

Estimación de Extremos en el dominio de atracción de la Distribución Gumbel

POR

Víctor Fabio Lazarte

Tesis para alcanzar el grado de Doctor en Estadística.

Facultad de Ciencias Económicas - Facultad de Ciencia Exactas y
Tecnología



Universidad Nacional de Tucumán

Director: Dr. Aldo José Viollaz

Comisión de Supervisión:

- Dra. Ana María Sfer – Prof. Titular. Facultad de Ciencias Exactas y Tecnología. UNT.
- Dra. Viviana Beatriz Lencina – Prof. Titular. Facultad de Ciencias Económicas. UNT.

San Miguel de Tucumán – Diciembre de 2017

Dedicada a mi padre: Ernesto Hugo Lazarte.

*El mayor y mejor ejemplo de esfuerzo, trabajo y dedicación que he tenido
en toda mi vida.*

INDICE DE CONTENIDOS

Resumen:.....	3
Introducción:.....	5
Capítulo 1: Distribución de valores extremos.....	11
1.1: Introducción.....	12
1.2: Distribución del máximo o mínimo de una muestra.....	15
1.3: Distribuciones límites del máximo.....	25
1.4: Máximo Dominio de Atracción.....	31
Capítulo 2: Aplicación de la transformada Box-Cox para aproximar una distribución Gumbel.....	41
2.1: Introducción.....	42
2.2: La distribución Gumbel.....	45
2.3: Aplicación de la transformación Box Cox.....	48
Capítulo 3: Distribución Gumbel Ampliada.....	53
3.1. Introducción.....	54
3.2. Distribución Potencial Gumbel.....	55
3.3. Aproximaciones a las familias Normal, Lognormal, Weibull y Gamma.....	63
Capítulo 4: Estimación de parámetros y percentiles superiores.....	65
4.1. Introducción y ejemplo.....	66
4.2. Estimación de parámetros en el modelo Gumbel.....	72
4.3. Estimación de parámetros en el modelo Potencial Gumbel.....	75
4.4: Aplicación a las series de vientos máximos de Argentina.....	79
Capítulo 5: Mapa de vientos de Argentina.....	91
5.1: Introducción.....	92
5.2: Estimación de parámetros y percentiles.....	98
5.3: Trazado de isolíneas.....	102
Conclusiones:	105
Referencias:	108
Bibliografía:	109

Resumen

En la teoría de valores extremos existe un resultado central que afirma que bajo ciertas condiciones, el máximo de variables aleatorias independientes e idénticamente distribuidas, convenientemente normalizado, converge sólo a tres tipos de distribuciones llamadas Distribuciones de Extremos. Estas distribuciones límites son: la Weibull, la Gumbel y la Fréchet. En las aplicaciones, en general sólo dos de estas distribuciones asintóticas, la Gumbel y la Fréchet, proporcionan modelos apropiados para la estimación de sucesos extremos. A su vez, es muy recomendada la distribución Fréchet por tratarse de una distribución de colas pesadas que usualmente conduce a estimaciones muy conservadoras de los percentiles superiores. Sin embargo es claro que una ganancia en seguridad puede causar una pérdida en precisión, por lo que vale la pena estudiar casos en los que se pueda ganar en precisión sin descuidar la seguridad.

Por otro lado, se sabe que el dominio de atracción de la distribución Gumbel, contiene distribuciones de colas que van desde moderadamente pesadas hasta muy livianas, y es por esta riqueza que se considera ventajoso proponer modelos de estimación de extremos que permitan ajustar distribuciones dentro de este dominio de atracción.

En el presente trabajo se define el modelo Potencial-Gumbel que surge de la aplicación de la transformación de Box-Cox y que permite estimar distribuciones dentro del dominio Gumbel. El modelo Potencial-Gumbel es una familia paramétrica de distribuciones que pertenecen al dominio Gumbel dado que posee un parámetro de forma que permite variar el peso de la cola de la distribución sin salirse del mismo.

En el ajuste del modelo Potencial-Gumbel a los datos de series de vientos máximos mensuales de Argentina puede verse que en cada estación meteorológica hay un comportamiento diferente en cuanto al peso en las colas derechas de las distribuciones que es captado por el modelo. Esto permite estimar con mayor precisión los percentiles superiores de cada estación, y en caso que se quiera uniformizar el modelo asignando un único parámetro de forma a todas las estaciones, se puede adoptar un criterio conservador eligiendo de entre los valores estimados, un valor mínimo para λ . En las estimaciones realizadas se puede evaluar la pérdida de precisión que se da por obtener estimaciones más conservadoras de los percentiles.

En definitiva, el modelo Potencial-Gumbel propuesto en esta tesis tiene la capacidad de discriminar distribuciones dentro del dominio de atracción de la distribución Gumbel, lo que permitirá estimaciones más precisas de los percentiles superiores, y dada la riqueza en distribuciones del dominio Gumbel el modelo propuesto aquí servirá para resolver un sinfín de casos aplicados.

Introducción

El objetivo final de la teoría de valores extremos es encontrar y estimar modelos apropiados para la distribución de probabilidades del máximo, o mínimo, de una sucesión de variables aleatorias con el fin de obtener información acerca de las probabilidades de ocurrencia de ciertos sucesos extremos y también determinar los periodos de recurrencia estimados de tales sucesos. Tal información será de gran valor para la planificación de acciones futuras como ser la prevención de catástrofes naturales, la construcción de obras civiles, cálculo de costos y réditos, etc.

El resultado principal de esta teoría afirma que, bajo ciertas condiciones, la distribución de probabilidades del máximo (o mínimo), convenientemente estandarizada, de una sucesión de variables aleatorias independientes e idénticamente distribuidas (iid), converge para n suficientemente grande solo a ciertas distribuciones especiales, una de las cuales es de especial interés en este trabajo y se conoce como la Distribución Gumbel. En otras palabras: la distribución de probabilidades del máximo (o mínimo) de una sucesión de variables aleatorias iid, cuando converge, lo hace solo a tres distribuciones posibles, estas son La distribución Weibull, la distribución Gumbel y la distribución Fréchet.

Esencialmente, la Teoría de Valores Extremos se ocupa de la predicción de sucesos raros que lógicamente estarán fuera del rango de datos, es decir, esta teoría se basa en la extrapolación de información, a pesar de ello científicos muy reconocidos (como ser Richard Smith y Jonathan Tawn) consideran que esta teoría hace el mejor uso que se le puede dar a la información disponible en los datos para resolver el problema de la ocurrencia de sucesos extremos.

En su versión más sencilla, el problema es el siguiente: Dada una muestra aleatoria X_1, X_2, \dots, X_n de una variable aleatoria X con función de distribución acumulada desconocida F , se desea estimar, en el caso del máximo, la cola derecha de la distribución, es decir $\bar{F}(x) = 1 - F(x)$ para valores muy grandes de x y por lo tanto, tendrán mayor importancia los valores grandes observados.

de obras, y este fue el problema que motivo este y otros trabajos del autor, dado que el Dr. Aldo J. Viollaz (Director) fue integrante de una de las comisiones para el diseño de las normas civiles del (CIRSOC) Centro de Investigación de los Reglamentos Nacionales de Seguridad para las Obras Civiles. Este caso aplicado a datos reales de las velocidades máximas del viento será tratado en varias secciones de este trabajo.

En la Argentina se registran las velocidades máximas mensuales de los vientos, en nudos, en cerca de 120 estaciones meteorológicas desde hace muchos años. Los datos son registrados por el Servicio Meteorológico Nacional y será de gran valor para la ingeniería y otras ciencias, diseñar y estudiar métodos y procedimientos que permitan estimar con razonable precisión y seguridad los periodos de recurrencia de los vientos más destructivos.

Los registros de las velocidades máximas de vientos con que se cuenta lamentablemente en muchos casos no cuentan con la precisión aceptable dado que se usaron diferentes métodos de medición no siempre confiables, en este trabajo para la estimación del modelo propuesto se seleccionaron 20 estaciones con datos confiables, pero para la confección del "mapa de vientos de Argentina" se usó un modelo más sencillo (Distribución Gumbel) y todos los datos disponibles con el fin de cubrir la máxima región. Como muestra, se presenta a continuación el gráfico de las velocidades máximas mensuales del viento en la estación meteorológica Aeroparque durante el periodo 1968-1991.

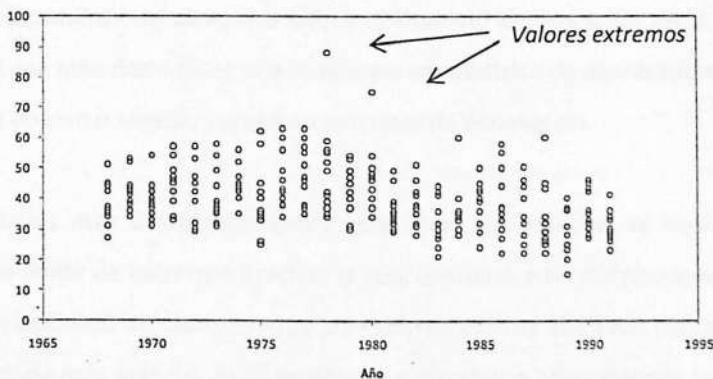


Figura 1: Velocidades máximas mensuales del viento, en nudos, en la Estación Meteorológica Aeroparque. Buenos Aires. Argentina. En el periodo 1968 – 1991.

Como ya se mencionó antes, bajo ciertas condiciones, la distribución del máximo converge a tres distribuciones límites lo cual significa que existen tres familias de distribuciones, llamadas dominios de atracción, que convergen respectivamente una a cada una de las tres distribuciones límites. Por lo tanto en la estimación de la distribución del máximo una propuesta es utilizar como modelo de ajuste a una de estas tres distribuciones asintóticas según el dominio en el que se encuentre, es decir, si la variable que genera los datos está en el dominio de la distribución Gumbel se usa como modelo para el máximo directamente esta distribución. Sin embargo, también se ha demostrado que la convergencia a las distribuciones asintóticas puede ser muy lenta en el caso que aquí se trata (el caso Gumbel), y al utilizar la distribución límite como modelo de ajuste puede conducir a sesgos en las estimaciones, por ejemplo, se podría estar cayendo en una subestimación de los percentiles superiores de la distribución real.

En el presente trabajo se utiliza la conocida transformación de Box-Cox (con cierta modificación adecuada) para mejorar la aproximación de los datos al modelo teórico, es decir el correspondiente a la Distribución Gumbel, también se muestra que la aplicación de la transformación mencionada en este sentido, conduce de un modo muy natural a definir un modelo más general que el modelo Gumbel (ya que se adiciona un parámetro de forma) para obtener mejores ajustes a los datos especialmente en las colas. Esta nueva familia paramétrica de distribuciones será llamada "Distribución Potencial-Gumbel" y es más rica que la Distribución Gumbel para aproximar distribuciones en su dominio, obviamente no alcanza a cubrir el Dominio de Atracción de la Distribución Gumbel ya que este dominio es una familia no paramétrica de distribuciones, pero si lo representa en cierto sentido específico con fines de estimación.

En los trabajos más destacados sobre este tema los estudios se vuelven hacia la distribución límite de extremos Fréchet la cual contiene a las distribuciones conocidas como "distribuciones de colas pesadas", la razón de esto es que estas distribuciones son de algún modo más seguras en el sentido que tienden a sobreestimar los percentiles superiores. Es razonable pensar que esta ganancia en seguridad representa una pérdida de precisión en las estimaciones, y en algunos casos aplicados puede traducirse en

pérdida económica innecesaria. Lo que se afirma aquí es que tal vez no es siempre conveniente dejar de lado a las distribuciones dentro del dominio de atracción de la distribución Gumbel, pues es muy conocido que en este dominio existen distribuciones con colas que van desde moderadamente pesadas hasta muy livianas, incluyendo distribuciones con dominio finito. Esta variedad en las colas de las distribuciones del dominio Gumbel posiblemente permitirá ganar precisión en la estimación de valores extremos. Ahora, si se trata de estimar dentro de este dominio, se presenta el problema de tratar con una familia no paramétrica de distribuciones, y es por ello que el modelo Potencial-Gumbel constituye una familia paramétrica (de tres parámetros) dentro del dominio Gumbel que lo representa dado que haciendo variar el parámetro de forma ($\lambda > 0$) obtenemos distribuciones con colas que van desde moderadamente pesadas hasta muy livianas. Es decir, que dada una distribución dentro del dominio Gumbel se ganará mucho en precisión estimando con este modelo que está contenido en el dominio Gumbel.

Por otro lado, la aplicación del modelo Potencial-Gumbel a los datos (velocidades máximas del viento) mostró que en muchas estaciones el modelo propuesto ajusta muy bien (la bondad del ajuste en este trabajo se realiza mediante la observación de un gráfico Q-Qplot, pero puede aplicarse cualquier otro criterio de bondad del ajuste). También se observó que la estimación del parámetro de forma (y por lo tanto de los percentiles) tiene una gran dependencia de los valores extremos del conjunto de datos, principalmente del valor máximo. Por esta razón será de gran importancia tener información sobre la confiabilidad de los datos extremos, lamentablemente, en algunos casos el valor máximo observado puede estar subestimado debido a que los aparatos de medición tengan un rango limitado y no alcancen a medir valores más allá de cierto límite. En el ajuste a los datos de vientos se encontraron algunos casos sospechosos en los que parece que el valor máximo de la muestra no está en concordancia con los demás datos (parece subestimado) y entonces no queda más opción que estimar del modo más seguro, es decir, tomar el valor del parámetro que tienda a sobreestimar y no a subestimar los valores extremos de la distribución.

Capítulo 1

Distribución de Valores Extremos

En este capítulo se presentan los conceptos básicos de la teoría de valores extremos, necesarios para establecer los resultados de la tesis. Estos resultados corresponden a la teoría asintótica que establece las condiciones de convergencia del máximo o mínimo a una de las tres distribuciones de extremos, también se hace referencia a la propiedad de las distribuciones extremales de ser estables para el máximo, al concepto de distribuciones de colas equivalentes y de máximo dominio de atracción de las distribuciones de extremos mediante las condiciones de Von Mises, máximo dominio de atracción de la distribución Gumbel y su caracterización.

1.1: Introducción

En este capítulo se presentan los resultados de la teoría de valores extremos, necesarios para fundamentar gran parte de este trabajo. El resultado más importante de esta teoría es el Teorema de Fisher-Tippett que, en términos generales, establece que la distribución del máximo de una sucesión de variables independientes e idénticamente distribuidas (iid), convenientemente estandarizado, si converge, lo hace a una de tres distribuciones especiales. Estas son las llamadas distribuciones de valores extremos, a saber: La distribución Weibull, la distribución Gumbel y la distribución Fréchet.

Las condiciones necesarias para la convergencia de una sucesión de variables aleatorias con cierta distribución a una de las tres distribuciones de valores extremos, están referidas a la forma de las colas de la distribución. En la terminología de valores extremos y sin ser riguroso se puede decir que: A la distribución Weibull solo convergen distribuciones de "colas livianas"; a la distribución de Fréchet solo convergen distribuciones de "colas pesadas" y a la distribución de Gumbel convergen variables de colas de diferente tipo, desde colas livianas hasta colas moderadamente pesadas. Una vez establecidas las condiciones de convergencia de las variables, es posible identificar los tipos de familias de variables aleatorias que convergen a cada una de estas distribuciones. Estos tipos de familias de variables aleatorias son llamados Máximo Dominio de Atracción de la distribución de valores extremos correspondiente.

En este trabajo se está especialmente interesado en el Máximo Dominio de Atracción de la Distribución Gumbel el cual es una familia muy rica de distribuciones en el sentido de contener distintos tipos de distribuciones categorizadas según sus colas, por ejemplo: Distribución Normal (cola liviana), Exponencial (cola liviana), Log-normal (cola moderadamente pesada) y otras. El dominio de atracción de la distribución Gumbel, como se verá en este capítulo, es una familia no paramétrica de distribuciones y el resultado más importante de este trabajo es que este dominio de atracción puede ser representado, a fines de estimación, por una familia semi-paramétrica de distribuciones que se logra ampliando la familia Gumbel mediante una transformación potencial y la noción de distribuciones de cola equivalentes.

Los símbolos O , o y \sim

A continuación se definen los símbolos de O , o y \sim Landau, que serán necesarios en el desarrollo de la teoría.

