946	954	98	9400	502	2	8302	1696
-0.3	100	175	9408	417	179	9404	417	84	9033	883	142	9431	427	141	9433	426	2	8764	1234
-0.3	500	166	9511	323	166	9511	323	110	9424	466	155	9526	319	155	9526	319	38	9300	662

Figura 3.2: Coberturas empíricas para Q_{95}

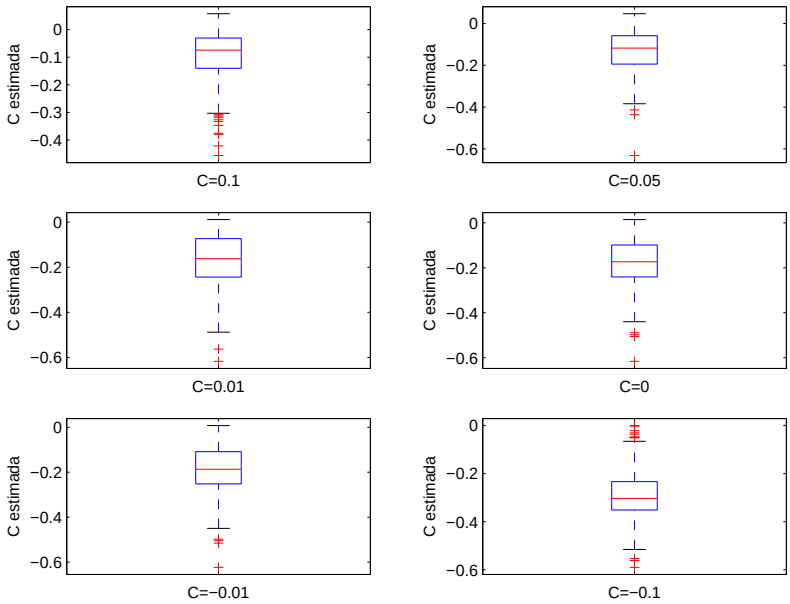


Figura 3.3: Estimaciones de c para los intervalos que subestimaron el verdadero valor de Q_{95} para $n = 50$ máximos

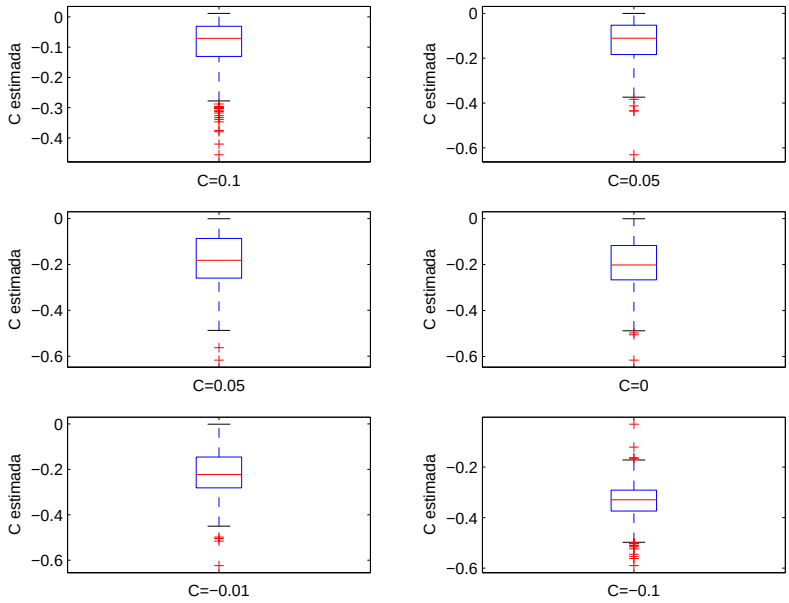


Figura 3.4: Estimaciones de c para los intervalos que subestimaron el verdadero valor de Q_{99} para $n = 50$ máximos.

casos de c mencionados con $n = 50$, corresponden a estimaciones de c menores al verdadero valor de c. La mayoría de las estimaciones están a 1 unidad por debajo del verdadero valor de c, considerando con ello modelos de colas más ligeras que el verdadero modelo. Es decir, se eligieron modelos Weibull cuando el verdadero modelo es un Fréchet, de cola más pesada.

3.4. Cociente de intervalos

Los diagramas de caja mostrados en las Figuras 3.5 , 3.6 y 3.7 representan la variabilidad del cociente de la longitud del intervalo de verosimilitud del submodelo elegido según el EMV $\hat{c}$ entre la longitud del intervalo de verosimilitud basado en la DVEG, para Q_{95} y Q_{99}, considerando los diferentes tamaños de muestra.