Sean f y φ dos funciones definidas sobre el conjunto de Números Naturales y suponga que φ es positiva y siempre creciente o siempre decreciente. En la mayoría de los casos φ tendrá una forma muy sencilla, por ejemplo: 1 , n , n^p , $1/n$, etc.

Definición 1.1.1

- (1) Si existe una constante k y un número natural n_0 tal que $|f(n)| \leq k\varphi(n)$, para $n \geq n_0$ entonces se escribe $f(n) = O(\varphi(n))$ y se dice que f es a lo sumo del orden de φ .
- (2) Si $\lim_{n \rightarrow \infty} f(n)/\varphi(n) = 0$ entonces se escribe $f(n) = o(\varphi(n))$ y se dice que f es de menor orden que φ .
- (3) Si $\lim_{n \rightarrow \infty} f(n)/\varphi(n) = l$, con $l \neq 0$ entonces se escribe $f(n) \sim l\varphi(n)$ y se dice que f es asintóticamente equivalente a $l\varphi$.

Ejemplos

$$2n^2 + 10n + 5 = O(n^2); \quad n = o(n^p) \text{ con } 1 < p < \infty; \quad \frac{n+10}{2} \sim \frac{n}{2}.$$

Tienen las siguientes propiedades:

- 1) $O(\varphi) + O(\psi) = O(\varphi + \psi)$.
- 2) $O(\varphi)O(\psi) = O(\varphi\psi)$.
- 3) $O(\varphi)o(\psi) = o(\varphi\psi)$.
- 4) Si $f \sim \varphi \Rightarrow f + o(\varphi) \sim \varphi$.

Convergencia de tipos de distribuciones

El siguiente concepto se necesitará para precisar las ideas sobre convergencia del máximo de variables aleatorias independientes e idénticamente distribuidas.

Se acuerda la siguiente notación, dadas dos variables aleatorias no degeneradas X e Y ,

si estas tienen la misma distribución de probabilidades escribiremos $X \stackrel{d}{=} Y$.

Definición 1.1.2

Se dice que las variables aleatorias X e Y y sus correspondientes distribuciones y densidades están en el mismo tipo o son del mismo tipo si es que existen constantes

a y $b \in \mathbb{R}$, $b > 0$ tales que $X \stackrel{d}{=} bY + a$

El siguiente teorema establece que la distribución límite de una sucesión de variables aleatorias está unívocamente determinada salvo cambios de posición y escala.

Teorema 1.1.1

Sean X, Y, X_1, X_2, \dots variables aleatorias y $a_n, b_n, \alpha_n, \beta_n \in \mathbb{R}$, $b_n > 0$ y $\beta_n > 0$, si se cumple que:

$$b_n^{-1}(X_n - a_n) \stackrel{d}{\rightarrow} X$$

Entonces se cumple también que:

$$\beta_n^{-1}(X_n - \alpha_n) \stackrel{d}{\rightarrow} Y,$$

si y solamente si:

$$\lim_{n \rightarrow \infty} \frac{b_n}{\beta_n} = b \geq 0 \quad \text{y} \quad \lim_{n \rightarrow \infty} \frac{a_n - \alpha_n}{\beta_n} = a \in \mathbb{R}.$$

Y en tal caso

$$Y \stackrel{d}{=} bX + a$$

Donde a y b son las únicas constantes que cumplen con esto.

1.2: Distribución del Máximo o Mínimo de una Muestra

De forma análoga al estudio de la suma de n variables aleatorias, se puede estudiar el máximo (o mínimo) de n variables aleatorias.

Considere una sucesión de variables aleatorias X_1, \dots, X_n iid como X , con función de distribución acumulada F no degenerada. Llámese M_n al máximo de las n primeras variables aleatorias, es decir:

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

Observe que el mínimo de las n primeras variables aleatorias se puede escribir en función del máximo que ya está definido, dado que:

$$\min\{X_1, \dots, X_n\} = -\max\{-X_1, \dots, -X_n\}.$$

Entonces, desde ahora sólo se estudiará la Distribución del Máximo de las n primeras variables aleatorias dado que gracias a esta última relación, no es difícil adaptar los resultados al caso de la Distribución del Mínimo.

Si la función de distribución acumulada F de la muestra es conocida, no hay dificultad en determinar la distribución del máximo, como se muestra a continuación:

Sea $x \in \mathbb{R}$, $n \in \mathbb{N}$,

$$F_{M_n}(x) = P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = F^n(x)$$

Y si F es absolutamente continua con densidad f entonces el máximo tendrá función de densidad dada por:

$$f_{M_n}(x) = n[F(x)]^{n-1}f(x).$$

De manera que, en principio si conocemos F conocemos la distribución exacta del máximo. Sin embargo, ocurre que en general las expresiones analíticas para F_{M_n} pueden ser muy complicadas y por otro lado, lo que es más probable es que en los casos aplicados F es desconocida y es necesario utilizar modelos estimados o aproximados.

Una posible solución, propuesta en la bibliografía, es tratar de conocer la forma asintótica de la distribución del máximo, para tener una distribución límite que sirva de aproximación a F_n . Esto es análogo al caso de la Distribución Normal que, bajo ciertas condiciones, sirve de aproximación a la distribución de la suma de n variables aleatorias independientes para n suficientemente grande.

Ejemplo 1.2.1:

Sea X variable aleatoria con distribución exponencial con parámetro $\lambda = 1$.

$$F_{M_n}(x) = (1 - e^{-x})^n \Rightarrow f(x) = n(1 - e^{-x})^{n-1}e^{-x}.$$

En el gráfico siguiente se muestran las curvas de las densidades del máximo para diferentes valores de n , en él se observa como la distribución se va desplazando, a medida que n crece, hacia los valores grandes del recorrido de la variable. Esto muestra claramente que la "forma" de la distribución del máximo dependerá de los valores grandes del recorrido de X , o próximos al extremo derecho de la distribución.

Más adelante se verá que este es un caso de distribución que converge al Modelo Gumbel.

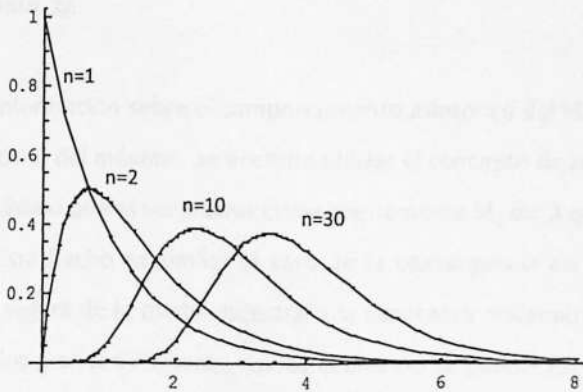


Figura 1.2.1: Densidades de las distribuciones del máximo de variables exponenciales para muestras de tamaño: 1, 2, 10 y 30.

Definición 1.2.1: Llamemos *punto final derecho* x_f de F al valor supremo del recorrido de X , es decir:

$$x_f = \sup\{x \in \mathbb{R} / F(x) < 1\}$$

Note que el punto x_f puede ser finito o infinito.

Los valores extremos de la variable aleatoria X son los valores "cercaños" a x_f por lo tanto su ley de probabilidades dependerá esencialmente de la forma de la cola derecha de la función de distribución F .

Además, se observa que:

- i) Si $x < x_f$ entonces $F(x) < 1$, luego $P(M_n \leq x) = F^n(x) \rightarrow 0$ si $n \rightarrow \infty$.
- ii) Si $x_f < \infty$ y $x \geq x_f$, $P(M_n \leq x) = F^n(x) = 1$ para cualquier n .

Así, se tiene que M_n converge en probabilidad a x_f cuando n tiende a ∞ , y como la sucesión M_n es no decreciente, se da la convergencia casi segura a x_f . Este hecho conlleva a que no sea posible ver la forma de la distribución límite dado que esta degenera en el punto x_f .

Par obtener más información sobre el comportamiento asintótico del M_n es decir, para ver la forma asintótica del máximo, se necesita utilizar el concepto de convergencia en distribución del máximo que al normalizar convenientemente M_n evita que se degenera en el punto x_f . Este hecho es similar al caso de la convergencia en probabilidad y convergencia casi segura de la media muestral a la esperanza matemática en las leyes débil y fuerte de los grandes números, en los cuales no se puede ver la forma de la distribución límite, mientras que en el Teorema Central del Límite que utiliza convergencia en distribución si se puede ver la forma asintótica de la distribución de la media muestral.

Distribución Límite del Máximo

En este punto será necesario mostrar cómo, bajo ciertas condiciones sobre la función de distribución acumulada F , es posible determinar la distribución límite de una adecuada "función lineal" del máximo M_n . Se comienza presentando el siguiente ejemplo:

Ejemplo 1.2.2: Sea X una variable aleatoria, con función de distribución F con cola derecha dada por:

$$\bar{F}(x) = 1 - F(x) \sim \frac{c}{x} \quad \text{para } x \rightarrow \infty, \quad c > 0.$$

Y sea X_1, \dots, X_n una muestra aleatoria de X , entonces, para $x > 0$.

$$\begin{aligned} P\left(\frac{M_n}{nc} \leq x\right) &= F_{M_n}(ncx) = F^n(ncx) \\ &\sim \left(1 - \frac{1}{nx}\right)^n \rightarrow e^{-\frac{1}{x}} \quad \text{para } n \rightarrow \infty. \end{aligned}$$

Significa que $e^{-\frac{1}{x}} = \exp(-x^{-1})$ es la distribución límite a la que converge en distribución el máximo normalizado de esta forma $\frac{M_n}{nc}$.

También se puede escribir este resultado generalizado para distribuciones con colas que decrecen potencialmente.

Proposición 1.2.1:

Sea la variable aleatoria X con función de distribución F tal que para algún $\alpha > 0$,

$$\lim_{n \rightarrow \infty} n^\alpha [1 - F(x)] = c, \quad c > 0$$

Y sea X_1, \dots, X_n una muestra aleatoria de X . Entonces el máximo normalizado $\frac{M_n}{(nc)^{1/\alpha}}$

converge débilmente a la una distribución límite dada por:

$$\Phi_\alpha(x) = \begin{cases} \exp(-x^{-\alpha}), & x > 0 \\ 0, & x \leq 0 \end{cases} \quad \alpha > 0$$

Demostración:

Por hipótesis tenemos que $x^\alpha[1 - F(x)] = c - o(1)$ para $x \rightarrow \infty$,

es decir

$$1 - F(x) = \frac{c + o(1)}{x^\alpha} \text{ para } x \rightarrow \infty,$$

luego

$$P\left(\frac{M_n}{(nc)^{1/\alpha}} \leq x\right) = F^n\left[(cn)^{1/\alpha} x\right] = \left[1 - \frac{c + o(1)}{cnx^\alpha}\right]^n \rightarrow \exp(-x^{-\alpha}),$$

para $(cn)^{1/\alpha} x = a_n x \rightarrow \infty$ con n . ■

Diferentes suposiciones para la cola de F conducirán a diferentes distribuciones límites, por ejemplo si la variable aleatoria X tiene soporte finito (es decir X es acotada), entonces para n grande M_n se acercará a x_f , pero dependiendo de la forma de F cerca de x_f existirá o no una distribución límite no degenerada para M_n .

Proposición 1.2.2:

Sea X una variable aleatoria con distribución F tal que para algún $\alpha > 0$,

$$\lim_{x \rightarrow x_f} (x_f - x)^{-\alpha} [1 - F(x)] = c, \quad c > 0,$$

entonces la distribución del máximo normalizado $(cn)^{1/\alpha}(M_n - x_f)$ converge débilmente a:

$$\Psi_\alpha(x) = \begin{cases} \exp\{-|x|^\alpha\} & x < 0 \\ 1, & x \geq 0 \end{cases} \quad \alpha > 0$$

Demostración:

Sea $x < 0$, por hipótesis tenemos que: $(x_f - x)^{-\alpha} [1 - F(x)] = c + o(1)$ para $x \rightarrow x_f$.

Lo que se puede escribir como:

$$1 - F(x) = \frac{c + o(1)}{(x_f - x)^{-\alpha}} \text{ para } x \rightarrow x_f$$

y entonces,

$$P\left(\frac{M_n - x_f}{(cn)^{-\frac{1}{\alpha}}} \leq x\right) = \left\{F\left[(cn)^{-\frac{1}{\alpha}}x + x_f\right]\right\}^n$$

$$= \left\{1 - \frac{c + o(1)}{cn(-x)^{-\alpha}}\right\} \rightarrow \exp(-|x|^\alpha) \text{ para } x \rightarrow x_f. \blacksquare$$

Hasta aquí solo se mostraron distribuciones con colas que decaen potencialmente (serían las conocidas como distribuciones de colas pesadas), consideremos ahora una distribución con cola que decrece exponencialmente, es decir que cae con más fuerza.

Proposición 1.2.3:

Sea X v.a. tal que para algún $c > 0$,

$$\lim_{x \rightarrow \infty} e^x [1 - F(x)] = c,$$

entonces

$$\lim_{n \rightarrow \infty} P[M_n - \ln(cn) \leq x] = \exp(-e^{-x}).$$

Demostración:

Por hipótesis

$$1 - F(x) = \frac{c + o(1)}{e^x}$$

$$P(M_n - \ln(cn) \leq x) = [F(x + \ln(cn))]^n = \left(1 - \frac{c + o(1)}{cn e^x}\right)^n \rightarrow \exp(-e^{-x}). \blacksquare$$

Note que hasta ahora se ha tratado con la distribución límite del máximo "apropiadamente normalizado" sólo para casos en los cuales las condiciones sobre la cola de la distribución hacen que las "constantes normalizadoras" adecuadas sean

obvias pero este hecho no ocurrirá siempre y muchas veces la elección de estas sucesiones de constantes no será para nada directa.

En definitiva, para ver la forma de la Distribución Límite del Máximo se necesita estandarizar convenientemente a M_n , que es similar a lo que se hace en las diferentes versiones del teorema central del límite para estudiar la convergencia de la suma de variables a la distribución Normal. En este sentido, el análogo al Teorema Central del Límite en la teoría de extremos es el Teorema de Fisher-Tippett (1928) el cual afirma que si para apropiadas constantes normalizadoras $c_n > 0$ y $d_n \in \mathbb{R}$, se cumple que:

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H, \quad n \rightarrow \infty,$$

para alguna función de distribución acumulada no degenerada H , entonces H pertenece a sólo a uno de tres tipos de distribuciones llamadas *Distribuciones de Valores Extremos*.

Para estudiar esta convergencia se puede escribir la función de distribución acumulada del máximo como:

$$F_{M_n}(u_n) = P(M_n \leq u_n).$$

Donde $u_n = c_n x + d_n$.

Las condiciones para la convergencia de esta sucesión de funciones de distribución están relacionadas principalmente a la forma de la cola derecha de F . Esta es una de las principales diferencias con las leyes límites para las sumas de variables aleatorias en las que, las condiciones para la convergencia se establecen en términos de los momentos centrales de las distribuciones.

Es muy importante aclarar que existen muchas familias de distribuciones para las cuales no existe una distribución límite no degenerada, cualquiera sea la sucesión de constantes normalizadoras que se utilice. Los ejemplos para estos casos se mostrarán más adelante.

Para continuar se presenta un resultado básico necesario para establecer las leyes límites para los máximos. Este resultado es conocido como la *Aproximación de Poisson*.

Proposición 1.2.4: (Aproximación de Poisson)

Sea X una variable aleatoria con función de distribución F , sean X_1, \dots, X_n una muestra aleatoria de X , una sucesión u_1, u_2, \dots, u_n de números reales y un número fijo $\tau \in [0, \infty)$.

Entonces las siguientes expresiones son equivalentes:

$$n\bar{F}(u_n) \rightarrow \tau, \quad (1.2.1)$$

$$P(M_n \leq u_n) \rightarrow e^{-\tau}. \quad (1.2.2)$$

Demostración

Sea τ tal $0 \leq \tau < \infty$.

La expresión (1.2.1) se puede reescribir como $n\bar{F}(u_n) = \tau + o(1)$ o lo que es lo mismo,

$$\bar{F}(u_n) = \frac{\tau + o(1)}{n} \quad \text{o} \quad F(u_n) = 1 - \frac{\tau + o(1)}{n}$$

donde o es la función "o pequeña" de Landau.