De estas gráficas notamos que en general, para valores de $|c| > 0.1$ los cocientes de los intervalos son idénticos, mientras que en el caso $|c| \leq 0.1$, notamos que el intervalo de verosimilitud del submodelo resulta ser de menor longitud. Para estos últimos valores de c, la mayor reducción del intervalo se obtiene para Q_{99}, mientras que para la muestra de tamaño 50 tambien se tiene reducción del intervalo. Además, se tiene mayor reducción en la longitud de los intervalos para los casos Weibull en relación a los casos Fréchet.

Para $n = 50$, notamos que se tiene una reducción en la longitud de los intervalos basados en la elección automática del submodelo correspondiente a $\hat{c}$ cuando $-0.1 \leq c \leq 0.1$. Esta reducción se hace más notoria para los casos Weibull y para Q_{99}. Hay que considerar en especial el caso $c = 0.1$ para Q_{99}, pues es el caso para el cual se registró la cobertura más baja para Q_{99}. En el diagrama de caja notamos que para este caso se tiene la menor ganancia en la longitud del intervalo al elegir automáticamente el submodelo asociado a $\hat{c}$, ya que el diagrama de caja es más angosto y más cercano a 1. La línea inferior de ese diagrama de caja muestra la extensión del resto de los cocientes menores a $\tilde{Q}_{25}$, los cuales alcanzan valores del cociente de 0.7.

La Figura 3.8 presenta la relación entre los cocientes de intervalos que fueron menores a 0.9 respecto a la estimación $\hat{c}$, basados en las muestras que subestimaron el verdadero valor de

Q_{99}. En ella vemos que estos cocientes corresponden a modelos Weibull, y el cociente alcanza valores hasta de 0.3.

Para $n = 100$ y $n = 500$, basándonos en los casos donde $|c| \leq 0.1$, en general se tiene ganacia en la gran mayoría de los cocientes, en especial para Q_{99}.

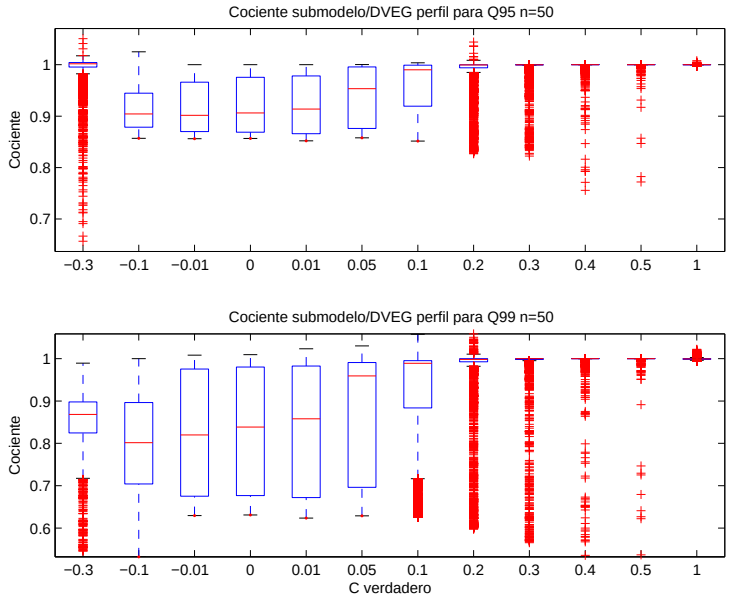


Figura 3.5: Cocientes de intervalos n=50

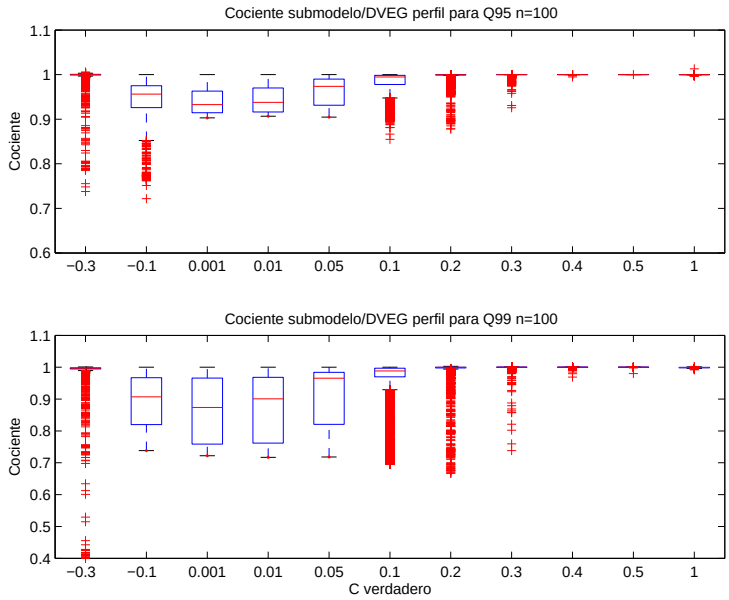


Figura 3.6: Cocientes de intervalos n=100

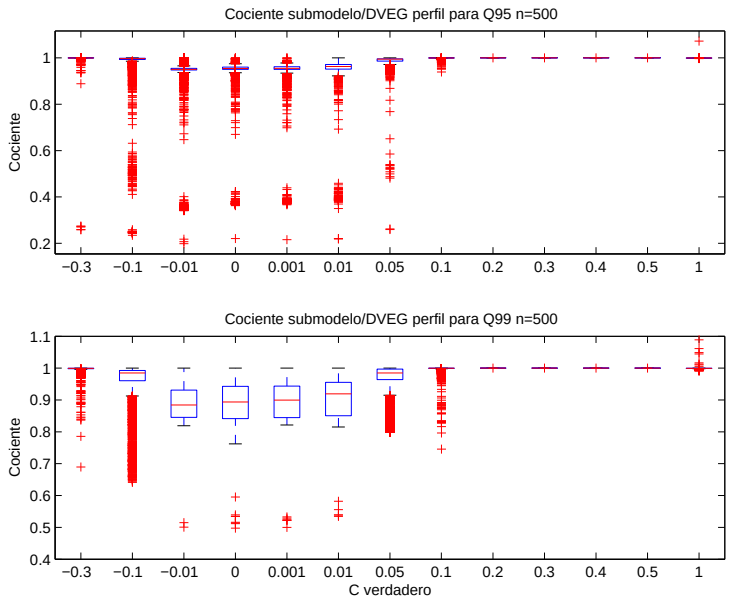


Figura 3.7: Cocientes de intervalos n=500

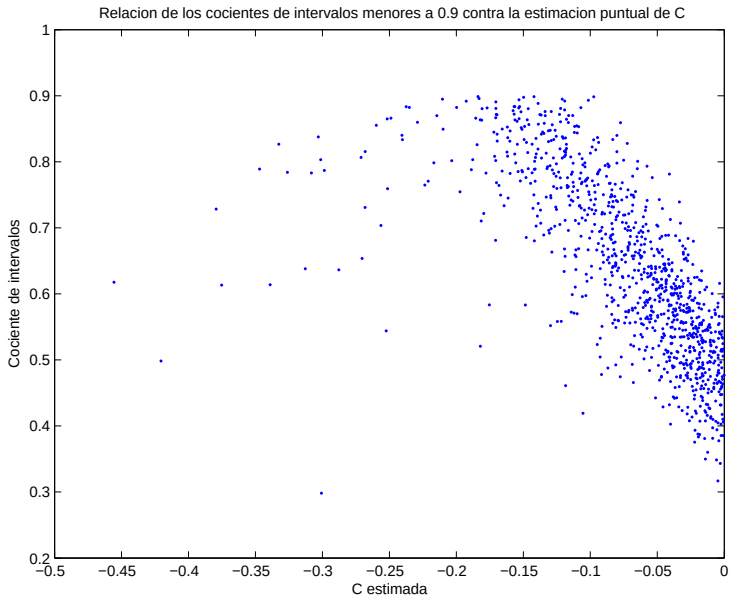


Figura 3.8: Cocientes de intervalos menores a 0.9 y su estimación puntual de c donde el valor verdadero de c es 0.1.

3.5. Comentarios finales

De las simulaciones realizadas obtuvimos que los intervalos de verosimilitud-confianza para los modelos no regulares considerados resultan ser más eficientes que los intervalos obtenidos a través del método delta en el sentido de que tienen una cobertura más cercana a la pretendida según los resultados asintóticos y que toman en cuenta la asimetría existente de la verosimilitud.

Además, la selección automática del submodelo sugerido resultante de $\hat{c}$, para los casos donde $|c| \leq 0.1$, conduce a intervalos de verosimilitud más angostos con coberturas razonables, obteniendo con esto estimaciones más precisas. En particular, el caso $c = 0.1$ con $n = 50$ nos

muestra que hay que tener cautela cuando tengamos como EMV de c un valor negativo cercano a cero, ya que podemos estar subestimando cuantiles grandes debido a la reducción del intervalo basados en el submodelo Weibull.

Por otro lado, para muestras de tamaño 50 las simulaciones muestran que se tienen coberturas razonables para Q_{95} basados en la función de verosimilitud perfil. Para Q_{99} se tienen coberturas buenas y los intervalos son más angostos que los obtenidos bajo el modelo DVEG.

Capítulo 4

Conclusiones

Mediante el uso de la función de verosimilitud perfil aplicada a la inferencia del parámetro c de la DVEG, se obtiene información relevante para sugerir submodelos apropiados.