Entonces, si (1.2.1) vale, tenemos que:

$$\begin{aligned} P(M_n \leq u_n) &= [F(u_n)]^n = (1 - \bar{F}(u_n))^n \\ &= \left(1 - \frac{\tau + o(1)}{n}\right)^n \rightarrow e^{-\tau} \quad \text{si } n \rightarrow \infty. \end{aligned}$$

Por otro lado, la condición (1.2.2) se puede escribir como $[1 - \bar{F}(u_n)]^n \rightarrow e^{-\tau}$ y si esto vale se cumple que $\bar{F}(u_n) \rightarrow 0$.

Si no fuese así existiría una subsucesión u_{n_k} tal que $\bar{F}(u_{n_k})$ se mantiene a una cierta distancia de cero, es decir, $\exists \varepsilon > 0$ tal que $\bar{F}(u_{n_k}) > \varepsilon$, para todo n_k mayor o igual que un cierto n_0 luego :

$$P(M_{n_k} \leq u_{n_k}) = [1 - \bar{F}(u_{n_k})]^{n_k} \leq [1 - \varepsilon]^{n_k} \rightarrow 0.$$

Lo que es una contradicción.

Ahora de (1.2.2) tenemos

$$[1 - \bar{F}(u_n)]^n \rightarrow e^{-\tau}$$

Tomando logaritmos

$$-n \ln(1 - \bar{F}(u_n)) \rightarrow \tau.$$

Utilizando el hecho de que $-\ln(1 - x) \sim x$ para $x \rightarrow 0$ se tiene que $n\bar{F}(u_n) = \tau + o(1)$ que implica 1.1. ■

Observe que este resultado sigue siendo válido para $\tau = \infty$.

Corolario

Sea X una variable aleatoria con recorrido finito, es decir $x_F < \infty$, y con la propiedad de que:

$$\bar{F}(x_{F-}) = F(x_F) - F(x_{F-}) > 0.$$

Esto significa que la distribución tiene una masa acumulada en x_F .

Entonces para cualquier sucesión (u_n) que cumple que $P(M_n \leq u_n) \rightarrow \rho$, se tiene que: $\rho = 0$ o $\rho = 1$.

Demostración

Sea $\rho: 0 \leq \rho \leq 1$ entonces existe $\tau: 0 \leq \tau \leq \infty$ tal que $\rho = \exp(-\tau)$ y sea u_n tal que: $P(M_n \leq u_n) \rightarrow \rho = e^{-\tau}$, entonces vale que $n\bar{F}(u_n) \rightarrow \tau$ para $n \rightarrow \infty$.

- Si $u_n < x_F$ para infinitos valores de n entonces $\bar{F}(u_n) \geq \bar{F}(x_{F-}) > 0$ para esos infinitos valores de n por lo que $n\bar{F}(u_n) \rightarrow \infty \Rightarrow \tau \rightarrow \infty$.
- Si $u_n \geq x_F$ a partir de un cierto n , se tiene que $n\bar{F}(u_n) = 0 = \tau$.

Por lo tanto $\rho = 0$ o $\rho = 1$.

Con este resultado se está probando que para una distribución de probabilidades que tiene una masa acumulada en x_F , la función de distribución acumulada $P(M_n \leq u_n)$ no converge a ninguna distribución no degenerada, cualquiera sea la sucesión de constantes normalizadoras (u_n) . Es decir, no existe una distribución del máximo no degenerada.

Se pueden escribir las condiciones para la existencia de la distribución límite.

Proposición 1.2.5

Sea la variable aleatoria X con función de distribución F y $x_F \leq \infty$, y sea $\tau \in (0, 1)$.

Entonces existe una sucesión (u_n) que satisface $n\bar{F}(u_n) \rightarrow \tau$ si y solo si:

$$\lim_{x \rightarrow x_F} \frac{\bar{F}(x)}{\bar{F}(x^-)} = 1. \quad (1.2.3)$$

Observación:

Si X toma solo valores enteros, la condición (1.2.3) se escribe como:

$$\frac{\bar{F}(n)}{\bar{F}(n-1)} \rightarrow 1, \quad \text{para } n \rightarrow \infty.$$

Este resultado se aplica para estudiar la existencia de la distribución límite en variables aleatorias discretas con recorrido infinito.

Ejemplo 1.2.3:

Sea X con Distribución Geométrica de parámetro p : $0 < p < 1$, luego su función de masa está dada por $p_X(x) = (1-p)^{x-1}p$, $x \in \mathbb{N}$, entonces:

$$\begin{aligned} \frac{\bar{F}(k)}{\bar{F}(k-1)} &= \frac{P(X > k)}{P(X > k-1)} = \frac{\sum_{r=k+1}^{\infty} (1-p)^{r-1}p}{\sum_{r=k}^{\infty} (1-p)^{r-1}p} \\ &= \frac{(1-p)^k + (1-p)^{k+1} + \dots}{(1-p)^{k-1} + (1-p)^k + \dots} = \frac{(1-p)[(1-p)^{k-1} + (1-p)^k + \dots]}{(1-p)^{k-1} + (1-p)^k + \dots} \\ &= (1-p) \in (0, 1). \end{aligned}$$

El límite no es 1 por lo que se puede concluir que no existe una distribución límite no degenerada para $P(M_n \leq u_n)$.

1.3: Distribuciones Límites del Máximo

En este punto se está interesado en obtener una caracterización de las posibles distribuciones límites del máximo de una sucesión de variables aleatorias X_1, X_2, \dots, X_n independientes e idénticamente distribuidas (iid), convenientemente normalizado. Este problema tiene una cierta analogía con Teorema Central del Límite en el cual, bajo ciertas condiciones, la distribución límite de la suma de n variables aleatorias es una distribución Normal, en este caso es importante recordar que la distribución Normal tiene la notable propiedad de ser suma-estable, es decir, de mantenerse en la suma de variables independientes, ya que al sumar variables aleatorias normales independientes se obtiene una variable también con distribución normal y además de ello, mediante un cambio de posición y escala se recupera la distribución de los sumandos.

Simbólicamente esto se escribe:

$$\forall X_1, X_2 \text{ iid como } X \text{ se cumple que } \forall c_1 > 0, \forall c_2 > 0, c_1 X_1 + c_2 X_2 \stackrel{d}{=} bX + a,$$

para apropiadas constantes $b > 0$ y a .

Observe que la distribución Normal no es la única con esta propiedad, como ejemplo, se puede mencionar a la distribución Cauchy que también es estable bajo la suma y por lo tanto es la distribución límite de la suma de otras distribuciones. Una caracterización de todas las distribuciones suma-estables simétricas utilizando la función característica puede encontrarse en referencia [2] Embrechts pág 71.

Lo análogo a este resultado para el problema de la distribución del máximo sería, determinar cuáles son las distribuciones que se mantienen al tomar el máximo de variables independientes y que son invariantes ante cambios de posición y escala. Así luego, estas serán las únicas posibles distribuciones límites del máximo de variables aleatorias independientes, convenientemente normalizado. A las distribuciones que cumplen con las condiciones mencionadas se las llama Distribuciones Max-estables. Veamos cuál es su definición formal.

Definición 1.3.1: (Distribuciones Max-estables)

Se dice que, una variable aleatoria no degenerada X tiene distribución Max-estable si y solo si satisface que el $\max\{X_1, \dots, X_n\}$ tiene la misma distribución de probabilidades que $c_n X + d_n$, para n variables aleatorias independientes e idénticamente distribuidas X_1, \dots, X_n y apropiadas constantes normalizadoras $c_n > 0$ y $d_n \in \mathbb{R}$, para todo $n \geq 2$.

Lo que se puede escribir como: $F^n(x) = F(c_n x + d_n)$.

Note que si X_1, \dots, X_n son independientes e idénticamente distribuidas con distribución max-estable, se tiene que:

$$\frac{M_n - d_n}{c_n} \stackrel{d}{=} X_1,$$

Y es obvio que cualquier distribución máx-estable es la distribución límite, no degenerado, del máximo de variables aleatorias iid, dado que por lo menos es límite de ella misma.

Ejemplo 1.3.1: La distribución Gumbel es máx-estable.

Sea X una variable aleatoria con función de distribución $F(x) = \exp\left(-\exp\left(-\frac{x-\xi}{\theta}\right)\right)$ para $-\infty < x < \infty$, y parámetros ξ y θ : $-\infty < \xi < \infty$ y $0 < \theta < \infty$.

Entonces ¿qué distribución tiene el máximo de n variables aleatorias iid como X ?

$$\begin{aligned} F_{M_n}(x) &= F^n(x) \\ &= \left[\exp\left(-\exp\left(-\frac{x-\xi}{\theta}\right)\right) \right]^n \\ &= \exp\left(-n \exp\left(-\frac{x-\xi}{\theta}\right)\right) = \exp\left(-n \exp\left(-\frac{x}{\theta}\right) \exp\left(\frac{\xi}{\theta}\right)\right) \\ &= \exp\left(-\exp\left(-\frac{x}{\theta}\right) \exp\left(\ln n + \frac{\xi}{\theta}\right)\right) = \exp\left(-\exp\left(-\frac{x - (\xi + \theta \ln n)}{\theta}\right)\right). \end{aligned}$$

$$= \exp\left(-\exp\left(-\frac{x - \xi^*}{\theta^*}\right)\right), \quad \text{donde } \theta^* = \theta \text{ y } \xi^* = \xi + \theta \ln n.$$

Entonces M_n tiene la misma distribución que la variable original salvo un cambio de posición y escala.

Es decir, si $Y = c_n X + d_n$ tenemos que:

$$F_Y(y) = \exp\left(-\exp\left(-\frac{y - d_n - \xi}{c_n \theta}\right)\right) = \exp\left(-\exp\left(-\frac{x - (d_n + c_n \xi)}{\theta c_n}\right)\right).$$

Entonces ¿Cuáles serían las apropiadas constantes normalizadoras?

$$\begin{cases} d_n - c_n \xi = \theta \ln n \\ c_n \theta = \theta \end{cases} \Rightarrow d_n = \theta \ln n \text{ y } c_n = 1.$$

Teorema 1.3.1: La clase de distribuciones máx-estables coincide con la clase de todas las posibles distribuciones límites, no degeneradas, del máximo normalizado de variables aleatorias iid.

Demostración

De la propia definición se tiene que si una distribución es max-estable, entonces es el límite de distribuciones pues al menos es el límite de ella misma.

Sean las constantes normalizadoras $c_n > 0$ y $d_n \in \mathbb{R}$ tales que para todo $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} F^n(c_n x + d_n) = H(x),$$

Para alguna función de distribución no degenerada H . Y sea $k \in \mathbb{N}$, entonces tenemos que:

$$H^k(x) = \left[\lim_{n \rightarrow \infty} F^n(c_n x + d_n) \right]^k = \lim_{n \rightarrow \infty} F^{nk}(c_n x + d_n),$$

pero por otro lado, también vale que:

$$\lim_{n \rightarrow \infty} F^{nk}(c_{nk}x + d_{nk}) = H(x).$$

Luego, por el Teorema 1.1.1 existen constantes $\tilde{c}_k > 0$ y $\tilde{d}_k \in \mathbb{R}$ tales que

$$\tilde{c}_k = \lim_{n \rightarrow \infty} \frac{c_{nk}}{c_n} \quad \text{y} \quad \tilde{d}_k = \lim_{n \rightarrow \infty} \frac{d_{nk} - d_n}{c_n}$$

Y tales que si Y_1, \dots, Y_k son iid con función de distribución H , se deberá cumplir que

$$\text{máx}\{Y_1, \dots, Y_k\} \stackrel{d}{=} \tilde{c}_k Y_1 + \tilde{d}_k$$

Teorema 1.3.2: (Fisher-Tippet)

Sea X_1, \dots, X_n una sucesión de variables aleatorias iid. Si existen constantes normalizadoras $c_n > 0$ y $d_n \in \mathbb{R}$ y alguna distribución no degenerada H tal que

$$\frac{M_n - d_n}{c_n} \stackrel{d}{\rightarrow} H,$$

entonces H pertenece a uno de los tres tipos de distribuciones siguientes:

$$\text{Fréchet:} \quad \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases} \quad \alpha > 0. \quad (1.3.1)$$

$$\text{Gumbel:} \quad G(x) = \exp(-e^{-x}), \quad x \in \mathbb{R} \quad (1.3.2)$$

$$\text{Weibull:} \quad \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\} & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0. \quad (1.3.3)$$

La demostración se basa también en el Teorema 1.1.1, pues este muestra de que manera surgen las tres distribuciones límites.

Esencialmente tenemos que, según la hipótesis, para todo $t > 0$ se cumple que:

$$F^{[nt]}(c_{[nt]}x + d_{[nt]}) \rightarrow H(x),$$

Donde los corchetes $[]$ denotan la función parte entera.

Pero también deberá cumplirse,

$$F^{[nt]}(c_n x + d_n) = [F^n(c_n x + d_n)]^{\frac{[nt]}{n}} \rightarrow H^t(x),$$

Luego por el Teorema 1.1.1 existen funciones $\beta(t) > 0$ y $\alpha(t) \in \mathbb{R}$ que satisfacen

$$\beta(t) = \lim_{n \rightarrow \infty} \frac{c_n}{c_{[nt]}} \quad \text{y} \quad \alpha(t) = \lim_{n \rightarrow \infty} \frac{d_n - d_{[nt]}}{c_{[nt]}}, \quad \text{con } t > 0,$$

y

$$H^t(x) = H[\beta(t)x + \alpha(t)].$$

De esta última ecuación se deduce que $\beta(st) = \beta(s)\beta(t)$ y $\alpha(st) = \beta(t)\alpha(s) + \alpha(t)$ y de las dos últimas ecuaciones funcionales surgen las tres distribuciones extremales. La demostración completa puede verse en Resnick [13]

La siguiente figura muestra las formas de las densidades de estas tres distribuciones, para la distribución Fréchet y Weibull se tomó $\alpha = 1$.

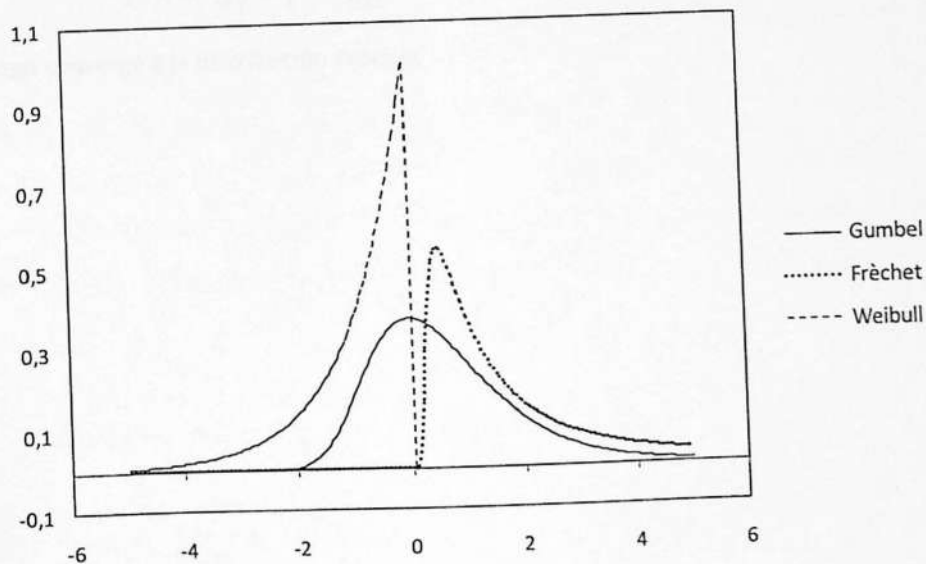


Figura 1.3.1: Densidades de las distribuciones de valores extremos estándares, se tomó $\alpha = 1$ para las distribuciones Fréchet y Weibull.

Ejemplo 1.3.2: Máximo de vs. as. Exponenciales.

Sean X_1, \dots, X_n iid con función de distribución $F(x) = 1 - e^{-x}$, $\bar{F}(x) = e^{-x}$.

$$P(M_n - \ln n \leq x) = P[X_1 \leq x + \ln n]^n = \left[1 - \frac{e^{-x}}{n}\right]^n \rightarrow \exp(-e^{-x}) = G(x)$$

Así converge a la distribución Gumbel.

Ejemplo 1.3.3: Máximo de vs. as. Cauchy.

Sean X_1, \dots, X_n iid con función de densidad

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}$$

sin dificultad se puede probar que

$$\bar{F}(x) \sim \frac{1}{\pi x}$$

luego por la Proposición 1.1 tenemos que:

$$P\left(M_n \leq \frac{nx}{\pi}\right) = \left[1 - \frac{1}{nx}\right]^n \rightarrow \exp(-x^{-1}) = \Phi_1(x), \quad x > 0.$$

Luego converge a la distribución Fréchet

1.4: Máximo Dominio de Atracción

En la sección anterior se dijo que la distribución del máximo de una sucesión de variables aleatorias iid, convenientemente normalizado, cuando converge, lo hace a una de sólo tres distribuciones posibles. También se puede preguntar ¿De qué manera se deben buscar o elegir las apropiadas constantes normalizadoras $c_n > 0$ y $d_n \in \mathbb{R}$ para que se cumpla que $c_n^{-1}(M_n - d_n) \xrightarrow{d} H$? Y si es posible cambiar la distribución límite del máximo cambiando las constantes normalizadoras. La respuesta a la segunda pregunta está en el hecho de que la convergencia a las distribuciones límites esta unívocamente determinada bajo transformaciones afines debido a la propiedad de ser máx-estables, sin embargo es posible mejorar la velocidad de convergencia cambiando las constantes normalizadoras. La respuesta a la primera pregunta se dará más adelante.

A continuación se buscará determinar las familias de distribuciones F tales que, el máximo normalizado de una sucesión de variables iid como F , converge a una determinada distribución de extremos. En palabras más sencillas, cuáles son las distribuciones cuyo máximo converge a la distribución Frechét, cuales convergen a la Gumbel y cuales a la Weibull.

A estas familias se las llama Máximo Dominio de Atracción de una distribución de extremos dada.

Definición 1.4.1: (Máximo Dominio de Atracción)

Se dice que la variable aleatoria X pertenece al máximo dominio de atracción de la distribución de valores extremos H , $MDA(H)$, (Donde H es una de las tres distribuciones de extremos), si es que existen constantes normalizadoras $c_n > 0$ y $d_n \in \mathbb{R}$ tales que:

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H$$

Se escribe $X \in MDA(H)$.

Proposición 1.4.1: Caracterización del máximo dominio de atracción.

La función de distribución F pertenece al MDA (H) con constantes normalizadoras $c_n > 0, d_n \in \mathbb{R}$ si y solo si

$$\lim_{n \rightarrow \infty} n\bar{F}(c_n x + d_n) = -\ln H(x), \quad x \in \mathbb{R}.$$

Es decir: $F \in \text{MDA}(H)$ si y solo si

$$P(M_n \leq c_n x + d_n) \rightarrow e^{-\tau} \text{ entonces } \tau = -\ln H(x) \text{ y así } n\bar{F}(c_n x + d_n) \rightarrow -\ln H(x).$$

Un concepto muy útil a la hora de especificar el máximo dominio de atracción de las distribuciones límites es el de "Distribuciones de colas equivalentes", este importante concepto también será utilizado en la descripción de la familia de distribuciones que son el objeto de este trabajo.

Definición 1.4.2: (Distribuciones de colas equivalentes)

Dos distribuciones F y G son de colas equivalentes si F y G tienen igual punto final derecho, esto es $x_F = x_G \leq \infty$ y para alguna constante real $c, 0 < c < \infty$, se cumple que

$$\lim_{x \rightarrow x_F^-} \frac{\bar{F}(x)}{\bar{G}(x)} = c.$$

Observaciones

- Si F y G son de colas equivalentes, $F \in \text{MDA}(H) \Leftrightarrow G \in \text{MDA}(H)$ lo que se deduce fácilmente de la caracterización de la Proposición 1.4.1.
- Si dos distribuciones son de colas equivalentes, entonces se pueden utilizar las mismas constantes normalizadoras para la convergencia del máximo.
- La determinación de las constantes normalizadoras puede llegar a ser muy complicada para algunos casos.

A continuación se presenta una descripción breve de los dominios de atracción correspondientes a las distribuciones Fréchet y Weibull, para luego tratar la que nos interesa especialmente, es decir, el caso Gumbel.

Máximo Dominio de Atracción de la Distribución Fréchet

Sea X con distribución Fréchet, con función de distribución:

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

Por Taylor

$$e^x = 1 + x + \frac{x^2}{2} + \dots$$

Así

$$\exp(-x^{-\alpha}) = 1 - x^{-\alpha} + \frac{x^{-2\alpha}}{2} \dots$$

por lo que la cola derecha,

$$\bar{F}(x) = 1 - \exp(-x^{-\alpha}) = x^{-\alpha} - \frac{x^{-2\alpha}}{2} \dots \sim x^{-\alpha}, \quad \text{para } x \rightarrow \infty.$$

Se dice en este caso que la cola derecha \bar{F} es de variación regular de índice $-\alpha$. En realidad, se puede probar que el Máximo Dominio de Atracción de la Distribución Fréchet está compuesto por las distribuciones cuya cola derecha es de variación regular de índice $-\alpha$ y se puede caracterizar mediante el siguiente resultado.

Proposición 1.4.2: (condición de Von Mises)

Sea F función de distribución absolutamente continua con densidad f que cumple con:

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

Entonces $F \in \text{MDA}(\Phi_\alpha)$.

Ejemplo 1.4.1: (Distribución de Pareto)

Sea la variable aleatoria X tal que:

$$\bar{F}(x) = \left(\frac{k}{k+x}\right)^\alpha, \text{ con } \alpha > 0 \text{ y } k > 0.$$

$$F(x) = 1 - \left(\frac{k}{k+x}\right)^\alpha \Rightarrow f(x) = \alpha \left(\frac{k}{k+x}\right)^{\alpha-1} \frac{k}{(k+x)^2}$$

Entonces el cociente,

$$\frac{xf(x)}{\bar{F}(x)} = \frac{\alpha x \left(\frac{k}{k+x}\right)^{\alpha-1} \frac{k}{(k+x)^2}}{\left(\frac{k}{k+x}\right)^\alpha} = \frac{x}{k+x} \alpha \rightarrow \alpha \text{ para } x \rightarrow \infty,$$

y así se tiene que la distribución de Pareto está en el MDA de la distribución Fréchet.

Se puede mostrar que cada F perteneciente al MDA de la distribución Fréchet es de cola equivalente con una distribución absolutamente continua que satisface la condición de Von Mises. Por lo cual, podemos decir que el $\text{MDA}(\Phi_\alpha)$ consta de todas las distribuciones que satisfacen la condición de Von Mises más sus distribuciones de colas equivalentes.

Máximo Dominio de Atracción de la Distribución Weibull

Sea X con distribución Weibull,

$$\psi_\alpha(x) = \exp\{-(-x)^\alpha\}, \quad x < 0, \quad \alpha > 0.$$

Si consideramos la relación entre ψ_α y Φ_α , esta es: $x > 0$, $\psi_\alpha(-x^{-1}) = \Phi_\alpha(x)$, la que significa que si $Y = -\frac{1}{X}$, donde X tiene distribución Fréchet entonces Y tiene distribución Weibull. Naturalmente tendremos que esperar que sus MDA se definan de manera similar.

Ciertamente, se tiene que $F \in \text{MDA}(\psi_\alpha) \Leftrightarrow x_F < \infty$ y $\bar{F}(x_F - x^{-1})$ es de variación regular de índice $-\alpha$.

Similar al caso Fréchet se presenta la siguiente condición, que también servirá para caracterizar este dominio.

Proposición 1.4.3: (condición de Von Mises)

Sea F una función de distribución absolutamente continua con densidad f positiva para algún intervalo finito (z, x_F) que cumple con:

$$\lim_{x \rightarrow x_F^-} \frac{(x_F - x)f(x)}{\bar{F}(x)} = \alpha > 0 \quad (1.4.1)$$

Entonces $F \in \text{MDA}(\psi_\alpha)$.

En resumen, se vio que utilizando los conceptos de funciones de variación regular y de distribuciones de colas equivalentes, es posible caracterizar los dominios de atracción de las distribuciones Fréchet Φ_α y Weibull Ψ_α . En cambio, el máximo dominio de atracción de la distribución Gumbel, que es el que nos importa especialmente, no es fácilmente caracterizable dado que cubre un amplio rango de distribuciones y no tiene una conexión directa con funciones de variación regular.

Máximo Dominio de Atracción de la Distribución Gumbel

Utilizando una expansión en serie de Taylor se puede demostrar que la cola derecha de la distribución Gumbel decrece exponencialmente, es decir:

$$\bar{G}(x) = 1 - \text{Exp}(-e^{-x}) \sim e^{-x} \quad \text{para } x \rightarrow \infty.$$

Sin embargo existen distribuciones con colas diferentes a la cola de la distribución exponencial y aun así están dentro del Dominio de Atracción de la Distribución Gumbel, estas se pueden caracterizar mediante una extensión de la representación propuesta por Von Mises.

Definición 1.4.3: (Función de von Mises)

Sea F una función de distribución con punto final derecho $x_f \leq \infty$ y supongase que existe un punto $z < x_f$ tal que F se puede representar como:

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

Donde c es una constante positiva, la función auxiliar $a(\cdot)$ es positiva y absolutamente continua. Además su derivada a' es tal que:

$$\lim_{x \rightarrow x_f^-} a'(x) = 0.$$

Entonces se dice que F es una función de von Mises.

Ejemplo 1.4.2: Sea X una variable aleatoria con distribución exponencial, entonces su función de sobrevivida será:

$$\bar{F}(x) = 1 - F(x) = e^{-\lambda x}, \quad \text{para } x > 0, \lambda > 0.$$

Esta función se puede escribir de la forma:

$$\bar{F}(x) = \exp\{-\lambda x\} = \exp \left\{ - \int_0^x \lambda dt \right\}.$$

Luego F es una función de von Mises con función auxiliar $a(x) = \frac{1}{\lambda}$.

Ejemplo 1.4.3: Sea X una variable aleatoria con distribución Weibull con función de supervivencia dada por:

$$\bar{F}(x) = \exp\{-cx^p\} = \exp\left\{-\int_0^x cpt^{p-1} dt\right\}$$

Luego es una función de von Mises con función auxiliar $a(x) = \frac{x^{1-p}}{cp}$.

Ejemplo 1.4.4: Sea X una variable aleatoria con distribución de Pareto, entonces la cola de la distribución tiene la siguiente forma:

$$\bar{F}(x) = \left(\frac{k}{k+x}\right)^\alpha, \quad \text{con } \alpha > 0, k > 0.$$

Entonces

$$\begin{aligned} \bar{F}(x) &= \exp\{\alpha(\ln k - \ln(k+x))\} \\ &= \exp\{-\alpha \ln(k+x) + \alpha \ln k\} \\ &= \exp\left\{-\int_0^x \frac{\alpha}{k+t} dt\right\} \end{aligned}$$

Luego es una función de von Mises con función auxiliar dada por $a(x) = \frac{k+x}{\alpha}$.

Ver otros ejemplos de funciones de este tipo en Embrechts (1997), Sección 3.3.3.

Observe que, de la última definición $\ln(\bar{F}(x)) = \ln(c) - \int_z^x \frac{1}{a(t)} dt$, y por lo tanto

$$\frac{d}{dx}(-\ln \bar{F}(x)) = \frac{f(x)}{\bar{F}(x)} = \frac{1}{a(x)}, \quad z < x < x_f$$

Luego ocurre, que para toda función de von Mises F con función de densidad positiva $F' = f$, la función auxiliar puede ser elegida como,

$$a(x) = \frac{\bar{F}(x)}{f(x)}.$$

Y además cumple con las siguientes condiciones,

i) Si $x_f = \infty$ entonces

$$\lim_{x \rightarrow \infty} \frac{xf(x)}{\bar{F}(x)} = \infty.$$

ii) Si $x_f < \infty$, entonces

$$\lim_{x \rightarrow x_f} \frac{(x_f - x)f(x)}{\bar{F}(x)} = \infty.$$

Además note que $a^{-1}(x) = \frac{f(x)}{\bar{F}(x)}$ es el llamado "hazard rate" o "tasa de mortalidad" de

F.

También se puede probar que una función de von Mises pertenece al Dominio Gumbel y además esta representación permite obtener las constantes normalizadoras como:

$$d_n = F^{-1}\left(1 - \frac{1}{n}\right) \quad \text{y} \quad c_n = a(d_n)$$

Donde se denota por F^{-1} a la función inversa generalizada de F y "a" es la función auxiliar de la representación de von Mises. Embrechts (1997) sección 3.3.3.

Observación:

Sea F una función de distribución con punto final derecho $x_f \leq \infty$ y suponga que existe algún $z < x_f$, tal que F tiene derivada hasta orden 2 en (z, x_f) con densidad positiva $f(x) = F'(x)$ y $F''(x) < 0$ para $z < x < x_f$. Entonces F es una función de von Mises con función auxiliar dada por $a = \bar{F}/f$ si y solo si

$$\lim_{x \rightarrow x_f^-} \frac{\bar{F}(x)F''(x)}{f^2(x)} = -1.$$

Esta condición se puede aplicar para demostrar que una determinada función de distribución es una función de von Mises y por lo tanto pertenece al Dominio Gumbel.

Caracterización del Dominio de Atracción Gumbel

El conjunto de las funciones de von Mises no alcanza a cubrir completamente al Máximo Dominio de Atracción de la Distribución Gumbel, pero una simple generalización de esta propiedad nos da una completa caracterización del MDA(G).

Proposición 1.4.4:

Una función de distribución F con punto final derecho $x_f \leq \infty$ pertenece al Máximo Dominio de Atracción de la Distribución Gumbel si y sólo si existe un punto $z < x_f$ tal que F se puede representar como:

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

Donde c y g son funciones medibles que satisfacen que $c(x) \rightarrow c$, $g(x) \rightarrow 1$ cuando $x \rightarrow x_f$ por valores menores que x_f , y la función $a(x)$ es positiva y absolutamente continua. Además su derivada a' es tal que:

$$\lim_{x \rightarrow x_f^-} a'(x) = 0.$$

Esta modificación a la definición de una función de von Mises es simple en sentido que agrega funciones cola-equivalentes. Entonces, otra manera de expresar lo mismo sería

diciendo que: El Dominio Gumbel es igual al conjunto de las funciones de von Mises más sus distribuciones cola-equivalentes.

Puede mostrarse que el MDA(G) contiene una gran variedad de distribuciones cuyas colas pueden ser verdaderamente diferentes, desde colas moderadamente pesadas hasta colas muy livianas incluso con punto final derecho finito.

Es por esta razón que se considera que esta familia de distribuciones es lo suficientemente rica, como para considerarse como modelo de estimación para muchos fenómenos, el problema es que, como ya se mencionó antes, tiene la dificultad de ser una familia no paramétrica de distribuciones. En este trabajo se propone aproximar el Dominio Gumbel mediante una familia paramétrica obtenida de ampliar la familia Gumbel mediante una transformación Potencial (Capítulo 3).

Ejemplo 1.4.5: La Distribución Normal pertenece al Dominio Gumbel

Teniendo en cuenta que para la distribución Normal ocurre que $-\frac{\varphi'(x)}{x} = \varphi(x)$ y observando que

$$\lim_{x \rightarrow \infty} \frac{\bar{\Phi}(x)}{\frac{\varphi(x)}{x}} = \lim_{x \rightarrow \infty} \frac{-\varphi(x)}{\frac{\varphi'(x)}{x} - \frac{\varphi(x)}{x^2}} = \lim_{x \rightarrow \infty} \frac{1}{1 + \frac{1}{x^2}} = 1$$

Obtenemos la razón de Mill, $\bar{\Phi}(x) \sim \varphi(x)/x$. Entonces

$$\lim_{x \rightarrow x_f^+} \frac{\bar{F}(x)F''(x)}{f^2(x)} = \lim_{x \rightarrow \infty} \frac{\bar{\Phi}(x)\varphi'(x)}{\varphi^2(x)} = -1$$

Luego, la función de distribución acumulada de la Distribución Normal es una función de von Mises y por lo tanto pertenece al Máximo Dominio de Atracción de la Distribución Gumbel.

Capítulo 2

Aplicación de la Transformada Box-Cox para aproximar una distribución Gumbel

En este capítulo se trata principalmente, con los detalles del procedimiento de transformar los datos para aproximarlos mejor a la distribución de extremos Gumbel. Esta necesidad surge, porque en los casos aplicados muchas veces el modelo Gumbel puede ser demasiado rígido y no captar bien la verdadera forma de las colas de las distribuciones involucradas.