En los submodelos Weibull y Fréchet es más fácil implementar restricciones con respecto al parámetro umbral, en contraste a la DVEG, cuyo soporte lo delimita la combinación de sus parámetros.

A través de restricciones adecuadas para el parámetro umbral se abre la posibilidad de considerar modelos más sencillos que no resultan evidentes al modelar directamente con la DVEG, como por ejemplo el modelo Fréchet de dos parámetros.

Cuando la función de verosimilitud hace plausibles a las tres familias, la estimación de los cuantiles de los submodelos de la DVEG (sin restricciones del parámetro umbral en el caso Fréchet y Weibull) nos lleva a intervalos de verosimilitud más angostos en comparación a la DVEG, con una cobertura razonable.

En estos casos, los intervalos de verosimilitud de los cuantiles de la DVEG contemplan incorrectamente a los tres submodelos en cuestión.

En general, los intervalos de verosimilitud son mejores que los intervalos calculados a través del método delta, debido a la asimetría en la superficie de verosimilitud y a la mejor cobertura.

La caída en la cobertura de los intervalos de verosimilitud de los modelos Fréchet cercanos a cero para Q_{99} con un tamaño de muestra de 50 máximos, en especial el caso c=0.1, indican la precaución que hay que tener al estimar cuantiles grandes cerca de $c = 0$ debido a la confusión de la familia límite.

Conjuntando los comentarios aquí presentes, se recomienda llevar a cabo la selección de un submodelo de la DVEG a través del proceso de modelación propuesto debido a una mejor representatividad del fenómeno en cuestión y/o a la ganancia en la longitud del intervalo de verosimilitud, manteniendo una cobertura razonable. Se recomienda el uso de la función de verosimilitud perfil relativa para fines de estimación, pues proporciona información relevante respecto a las cantidades de interés.

La metodología propuesta condujo a modelos adecuados aún con muestras de 40 máximos. Si el ajuste de una de las DVE o la DVEG misma no resulta adecuado, puede deberse tanto al tamaño de muestra chico como a la violación de los supuestos referentes a la distribución límite del máximo.

Bibliografía

[1] Box, G. E. P. (1980). *Sampling and Bayes' Inference in Scientific Modelling and Robustness.* Journal of the Royal Statistic Society, Series A, pp. 383-430.

[2] Box, G. E. P.; D. R. Cox. (1964). *An Analysis of Transformations.* Journal of the American Statistical Association, Vol. 77, No. 377.

[3] Castillo, E.; A.S. Hadi, N. Balakrishnan, J.M. Sarabia (2005). *Extreme Value and Related Models with Applications in Engineering and Science.* John Wiley & Sons, INC.

[4] Castillo, E. (1988). *Extreme Value Theory in Engineering.* Boston, Mass.: Academic Press.

[5] Cheng, R. C. H.; T. C. Iles (1990). *Embeeded Models in Three-Parameter Distributions and Their Estimation.* Journal of the Royal Statistical Society. Series B, Vol. 52, No. 1, pp. 135-149.

[6] Coles, S. (2004). *An Introduction to Statistical Modeling of Extreme Values.* Springer-Verlag.

[7] D'Agostino, R. B.; M. A. Stephens. (1986) *Goodness of Fit Techniques.* New York; Marcel Dekker, INC.

[8] Embrechts, P.; C. Klüppelberg, T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag.

[9] Finkenstädt, B.; H. Rootzén (2004). *Extreme Value in Finance, Telecomunications, and the Environment.* Chapman & Hall.

[10] Gutierrez, A. (2004). *Estudio del manejo de aguas pluviales en la zona metropolitana de Morelia (ZMM), Estado de Michoacán.* Instituto Mexicano de Tecnología del Agua, pp. 29.

[11] Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data.* New Jersey: John Wiley & Sons.

[12] Leadbetter, M. R.; G. Lindgren; H Rootzén (1983).*Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag.

[13] Meeker, W. Q.; L. Escobar (1998). *Statistical Methods for Reliability Data.* John Wiley & Sons, INC.

[14] Pawitan, Y. (2004). *In all likelihood.* Oxford University Press.

[15] Pérez Miranda, M. (2005). *Valores Extremos Aplicados a Datos Ambientales*. Tesis de Maestría no publicada. Programa de Maestría en Ciencias con especialidad en Probabilidad y Estadística, CIMAT.

[16] Smith, R. (1985). *Maximum Likelihood Estimation in a Class of Nonregular Cases*. Biometrika, Vol. 72, No. 1, pp. 67-90.

[17] Sprott, D. A. (2000). *Statistical Inference in Science*. Springer-Verlag.

[18] Resnick, S. (1987). *Extreme Value, Regular Variation and Point Process*. Springer-Verlag.