Se discute en qué sentido evaluar la aproximación y se muestra gráficamente la capacidad de la transformación para realizar la aproximación. Si se tiene en cuenta que la aproximación es con el fin de calcular percentiles superiores, entonces el interés está en tener una aproximación lo más exacta posible en la cola derecha de la distribución.

2.1: Introducción

En este capítulo se tratan los pormenores de la aplicación de una transformación potencial a los datos para aproximarlos mejor a la distribución de extremos Gumbel. Esta necesidad surge del hecho que el Dominio de Atracción de la Distribución Gumbel, en el cual se está interesado, es muy amplio en sentido de contener distribuciones de colas muy variadas, pues existen distribuciones en este dominio que tienen velocidad de convergencia muy lenta a la asíntota, como ser por ejemplo el caso de la distribución Normal.

En la teoría de extremos se muestra que la distribución del máximo de una muestra de variables aleatorias X_1, \dots, X_T independientes e idénticamente distribuidas, cuando converge para $T \rightarrow \infty$, lo hace a una de las tres distribuciones extremales, Fréchet, Gumbel y Weibull, Capítulo 1.

En particular en este trabajo el interés está puesto en las distribuciones que tienen convergencia a la distribución Gumbel. En estadística de valores extremos, que trata con la estimación de la distribución del máximo (o lo que es lo mismo, la estimación de los percentiles superiores de la distribución) puede ocurrir que, en los casos aplicados la cantidad de variables T considerada no sea lo suficientemente grande como para suponer, invocando a la convergencia a las distribuciones extremales, que la distribución de los datos sea muy próxima a la distribución límite (Gumbel), provocando sesgos importantes en las estimaciones.

Por ejemplo, en nuestro caso se cuenta con datos que son velocidades máximas mensuales de vientos. De éstas velocidades máximas mensuales se toman las velocidades máximas anuales (para reforzar la hipótesis de independencia entre las observaciones y para mejorar la proximidad a la asíntota, la distribución Gumbel) que está muy recomendado en la bibliografía de extremos, obteniéndose así $T = 12$ (12 meses) y aun así tomar la distribución límite puede ser poco sustentable y además producir sesgos en las estimaciones de los percentiles superiores dado que el modelo ajustado (distribución Gumbel) puede ser demasiado rígido y no ajustarse bien a los datos, especialmente en las colas. Esto restringe significativamente la aplicabilidad del método, por lo que sería entonces interesante dar al modelo cierta flexibilidad a partir

del Modelo Gumbel para que pueda captar estas diferencias en los índices de las colas sin salirse del Dominio de Atracción de la Distribución Gumbel.

Una alternativa muy atractiva es debilitar significativamente la hipótesis de que la variable X tiene Distribución Gumbel, permitiendo que la distribución de X esté en el Máximo Dominio de Atracción de la Distribución Gumbel, lo que representaría una importante ampliación de la aplicabilidad de los métodos debido a la riqueza del dominio Gumbel, que contiene distribuciones con colas desde muy livianas, como la distribución Normal, hasta moderadamente pesadas como por ejemplo la distribución Lognormal. El problema que surge de inmediato es que al ser el dominio Gumbel una familia no paramétrica de distribuciones, (no existe un modelo paramétrico a estimar) dificulta el ajuste de este modelo a los datos reales.

Un procedimiento propuesto aquí para determinar un modelo paramétrico que generalice al modelo Gumbel, que esté dentro del Máximo Dominio de Atracción Gumbel y que de algún modo lo represente con fines de estimación, se basa en la siguiente idea: Supongamos que ciertos datos son generados por una distribución perteneciente al Máximo Dominio de Atracción Gumbel, entonces, mediante una transformación adecuada aproximo estos datos a la distribución Gumbel y luego estimo con estos datos transformados los parámetros de la distribución. Entonces estamos ampliando la distribución Gumbel a la "familia de distribuciones" tales que transformadas se aproximan bien, *en cierto sentido específico*, a la distribución Gumbel.

Esta idea, similar a la de Box-Cox (1964), que utiliza la transformación potencial,

$$y = g(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda > 0, \\ \ln(x) & \lambda = 0 \end{cases}, \quad x > 0.$$

para lograr normalidad, será utilizada aquí para obtener proximidad a la distribución Gumbel, donde el parámetro λ , que llamaremos parámetro de forma, es quien proporciona la flexibilidad adicional para mejorar el ajuste en las colas de las distribuciones.

Observe que la distribución Gumbel tiene recorrido en $(-\infty, +\infty)$ y la transformación sólo toma valores mayores que $-1/\lambda$, esta dificultad no será tenida en cuenta dado que la variable original X será corregida adecuadamente mediante parámetros de posición y escala para no tener problemas de definición y mejorar eficiencia de la transformación. Observe también que, al variar λ el modelo de ajuste tendrá colas más próximas o más alejadas de la distribución Gumbel, sin salirse del Máximo Dominio de Atracción de la Distribución Gumbel.

Lo que interesa, entonces, es estudiar si esta familia de distribuciones proporciona un modelo lo suficientemente rico para ajustar bien a los datos, si es realmente "más amplio" que el modelo Gumbel para la estimación de la distribución de extremos, y si de alguna manera representa al Máximo Dominio de Atracción de la Distribución Gumbel.

Es claro que dada una distribución perteneciente al Máximo Dominio de Atracción de la Distribución Gumbel en general no existirá ningún valor del parámetro de forma λ tal que esta distribución transformada tenga exactamente una distribución Gumbel, por ejemplo, Si X tiene distribución Normal estándar no existirá un parámetro $\lambda > 0$ tal que

$$Y = \frac{x^\lambda - 1}{\lambda}$$

tenga distribución Gumbel. Sin embargo, teniendo en cuenta que el objetivo final es la estimación del máximo y por lo tanto lo que realmente importa es el comportamiento de la distribución en las colas, podemos buscar un $\lambda > 0$ tal que X e Y sean distribuciones de colas lo suficientemente próximas, lo que significa en otras palabras lograr estimaciones precisas de los percentiles superiores de la distribución de interés. Por ejemplo, para el caso de aproximar la distribución Normal por una distribución Gumbel el valor óptimo es de $\lambda \approx 2$. Ver sección 3.3.

2.2: La Distribución Gumbel

Uno de los primeros científicos en aplicar esta teoría fue el matemático alemán Emil Gumbel (1891-1966). Se centró principalmente en aplicaciones de la teoría de valores extremos a problemas de ingeniería, en particular de modelado de fenómenos meteorológicos tales como flujos de inundación anual.

La distribución de Gumbel, también es conocida como la distribución de valores extremos de Tipo I, está definida sobre toda la recta real y tiene la siguiente función de distribución acumulada:

$$G(x) = \exp\left(-\exp\left(-\frac{x-\xi}{\theta}\right)\right), \quad -\infty < x < \infty. \quad (2.2.1)$$

Donde $-\infty < \xi < \infty$ y $\theta > 0$, μ es el parámetro de localización, y σ es el parámetro de escala de la distribución. La forma del modelo Gumbel no depende de los parámetros de la distribución.

Para el caso especial $\xi = 0$ y $\theta = 1$, es decir

$$G(x) = \exp(-\exp(-x)), \quad -\infty < x < \infty, \quad (2.2.2)$$

la llamaremos Distribución Gumbel Estándar.

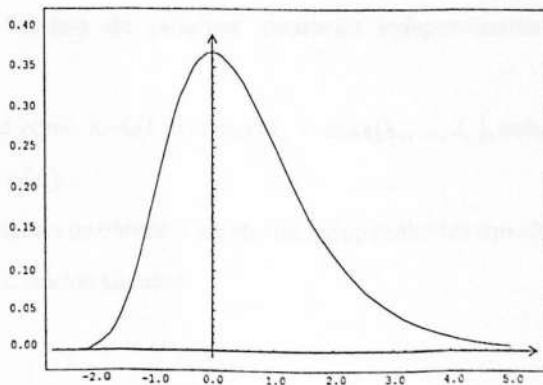


Figura 2.2.1: Densidad de la distribución Gumbel Estándar, $\xi = 0$ y $\theta = 1$.

A continuación se muestran sus parámetros de posición y escala:

Si M es la mediana de la distribución entonces $\exp\left(-\exp\left(-\frac{M-\xi}{\theta}\right)\right) = 1/2$, entonces

$$M = \xi - \theta \ln\left(\ln\left(\frac{1}{2}\right)\right)$$

Para encontrar la esperanza, recuerde que

$$E(X) = \int_0^1 F^{-1}(y) dy = \int_0^1 \xi - \theta \ln(-\ln y) dy = \xi + \theta \gamma,$$

donde

$$\gamma = \int_0^1 -\ln(-\ln y) dy \cong 0.57722,$$

se llama constante de Euler-Mascheroni.

La moda es ξ , y también se puede probar que su desviación estándar es:

$$\sigma = \frac{\theta \pi}{\sqrt{6}}$$

Otra propiedad importante, es que el máximo de T variables aleatorias independientes con distribución Gumbel tiene también distribución Gumbel con diferentes parámetros.

Teorema 2.2.1: Máximo de variables aleatorias independientes con distribución Gumbel.

Sean X_1, \dots, X_n iid como $X \sim G(\xi, \theta)$ y sea $M_n = \max\{X_1, \dots, X_n\}$, entonces $M_n \sim G(\xi^*, \theta)$ donde $\xi^* = \xi + \theta \ln(n)$.

Es decir que el máximo de variables aleatorias independientes con distribución Gumbel tiene también distribución Gumbel.

Demostración

$$\begin{aligned}F^n(z) &= \left[\exp\left(-\exp\left(-\frac{z-\xi}{\theta}\right)\right) \right]^n \\&= \exp\left(-n \exp\left(-\frac{z-\xi}{\theta}\right)\right) \\&= \exp\left(-\exp\left(\ln(n) - \frac{z-\xi}{\theta}\right)\right) \\&= \exp\left(-\exp\left(-\frac{z - (\xi - \theta \ln(n))}{\theta}\right)\right).\end{aligned}$$

Luego $M_n = \max\{X_1, \dots, X_n\}$ tiene distribución Gumbel con parámetro de posición $\xi^* = \xi - \theta \ln(n)$, y parámetro de escala $\theta^* = \theta$. ■

Note que $F^n(z) = F(z + \theta \ln(n)) = F(c_n z + d_n)$ donde $c_n = 1$ y $d_n = \theta \ln(n)$, por lo tanto la distribución Gumbel es máx-estable como era de esperarse.

2.3: Aplicación de la Transformación Box-Cox.

Suponga que X_1, \dots, X_n es una muestra aleatoria de una variable aleatoria X con distribución en el Máximo Dominio de Atracción de la Distribución Gumbel.

Se desea utilizar una modificación de la siguiente función para transformar los datos:

$$y = g(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda > 0, \\ \ln(x) & \lambda = 0 \end{cases}, \quad x > 0 \quad (2.3.1)$$

Esta fue propuesta por Box-Cox (1964) para resolver problemas de falta de normalidad y de heterocedasticidad en los datos.

En el presente caso el objetivo será aproximar la distribución de los datos a la distribución Gumbel y así lograr un mejor ajuste especialmente en las colas. Este procedimiento conduce de una manera natural a extender el modelo de ajuste a partir del modelo Gumbel a una familia más rica de distribuciones para obtener una mejor estimación de los percentiles superiores o periodos de recurrencia de sucesos extremos. Sin embargo, si se aplica (2.3.1) directamente para aproximar los datos a una distribución Gumbel de parámetros ξ, θ surge un problema de identificación entre los parámetros de posición y escala de la variable original y los de la variable transformada respectivamente, es por ello que la transformación más apropiada utilizada aquí deberá tener la forma:

$$Y = G(X) = \frac{c}{\lambda} \left[\left| 1 + \frac{X - \xi}{c\theta} \right|^\lambda \operatorname{sgn} \left(1 + \frac{X - \xi}{c\theta} \right) - 1 \right], \quad -\infty < X < +\infty \quad (2.3.2)$$

Donde los parámetros son tales que: $\lambda > 0, -\infty < \xi < \infty$ y $\theta > 0$ y la constante c se elige entre 2 y 5 de modo que la condición $X \geq \xi - c\theta$ se verifique en el caso especial que se esté tratando, es decir que la función signo que se ve en (2.3.2) tomará siempre el valor positivo.

La transformación (2.3.2) surge de (2.3.1) luego de realizar una estandarización de la variable original X realizada para solucionar el problema de falta de identificación de los

parámetros y así obtener una mejor eficiencia en la estimación e interpretación de los mismos.

Es claro que con la transformación así planteada, los parámetros se estiman de modo que la variable transformada Y se aproxime a una distribución Gumbel Estándar. Este procedimiento se explica en el capítulo siguiente.

Ahora se pueden considerar las siguientes preguntas

¿Cuáles son las distribuciones que transformadas se aproximan bien (en un sentido específico) a la distribución Gumbel?

Y

¿Cuáles son las distribuciones que transformadas tienen exactamente una distribución Gumbel?

Con respecto de la segunda pregunta, se puede identificar a las distribuciones que transformadas tienen exactamente una distribución Gumbel, como la Familia de Distribuciones Anti-transformadas mediante (2.3.2) de la distribución Gumbel Estándar. Pero, teniendo en cuenta que lo que interesa realmente es estimar los percentiles superiores y por lo tanto sólo el comportamiento en la cola derecha de la distribución es importante, se puede ampliar considerablemente esta familia a todas las distribuciones cuya cola derecha transformada es equivalente a la cola de la distribución Gumbel. Lo que responde a la pregunta 1, pues el modelo será el conjunto de todas las distribuciones tales que transformadas tienen una cola equivalente a la distribución Gumbel. Si observamos la Figura 2.3.1 (a) notaremos que la distribución normal transformada se ajusta bien a la Gumbel en la parte derecha, línea punteada, y tiene un desajuste en los valores menores (se aleja de la linealidad). En un sentido práctico, esto significa que en ciertos casos será más eficiente utilizar solo los datos en la parte derecha de la distribución.

En las figuras siguientes se presentan dos ejemplos, caso Normal y Lognormal, que muestran la capacidad de la transformación propuesta para aproximar distribuciones del Dominio Gumbel.

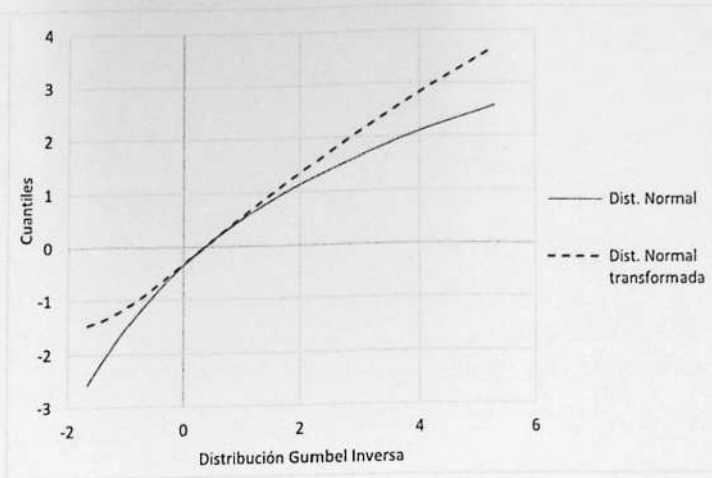
Lo que se presenta es el Q-Q plot de valores esperados, es decir, los valores esperados transformados y sin transformar de las distribuciones Normal y Lognormal en función de los valores esperados de la Gumbel estándar. Es decir que la proximidad a la línea recta representa la proximidad de la distribución transformada a la distribución Gumbel. Este método permite apreciar gráficamente cómo la transformación deforma a la distribución de interés para aproximarla a la Distribución Gumbel y un método basado en esta idea será utilizado para la estimación de los parámetros.

En la Figura 2.3.1 (b) se ve como la aproximación de la distribución Log-normal es muy buena en todo el dominio, línea punteada.

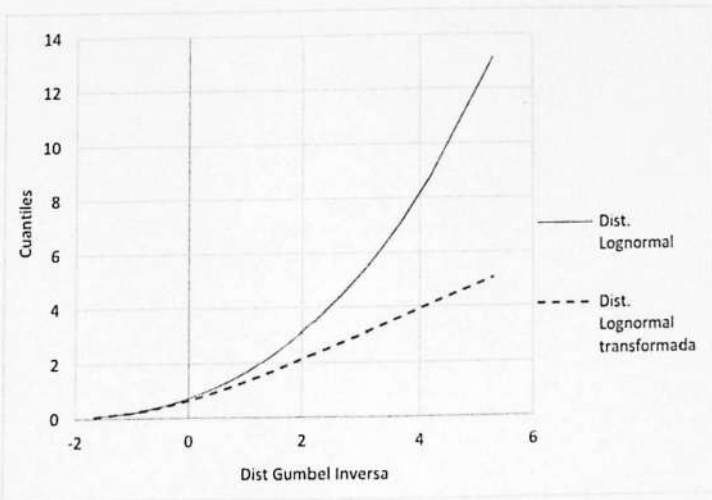
Estos gráficos también permiten ver si la cola de la distribución original es más o menos pesada que la cola de la distribución Gumbel, en la Figura 2.3.1 (a) (caso normal) la línea azul indica una cola más liviana que la Gumbel y la Figura 2.3.1 (b) (caso log-normal) muestra lo contrario.



Figura 2.3.1: a) Q-Q plot de valores esperados de la distribución Normal transformada y sin transformar en función de los valores esperados de la Gumbel estándar. b) Q-Q plot de valores esperados de la distribución Log-normal transformada y sin transformar en función de los valores esperados de la Gumbel estándar.



(a)



(b)

Figura 2.3.1: a) Q-Q plot Distribuciones Normal versus distribución Gumbel y Q-Q plot Distribuciones Normal transformada ($\lambda = 2$) vs distribución Gumbel.
 b) Q-Q plot Distribuciones Lognormal versus distribución Gumbel y Q-Q plot Distribuciones Lognormal transformada ($\lambda \approx 0$) vs distribución Gumbel.

Capítulo 3:

Distribución Gumbel Ampliada

Este capítulo está dedicado a la Familia Potencial-Gumbel y cómo surge de aplicar la anti-transformada de una modificación de la transformada de Box-Cox a la Distribución Gumbel. Se muestra que esta familia está contenida dentro del Máximo Dominio de Atracción de la Distribución Gumbel; que tiene distribuciones de diferentes formas (simétricas y asimétricas) y de diferentes colas; que al igual que el Máximo Dominio de Atracción de la Distribución Gumbel contiene distribuciones con colas muy livianas que corresponden a valores de $\lambda > 1$, hasta distribuciones con colas moderadamente pesadas para valores de $\lambda \approx 0$.

Lo que se propone en definitiva es, visualizar a la familia Potencial-Gumbel como una semi-parametrización del Máximo Dominio de Atracción de la Distribución Gumbel.

3.1. Introducción

Un resultado clásico de la teoría de extremos afirma que la distribución del máximo de un conjunto de observaciones independientes e idénticamente distribuidas X_1, \dots, X_n de una variable aleatoria con distribución F , si converge cuando n tiende a infinito, lo hace a una de tres posibles distribuciones límites, una de las cuales es la distribución de una variable aleatoria Gumbel, que juega un papel central en este trabajo.

Es un hecho conocido que la distribución del máximo en este caso puede converger a su asíntota en forma muy lenta como lo hicieron notar Fisher y Tippett (1928). Ver al respecto Galambos (1987), Resnick (1987), Embrechts y otros (1997) y Leadbetter y otros (1982). En Galambos se prueba que la velocidad de convergencia de la distribución del máximo de observaciones Gaussianas i.i.d. es del orden de $1/\ln n$, mientras que para la distribución exponencial es del orden $1/n$.

Lo importante es que en general, la convergencia de la distribución del máximo a su límite puede ser muy lenta.

En este trabajo se propone modelar el Máximo Dominio de Atracción Gumbel mediante una familia de distribuciones que se obtiene mediante la aplicación de una modificación de la transformación de Box-Cox inversa a una distribución Gumbel.

Sean X_1, \dots, X_n observaciones de una variable aleatoria X con función de distribución F_X , sea $Y = g(X)$ donde g es una función estrictamente monótona.

Si g^{-1} representa la función inversa de g y si $Y_i = g(X_i)$, $i = 1, \dots, n$, se tiene que:

$$\max(X_1, \dots, X_n) = g^{-1}[\max(Y_1, \dots, Y_n)],$$

Esta igualdad permite transformar las observaciones en un conjunto con mejores propiedades para la aplicación de la estadística de extremos. Note que no se requiere que las observaciones sean independientes.

Por otra parte si X_1, \dots, X_n son i.i.d. según F y F_Z es la función de distribución de $Z = \max\{X_1, \dots, X_n\}$ entonces $F_Z(z) = F(z)^n$. Por lo tanto los valores de $F_Z(z) > \frac{1}{2}$ están determinados por los valores

$$F(z) > \left(\frac{1}{2}\right)^{\frac{1}{n}} = 0.9330,$$

si $n = 10$ por ejemplo, lo cual muestra que sólo es crítica la transformación de la cola derecha.

3.2. Distribución Potencial Gumbel

La distribución Potencial-Gumbel, es la familia de distribuciones que surge de aplicar la transformada (2.3.2) inversa a una distribución Gumbel Estándar (es decir de parámetros $\xi = 0$ y $\theta = 1$). Esta familia de distribuciones que se obtiene al agregar el tercer parámetro λ de la transformación, tendrá una mayor flexibilidad que le permitirá un mejor ajuste en las colas sin complicar demasiado el modelo.

Definición 3.2.1 (Distribución Potencial Gumbel)

Se dice que una variable aleatoria X tiene distribución *Potencial Gumbel* de parámetros ξ, θ, λ con $-\infty < \xi < \infty, \theta > 0$ y $\lambda > 0$, si la variable aleatoria

$$Y = g(X) = \frac{c}{\lambda} \left[\left| 1 + \frac{X - \xi}{c\theta} \right|^\lambda \operatorname{sgn} \left(1 + \frac{X - \xi}{c\theta} \right) - 1 \right], \quad -\infty < X < +\infty,$$

tiene distribución Gumbel Estándar, es decir tiene función de distribución dada por:

$$F_Y(y) = \exp(-\exp(-y)).$$

Y por lo tanto, la función de distribución acumulada de X será:

$$F_X(x) = P(X \leq x) = P(Y \leq g(x)) = F_Y(g(x))$$

Entonces

$$F_X(x) = \exp \left(-\exp \left(-\frac{c}{\lambda} \left[\left| 1 + \frac{x - \xi}{c\theta} \right|^\lambda \operatorname{sgn} \left(1 + \frac{x - \xi}{c\theta} \right) - 1 \right] \right) \right)$$

En un caso aplicado, la elección de la constante c permite que el total de los datos se localicen dentro la primera rama y entonces bajo este supuesto se puede considerar, que la variable aleatoria X es absolutamente continua con función de densidad:

$$f_x(x) = \exp(-\exp(-g(x)))\exp(-g(x))g'(x)$$

Fearn y Nebenzahl [7] y S. Nadarajah [10] propusieron el uso de transformaciones de potencia para la aproximación de cuantiles. Vanroelen [8] y Teguel y Vanroelen [9] han estudiado el efecto de las transformaciones de tipo Box-Cox en la aproximación del máximo de variables aleatorias pertenecientes a los dominios de atracción Frechet y Gumbel.

En el presente trabajo se define la familia Potencial Gumbel como modelo del dominio de atracción Gumbel y se destaca la importancia que tiene para la estimación del máximo de observaciones pertenecientes a ese dominio de atracción.

Toda función de distribución de una variable absolutamente continua con recorrido $(0, \infty)$ se puede expresar en función de la tasa de riesgo, $h(x) = f(x)/(1 - F(x))$, como:

$$\bar{F}(x) = 1 - F(x) = \exp\left(-\int_0^x h(t)dt\right)$$

Entonces, como se vio en el Capítulo 1, una representación de las funciones de distribución del dominio Gumbel es de la forma:

$$\bar{F}(x) = c(x) \exp\left(-\int_z^x \frac{1}{a(t)} dt\right), \quad z < x < \infty. \quad (3.2.1)$$

donde z es una constante tal que, $z < x < x_f$, $a(x)$ una función positiva absolutamente continua con derivada $a'(x) \rightarrow 0$ cuando $x \rightarrow \infty$ y $c(x)$ es una función que cumple que $c(x) \rightarrow c > 0$ cuando $x \rightarrow \infty$.

Como se vio en el Capítulo 2, si una función F tiene representación (3.2.1) la función auxiliar $a(x)$ se puede elegir como la recíproca de la tasa de riesgo (hazard rate),

$$a(x) = h^{-1}(x) = \frac{\bar{F}(x)}{f(x)}$$

Dado que, para $x > \xi - c\theta$,

$$g'(x) = \frac{1}{\theta} \left(1 + \frac{x - \xi}{c\theta}\right)^{\lambda-1} \sim \frac{x^{\lambda-1}}{k}$$

con $k = \theta^\lambda c^{\lambda-1}$.

Tenemos que la función auxiliar $a(x)$ para la distribución F_X puede elegirse como:

$$a(x) = \frac{\bar{F}_X(x)}{f_X(x)} = \frac{\bar{F}_Y(g(x))}{f_Y(g(x))g'(x)} \sim kx^{1-\lambda}$$

y así, puesto que $\lambda > 0$, $a'(x) \sim k(1-\lambda)x^{-\lambda} \rightarrow 0$, cuando $x \rightarrow \infty$.

Por lo tanto la distribución F pertenece al Máximo Dominio de Atracción de la Distribución Gumbel.

Además note que si $\lambda = 0$, $a'(x) = k$, no converge a 0 y por lo tanto la distribución ya no pertenece al dominio Gumbel. Puede probarse que en este caso la variable transformada $X = \exp(Y)$ tiene distribución Fréchet.

Una propiedad general de las funciones $a(x)$ es que:

$$\lim_{x \rightarrow \infty} \frac{a(x)}{x} = 0.$$

O sea que $a(x)$ tiende a crecer más lentamente que x , y precisamente esto ocurre con $a(x) = kx^{1-\lambda}$. El valor $\lambda = 1$ corresponde a $a(x) = k$, constante que a su vez corresponde a la distribución Gumbel.

A valores de λ grandes le corresponden valores grandes de a^{-1} y por lo tanto valores grandes de su integral entre z y x (que son, dentro del dominio Gumbel, las distribuciones con colas más livianas).

Por otro lado, la familia de funciones

$$a(x) = kx^{1-\lambda}, \quad \lambda > 0, \tag{3.2.2}$$

tiene como función frontera a la función $a(x) = kx$ la cual no pertenece a la familia.

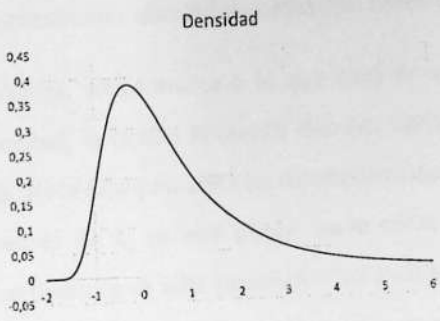
Además suponiendo que las funciones $a(x)$ son de variación regular. Esto es:

$$a(x) = L(x)x^p, \quad p > 0 \tag{3.2.3}$$

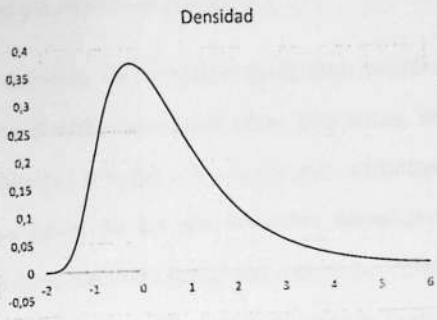
donde $L(x)$ es una función de variación lenta. En la expresión (3.2.2) se reemplaza la función $L(x)$ por una constante y bajo este supuesto la familia (3.2.3) parece ser lo suficientemente rica como para aproximar la familia de todas las posibles funciones $a(x)$.

Distribuciones pertenecientes a la familia Potencial Gumbel

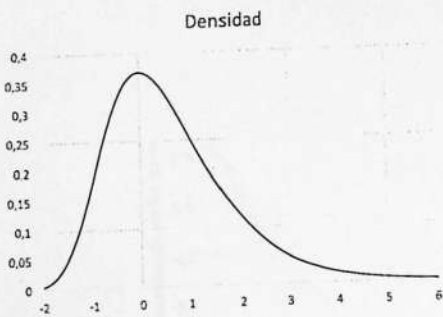
A continuación se presenta como ejemplo, las funciones de densidad de probabilidad de algunos miembros de la familia Potencial Gumbel para ciertos valores de los parámetros.



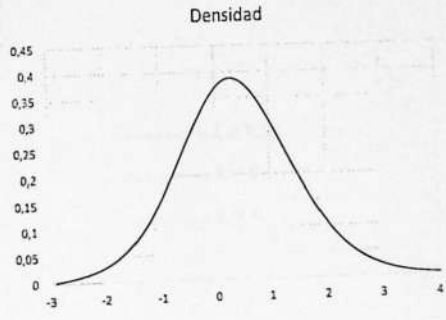
$\xi = 0, \theta = 1, \lambda = 0.01$ es el miembro de cola derecha más pesada



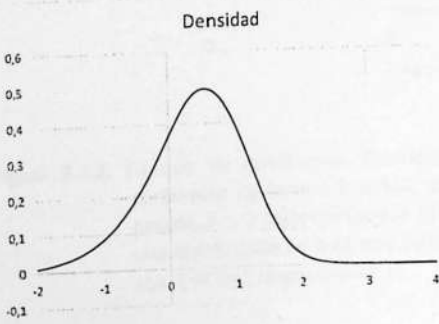
$\xi = 0, \theta = 1, \lambda = 0.5$



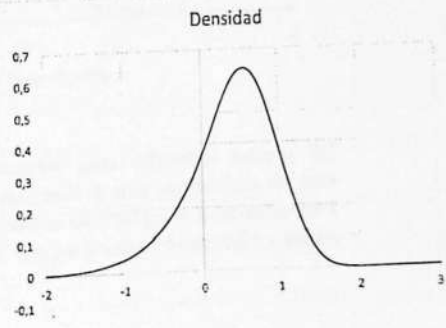
$\xi = 0, \theta = 1, \lambda = 1$, Corresponde a la distribución Gumbel



$\xi = 0, \theta = 1, \lambda = 2$, es el miembro que mejor se aproxima a la distribución Normal



$\xi = 0, \theta = 1, \lambda = 4$ miembro de cola más liviana que la Normal



$\xi = 0, \theta = 1, \lambda = 8$ miembro de cola muy liviana

Figura 3.2.1: Miembros de la familia Potencial Gumbel para valores seleccionados de los parámetros

En estos ejemplos puede notarse que la familia potencial Gumbel contiene distribuciones asimétricas negativas, simétricas y asimétricas positivas.

Además, similarmente a lo que pasa en el dominio de atracción de la distribución Gumbel, la familia Potencial Gumbel contiene distribuciones con colas diferentes, la siguiente figura muestra las curvas correspondientes a $\bar{F}_X(x) = 1 - F_X(x)$ para distintos valores de λ , en ella puede verse como las colas de las distribuciones dependen fuertemente de este parámetro; para valores de λ menores (próximos a cero) las colas son más pesadas y para valores muy grandes de λ las colas son muy livianas como aproximándose a variables de punto final finito.

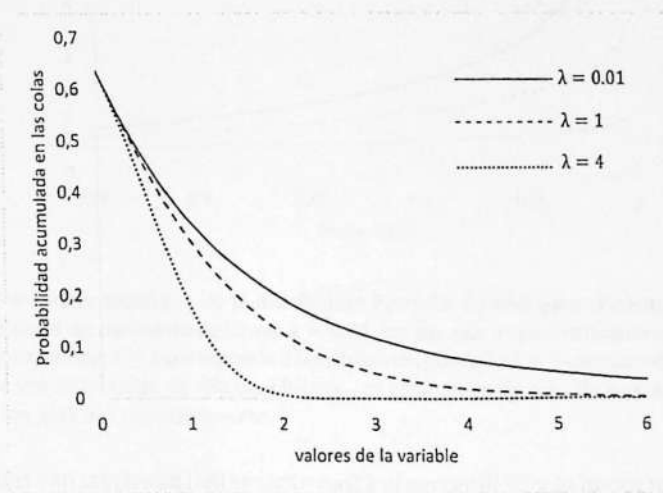


Figura 3.2.2: Función de distribución Potencial-Gumbel para diferentes valores del parámetro de forma: $\lambda = 0.01$ corresponde a una distribución de cola pesada, $\lambda = 1$ corresponde a la distribución Gumbel y $\lambda = 4$ corresponde a una distribución de cola muy liviana. Los parámetros de posición y escala son $\xi=0, \theta=1$ respectivamente.

Las diferencias observadas en las colas se reflejan nítidamente en los percentiles superiores de las distribuciones, en la figura siguiente se representaron los percentiles teóricos de distribuciones pertenecientes a la familia Potencial-Gumbel (y por lo tanto dentro del dominio Gumbel) para valores seleccionados del parámetro λ en comparación con los percentiles de la distribución Gumbel (correspondiente a $\lambda = 1$). En estos gráficos puede verse claramente las ventajas en precisión que puede ganarse al plantear éste modelo en vez de usar la distribución asintótica Gumbel.

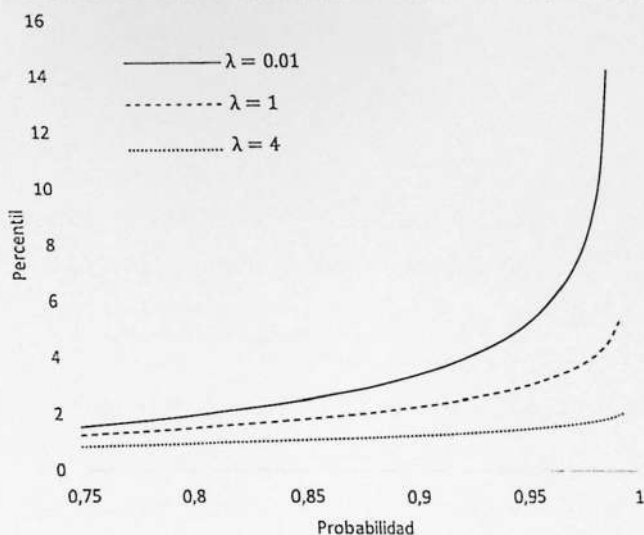


Figura 3.2.3: Percentiles superiores de la distribución Potencial-Gumbel para diferentes valores del parámetro de forma: $\lambda = 0.01$ corresponde a una distribución de cola pesada, $\lambda = 1$ corresponde a la distribución Gumbel y $\lambda = 4$ corresponde a una distribución de cola muy liviana. Los parámetros de posición y escala son $\xi=0$, $\theta=1$ respectivamente.

Las diferencias van creciendo lentamente hasta el percentil 95 y se hacen muy marcadas en los percentiles altos como ser 98 y más. Es claro, que por una cuestión de seguridad, muchas veces será recomendable plantear como modelo una distribución de cola más pesada que de algún modo tenderá a sobreestimar los percentiles superiores de la variable, especialmente si se trata con variables climáticas y se intenta prevenir desastres. Sin embargo, es muy importante no perder la precisión pues si nos volcamos por el lado de la seguridad ésta debe estar medida con razonable exactitud, y poder discriminar los percentiles de distribuciones dentro del dominio Gumbel puede ser muy valioso en este sentido.

El modelo propuesto aquí aspira a representar el dominio Gumbel con fines de estimación de los percentiles superiores, y por ello se habló de distribuciones de colas moderadamente pesadas, como la distribución Log-normal cuyo representante en el modelo propuesto sería de $\lambda \cong 0$ hasta distribuciones de colas muy livianas cuyo representante será de un valor de λ muy elevado. Esto nos da como una recorrida del dominio que se quiere representar y entonces es razonable pensar que cualquier otra distribución de este dominio tendrá un representante para algún valor de $\lambda > 0$ por ejemplo la Distribución Normal.

3.3. Aproximaciones a las Familias Normal, Lognormal, Weibull y Gamma.

La familia potencial Gumbel es un instrumento muy adecuado para aproximar las familias Normal, Lognormal, Weibull, Gamma, entre otras. Equivalentemente aplicando la transformación se puede aproximar las transformadas de estas familias mediante la familia Gumbel. Esta aproximación constituye el núcleo del método que se propone en este trabajo para acelerar la convergencia del máximo de una muestra aleatoria de una variable aleatoria así como mejorar la aproximación de los modelos estadísticos empleados en el análisis de los extremos, transformando los datos básicos de manera que su distribución se acerque a una distribución Gumbel. Ver Viollaz y Lazarte (2011).

Distribución Normal.

Considerando la razón de Mill, se tiene que: $\bar{F}(x) = \bar{\Phi}(x) \sim \varphi(x)/x$. Por lo tanto el "Hazard rate" o "tasa de mortalidad" correspondiente satisface:

$$h(x) = \frac{\varphi(x)}{\bar{\Phi}(x)} \sim x.$$

Por otra parte la tasa de riesgo de la distribución Potencial-Gumbel es asintóticamente igual a $k^{-1}x^{\lambda-1}$, por lo que teóricamente, tomado $\lambda = 2$ lograremos la mejor aproximación de la distribución normal por la distribución Potencial-Gumbel.

Distribución Lognormal.

Una variable aleatoria X tiene distribución lognormal si la variable $Y = \ln(X)$ tiene distribución normal $N(\mu, \sigma^2)$.

Entonces, tomando $\mu = 0$ y $\sigma^2 = 1$ resulta $F_X(x) = F_Y(\ln x)$, luego

$$f_X(x) = f_Y(\ln x) \frac{d(\ln x)}{dx}$$

y entonces la tasa de riesgo satisface:

$$h(x) = \frac{x^{-1}\varphi(\ln x)}{\bar{\Phi}(x)} \sim \frac{\ln x}{x}.$$

Esta función $h(x)$ está próxima a la frontera $\frac{1}{x}$ de las funciones h de la familia potencial Gumbel. Al realizar la estimación de los parámetros de la familia potencial Gumbel aproximante se obtendrá un valor próximo a cero para el exponente λ de la transformación.

Distribución Weibull.

En este caso, $\bar{F}(x) \sim \exp(-(\gamma x)^p)$ y $f(x) = \gamma p \exp(-(\gamma x)^p)(\gamma x)^{p-1}$, $\gamma > 0, 0 < p < \infty$. (Observe que aun que tenga el mismo nombre, no se trata de la distribución Weibull de extremos). Entonces la tasa de riesgo de esta distribución satisface:

$$h(x) = \gamma p (\gamma x)^{p-1}. \quad (3.3.1)$$

Por lo tanto tomando $\lambda = p$ se obtiene la mejor aproximación.

Distribución Gamma.

La densidad de la distribución Gamma está dada por:

$$f(x) = H\Gamma(\Gamma)^{-1}(HX)^{\Gamma-1} \exp(-HX) \quad \text{con } H \text{ y } \Gamma > 0. \quad (3.3.2)$$

Tomando límites para $x \rightarrow \infty$ y aplicando L'Hopital se tiene que:

$$\lim \frac{f(x)}{\bar{F}(x)} = \lim \frac{f'(x)}{-f(x)} = \eta.$$

Por lo tanto hay que tomar λ igual a uno (correspondiente a la transformación identidad) como era de esperar para aproximar una distribución a la distribución Gumbel, pues la distribución Gamma tiene colas equivalentes a la Gumbel.

Los ejemplos presentados sugieren que la familia potencial Gumbel es apta como modelo del Máximo Dominio de Atracción de la Distribución Gumbel.

Capítulo 4:

Estimación de Parámetros y Percentiles Superiores

En este capítulo, se trata con la elección correcta del modelo a utilizar así como con la estimación de sus parámetros. Para la estimación de parámetros, percentiles superiores de la distribución y de los periodos de recurrencia se usarán estadísticos de orden y un método propuesto aquí, que surge de la misma idea de transformar los datos para aproximarlos a la distribución Gumbel Estándar. Se parte de un breve repaso del tratamiento estadístico recomendado para datos de sucesos extremos utilizando el ejemplo de las velocidades máximas de vientos. Se menciona brevemente la estimación del modelo Gumbel por dos métodos diferentes, para luego proceder a la estimación de los parámetros de modelos dentro de la familia Potencial Gumbel y así lograr el objetivo final que es la estimación de los percentiles superiores de la distribución y los periodos de recurrencia. Por último, como ejemplo, se lleva a cabo una sencilla aplicación a series de velocidades máximas de vientos de la Rca. Argentina.

4.1. Introducción y Ejemplo

En los capítulos previos se trabajó en la formulación de la familia Potencial-Gumbel a partir de la distribución Gumbel y mediante la anti-transformada de la transformada Box-Cox modificada, con el fin de obtener modelos que se aproximen mejor, especialmente en las colas, a las distribuciones dentro del Dominio Gumbel. Se llegó a la conclusión que los miembros de esta familia serían más apropiados para ajustar distribuciones dentro del Dominio Atracción de la Distribución Gumbel que la misma asíntota, es decir, la distribución Gumbel. En este sentido diremos que esta familia representa al Dominio de Atracción de la Distribución Gumbel de un modo más eficaz que la propia distribución Gumbel, dado que admite distribuciones de diferentes colas (más livianas y más pesadas que la Gumbel) que se aproximan mejor a sus miembros. En este punto, se considera conveniente señalar que la elección óptima del parámetro de forma λ del modelo dentro de la familia Potencial-Gumbel y la estimación de los parámetros de posición y escala, ξ y θ respectivamente se realizan con el fin de estimar eficazmente los percentiles superiores de la distribución.

Selección del modelo

Dado que se trata de una familia paramétrica se pueden estimar los tres parámetros λ , ξ y θ simultáneamente, mediante métodos de estimación convencionales como ser máxima verosimilitud, y estadísticos de orden. Sin embargo se considera más conveniente utilizar estadísticos de orden dado que éste en nuestro caso nos permite apreciar la bondad del ajuste visualmente en el diagrama Q-QPlot generado en el proceso de estimación.

Recuerde que la transformación óptima a aplicar a los datos es:

$$Y = \frac{c}{\lambda} \left[\left(1 + \frac{X - \xi}{c\theta} \right)^\lambda - 1 \right]$$

donde la constante c puede tomar valores entre 2 y 5 dependiendo de la cantidad de datos y de su varianza, aquí se toma $c = 3$. Entonces en este caso, la variable transformada Y se supone con distribución Gumbel Estándar, es decir de parámetros 0 y 1. Esto es equivalente a ajustar a los datos el modelo propuesto en el Capítulo 3, es decir el Modelo Potencial-Gumbel.

Durante este capítulo se desarrollará como ejemplo, una sencilla aplicación a las series de vientos máximos mensuales utilizadas en todo este trabajo.

Ejemplo 4.1.1: Velocidades máximas del viento.

Una base de datos tratada reiteradas veces por el autor y su director, que generó publicaciones y presentación a congresos, es la de velocidades máximas mensuales de vientos en varias estaciones meteorológicas de Argentina. Estas bases contienen datos mensuales en un periodo de 24 años, es decir 288 datos por estación, y serán utilizadas en este capítulo para ejemplificar la aplicación de los modelos mencionados.

Como ya se dijo, el conjunto de datos contiene las velocidades máximas mensuales del viento, que naturalmente poseen una fuerte componente estacional intra-anual por lo que se estará "lejos" del supuesto de independencia. Para salvar este problema, se utilizará la técnica de tomar máximos anuales eliminando la dependencia intra-anual. Figuras 4.1.1 – 4.1.4. Este procedimiento es comúnmente utilizado en los métodos estadísticos de extremos ver referencia [2] página 284. En el caso de la estimación del modelo Gumbel, el procedimiento de tomar el máximo anual lleva a que la distribución de los datos se aproxime mejor a su distribución límite (Gumbel) basándose en el Teorema 1.3.2 del Capítulo 1 y por lo tanto a un mejor ajuste, siempre bajo el supuesto que los datos originales pertenecen al Dominio de Atracción Gumbel.

De este modo, es razonable suponer que los datos anuales x_1, \dots, x_{24} son observaciones de la sucesión de variables aleatorias X_1, \dots, X_{24} iid con Distribución de Extremos Gumbel.

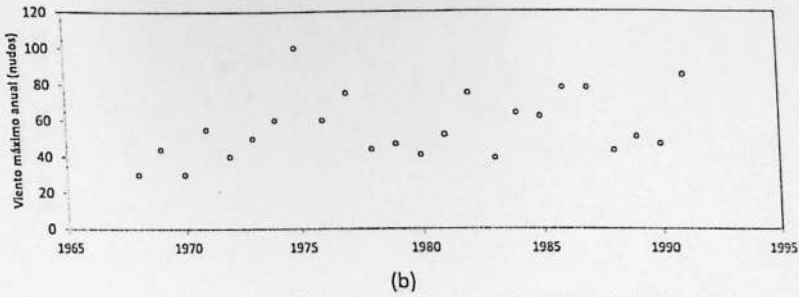
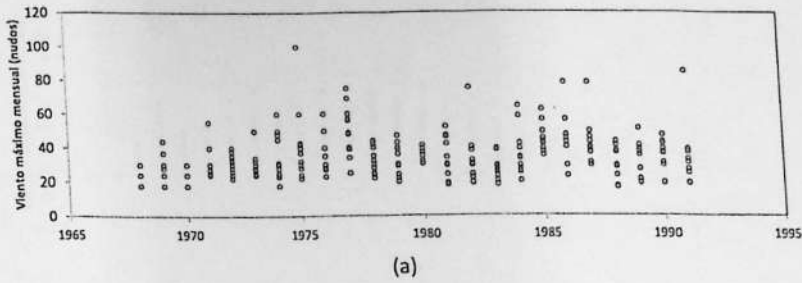


Figura 4.1.1: Estación Meteorológica Concordia, Argentina. (a) Velocidades máximas mensuales del viento. (b) Velocidades máximas anuales del viento.

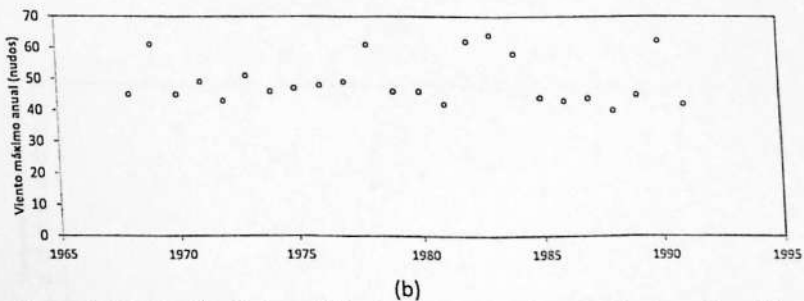
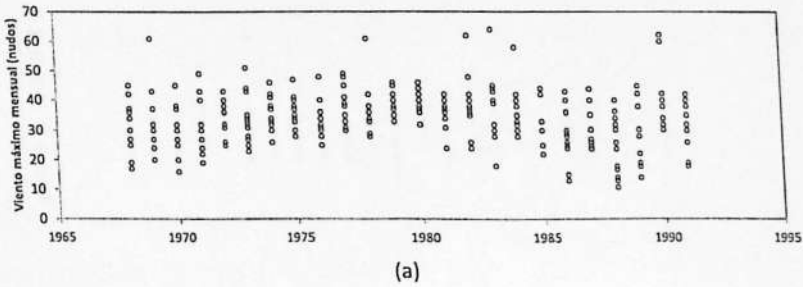
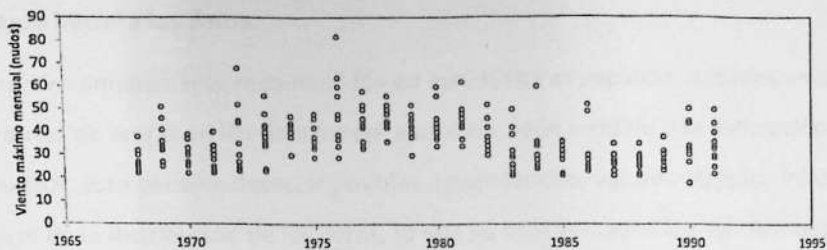
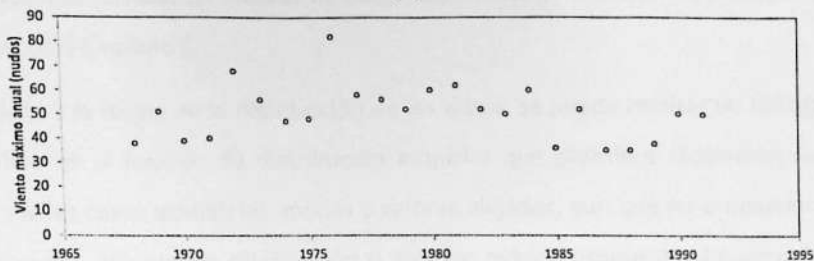


Figura 4.1.2: Estación Meteorológica Buenos Aires Observatorio, Argentina. (a) Velocidades máximas mensuales del viento. (b) Velocidades máximas anuales del viento.

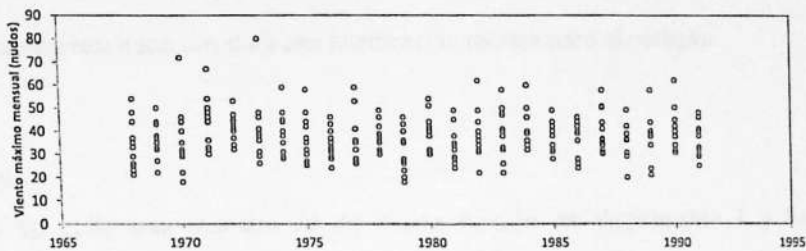


(a)

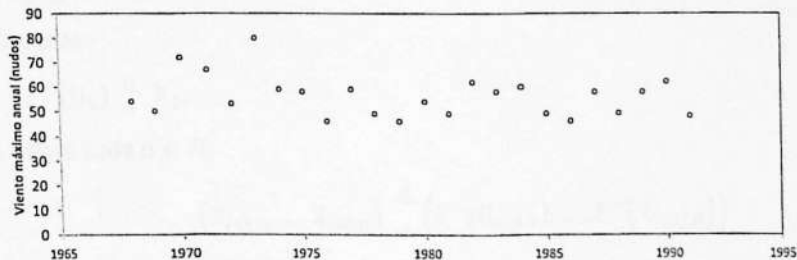


(b)

Figura 4.1.3: Estación Meteorológica Córdoba, Argentina. (a) Velocidades máximas mensuales del viento. (b) Velocidades máximas anuales del viento.



(a)



(b)

Figura 4.1.4: Estación Meteorológica Ezeiza, Buenos Aires, Argentina. (a) Velocidades máximas mensuales del viento. (b) Velocidades máximas anuales del viento.

Un vistazo inicial a los datos

Una práctica ampliamente recomendada en estadística es explorar cuidadosamente los datos antes de entrar en los detalles del ajuste de algún modelo y la estimación de los parámetros, esto permite detectar posibles dependencias, valores alejados influyentes y formas de la distribución de los datos, lo que se aplica totalmente en distribución de extremos. Las técnicas de análisis de datos específicas para extremos pueden verse en referencia[2] Capítulo 6.

En cuanto a la forma de la distribución de los datos, se puede realizar un histograma o el gráfico de la función de distribución empírica que permiten fácilmente visualizar propiedades como asimetrías, modas y valores alejados, aun que no proporcionan una comparación con alguna distribución o modelo teórico propuesto. En algunos casos, como el que se trata aquí, el problema se reduce a decidir si la distribución empírica de los datos se aproxima lo suficiente al modelo surgido de la teoría y por tal razón se destaca el gráfico conocido como Q-Qplot. También es posible utilizar un P-Pplot o algunos de sus variantes. A continuación se dará una breve descripción.

El siguiente resultado nos dará una justificación teórica para el método.

Lema:

Sean X_1, \dots, X_n una muestra iid de X con función de distribución F y llamemos $X_{(1),n}, \dots, X_{(n),n}$ a la muestra ordenada de menor a mayor. Sea U_1, \dots, U_n variables aleatorias iid con distribución uniforme en el intervalo $(0, 1)$ y denotemos por $U_{(1),n} < \dots < U_{(n),n}$ a sus correspondientes estadísticos de orden. Entonces valen las siguientes afirmaciones:

- $F^-(U_1) \stackrel{d}{=} X_1$.
- Para cada $n \in \mathbb{N}$,

$$(X_{(1),n}, \dots, X_{(n),n}) \stackrel{d}{=} (F^-(U_{(1),n}), \dots, F^-(U_{(n),n}))$$

- La variable aleatoria $F(X_1)$ tiene distribución uniforme sobre el intervalo $(0, 1)$ si y solo si F es una función continua.

Donde el símbolo $\stackrel{d}{=}$ significa "igual distribución" y F^{\leftarrow} es la inversa generalizada de una función monótona, definida por: $F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} \mid F(x) \geq t\}$.

Este lema dice entonces que si F es continua $U_i = F(X_i)$ para $i = 1, \dots, n$ son variables aleatorias iid con distribución uniforme en $(0, 1)$. Y por lo tanto se puede probar que:

$$E(F(X_{(j),n})) = \frac{j}{n+1} \text{ para } j = 1, \dots, n.$$

De este modo se puede inspeccionar el gráfico de los puntos

$$\left\{ \left(X_{(j),n}, F^{\leftarrow} \left(\frac{j}{n+1} \right) \right) \mid j = 1, \dots, n \right\}$$

Comúnmente llamado Q-Qplot para verificar si los datos están distribuidos según una función de distribución postulada F .

Sin embargo se recomienda utilizar una mejora, obtenida mediante corrección por continuidad, de este método que considera los puntos de la forma:

$$\left\{ \left(X_{(j),n}, F^{\leftarrow} \left(\frac{j-0.5}{n} \right) \right) \mid j = 1, \dots, n \right\}$$

Este método se usará más adelante para la estimación de parámetros en los modelos Gumbel y Potencial Gumbel.

En lo que sigue se realizará un repaso referido a la estimación del modelo Gumbel por máxima verosimilitud y por estadísticos de orden para luego entrar en la estimación del modelo Potencial-Gumbel que es más general.

4.2. Estimación de Parámetros en el Modelo Gumbel

En esta sección se desarrollan dos procedimientos de estimación de los parámetros, uno por máxima verosimilitud y otro basado en estadísticos de orden, para el caso del modelo Gumbel y del modelo Potencial-Gumbel propuesto aquí. En el caso de la estimación de los parámetros del Modelo-Potencial Gumbel deberán tenerse muy en cuenta las observaciones realizadas en el ítems Selección del Modelo de la sección 4.1. La teoría estándar de estimación de parámetros en general supone que los datos disponibles son una muestra

$$X_1, \dots, X_n \text{ iid de una variable aleatoria } X.$$

En este caso se supondrá que X tiene distribución Gumbel de parámetros ξ y θ , o Potencial-Gumbel de parámetros λ , ξ y θ .

Por un lado, se sabe que la suposición que las X_i tienen exactamente la distribución propuesta, en general no es del todo cierta. Por otro lado la hipótesis de independencia tampoco lo es, dado que en la mayoría de los casos, los datos disponibles son series cronológicas que siempre contienen datos correlacionados y con componentes estacionales.

En este trabajo ya se ha propuesto una mejoría para el primer problema referido a la verdadera distribución de los datos permitiendo que el modelo abarque otras distribuciones, es decir, ampliando el modelo Gumbel al Potencial-Gumbel.

Además, una metodología utilizada en hidrología y otras ciencias ambientales, que alivia estos dos problemas mencionados es la que utiliza el máximo anual (*Fitting of Annual Maxima*). En general los datos son mensuales y contienen componentes estacionales de periodo igual a un año entonces al tomar los máximos de periodos disjuntos iguales a un año se elimina este efecto aproximando los datos a la hipótesis de independencia, también por el teorema de convergencia del máximo, esta metodología producirá datos con distribución más próxima a su asíntota, en este caso la variable aleatoria Gumbel.

Tampoco hay que perder de vista, que al tomar el máximo anual se cuenta con menos datos para la estimación y así se tendrá mayor error en las estimaciones. Lo que se recomienda en la bibliografía en este sentido es tomar máximos de periodos tales que

se mantengan las propiedades de los estimadores de máxima verosimilitud, siempre que sea posible. Sin embargo históricamente se ha tomado el máximo anual.

Modelo Gumbel: Estimación por Máxima Verosimilitud.

Para el caso de la distribución Gumbel se tiene que la función de densidad de X es, para $-\infty < x < \infty$.

$$f_X(x) = \frac{1}{\theta} \exp\left(-\exp\left(-\frac{x-\xi}{\theta}\right)\right) \exp\left(-\frac{x-\xi}{\theta}\right).$$

Con ξ y θ tales que $-\infty < \xi < \infty$ y $\theta > 0$.

Luego si se cuenta con una muestra aleatoria X_1, \dots, X_n iid con función de distribución F_X , la función de verosimilitud será:

$l(\xi, \theta / x_1, \dots, x_n)$

$$= \frac{1}{\theta^n} \exp\left(-\sum_{i=1}^n \exp\left(-\frac{x_i-\xi}{\theta}\right)\right) \exp\left(-\sum_{i=1}^n \frac{x_i-\xi}{\theta}\right).$$

y así, la función log-verosimilitud será:

$$L(\lambda, \xi, \theta) = -n \ln(\theta) - \sum_{i=1}^n \exp\left(-\frac{x_i-\xi}{\theta}\right) - \sum_{i=1}^n \frac{x_i-\xi}{\theta}.$$

Derivando respecto de los 2 parámetros, se encuentra que las ecuaciones de verosimilitud para el caso Gumbel son:

$$\sum_{i=1}^n \left[\exp\left(-\frac{x_i-\xi}{\theta}\right) - 1 \right] = 0$$

$$\frac{1}{\theta} \sum_{i=1}^n \left[\exp\left(-\frac{x_i-\xi}{\theta}\right) - 1 \right] \left(\frac{x_i-\xi}{\theta} \right) + n = 0$$

No existe una solución explícita para estas ecuaciones, se deben recurrir a soluciones de optimización Numérica.

Modelo Gumbel: Estimación por estadísticos de orden.

Una Manera gráfica de estimar los parámetros de la distribución Gumbel, propuesta en la bibliografía para el caso de familias de posición y escala, es realizando el gráfico de los datos ordenados transformados en función de los valores esperados de la distribución Gumbel estándar (Q-Qplot). Los valores de los parámetros ξ y θ , se pueden estimar respectivamente, como la ordenada al origen y la pendiente de la recta ajustada por mínimos cuadrados.

Si la variable aleatoria X tiene distribución Gumbel, y X_1, \dots, X_n son iid como X , entonces se cumple que $F_X(x_{(j)}) \cong j/(n+1)$. Se han propuesto muchas mejoras de este procedimiento como ser

$$\exp\left(-\exp\left(-\frac{x_{(j)} - \xi}{\theta}\right)\right) \cong \frac{j - 0.5}{n}$$

que resulta en

$$x_{(j)} \cong \xi - \theta \ln\left(-\ln\left(\frac{j - 0.5}{n}\right)\right)$$

Graficando $x_{(j)}$ vs $-\ln\left(-\ln\left(\frac{j - 0.5}{n}\right)\right)$ se ajusta la recta por mínimos cuadrados para obtener los valores de los estimadores y sus correspondientes desviaciones estándar.

Para una elección apropiada de la constante c , tal que $1 + \frac{x-\xi}{c\theta} > 0$, que es la región que nos importa, tendremos:

$$g'(x) = \frac{1}{\theta} \left(1 + \frac{x-\xi}{c\theta} \right)^{\lambda-1}$$

(Nota: no hay problemas para hacer la derivada por ramas, pero la otra rama no es de interés dado que por un lado no acumula una probabilidad significativa y por otro lado sólo tiene importancia el lado derecho de la distribución)

y a función de verosimilitud estará dada por:

$$l(\xi, \theta, \lambda) = \prod_{i=1}^n f_X(x_i).$$

$$l(\xi, \theta, \lambda) = \prod_{i=1}^n [\exp(-\exp(-g(x_i))) \exp(-g(x_i)) g'(x_i)].$$

y denotemos por

$$L(\xi, \theta, \lambda) = \ln l(\xi, \theta, \lambda).$$

Entonces

$$L(\xi, \theta, \lambda) = - \sum_{i=1}^n [\exp(-g(x_i)) + g(x_i) - \ln(g'(x_i))]$$

El estimador de máxima verosimilitud $(\hat{\xi}, \hat{\theta}, \hat{\lambda})$ es el argumento que maximiza la función $L(\xi, \theta, \lambda)$ y que sería la solución del sistema:

$$\begin{cases} \frac{\partial L(\xi, \theta, \lambda)}{\partial \xi} = - \sum_{i=1}^n \left[- \exp(-g(x_i)) \frac{\partial g(x_i)}{\partial \xi} + \frac{\partial g(x_i)}{\partial \xi} - \frac{1}{g'(x_i)} \frac{\partial g'(x_i)}{\partial \xi} \right] = 0 \\ \frac{\partial L(\xi, \theta, \lambda)}{\partial \theta} = - \sum_{i=1}^n \left[- \exp(-g(x_i)) \frac{\partial g(x_i)}{\partial \theta} + \frac{\partial g(x_i)}{\partial \theta} - \frac{1}{g'(x_i)} \frac{\partial g'(x_i)}{\partial \theta} \right] = 0 \\ \frac{\partial L(\xi, \theta, \lambda)}{\partial \lambda} = - \sum_{i=1}^n \left[- \exp(-g(x_i)) \frac{\partial g(x_i)}{\partial \lambda} + \frac{\partial g(x_i)}{\partial \lambda} - \frac{1}{g'(x_i)} \frac{\partial g'(x_i)}{\partial \lambda} \right] = 0 \end{cases}$$

No existen soluciones explícitas para estos estimadores pero se pueden obtener sin mayores problemas por métodos de aproximación numérica. Por ejemplo, utilizando la aplicación "Mata" del programa Stata11 se calcularon estos estimadores para el ejemplo de series de vientos máximos. Los resultados de las estimaciones de los parámetros λ , ξ y θ se presentan en la tabla 4.4.1 y 4.4.2.

Se estiman las varianzas asintóticas de los estimadores mediante la inversa de la matriz de información de Fisher y por la normalidad asintótica de los estimadores de máxima verosimilitud* se calcularon los intervalos de confianza aproximados para estos parámetros.

Para obtener la matriz de información de Fisher y las varianzas asintóticas de los estimadores de máxima verosimilitud se calcularon las derivadas segundas:

$$\frac{\partial^2 L(\xi, \theta, \lambda)}{\partial \xi^2} = - \sum_{i=1}^n \left[- \exp(-g(x_i)) \left(\frac{\partial g(x_i)}{\partial \xi} \right)^2 - \exp(-g(x_i)) \frac{\partial^2 g(x_i)}{\partial \xi^2} + \frac{\partial^2 g(x_i)}{\partial \xi^2} - \frac{\frac{\partial^2 g'(x_i)}{\partial \xi^2} - \left(\frac{\partial g'(x_i)}{\partial \xi} \right)^2}{(g'(x_i))^2} \right]$$

$$\frac{\partial^2 L(\xi, \theta, \lambda)}{\partial \xi \partial \theta} = - \sum_{i=1}^n \left[- \exp(-g(x_i)) \frac{\partial g(x_i)}{\partial \xi} \frac{\partial g(x_i)}{\partial \theta} - \exp(-g(x_i)) \frac{\partial^2 g(x_i)}{\partial \xi \partial \theta} + \frac{\partial^2 g(x_i)}{\partial \xi \partial \theta} \right. \\ \left. - \frac{\frac{\partial^2 g'(x_i)}{\partial \xi \partial \theta} - \frac{\partial g'(x_i)}{\partial \xi} \frac{\partial g'(x_i)}{\partial \theta}}{(g'(x_i))^2} \right]$$

$$\frac{\partial^2 L(\xi, \theta, \lambda)}{\partial \xi \partial \lambda} = - \sum_{i=1}^n \left[- \exp(-g(x_i)) \frac{\partial g(x_i)}{\partial \xi} \frac{\partial g(x_i)}{\partial \lambda} - \exp(-g(x_i)) \frac{\partial^2 g(x_i)}{\partial \xi \partial \lambda} + \frac{\partial^2 g(x_i)}{\partial \xi \partial \lambda} \right. \\ \left. - \frac{\frac{\partial^2 g'(x_i)}{\partial \xi \partial \lambda} - \frac{\partial g'(x_i)}{\partial \xi} \frac{\partial g'(x_i)}{\partial \lambda}}{(g'(x_i))^2} \right]$$

$$\frac{\partial^2 L(\xi, \theta, \lambda)}{\partial \theta^2} = - \sum_{i=1}^n \left[- \exp(-g(x_i)) \left(\frac{\partial g(x_i)}{\partial \theta} \right)^2 - \exp(-g(x_i)) \frac{\partial^2 g(x_i)}{\partial \theta^2} + \frac{\partial^2 g(x_i)}{\partial \theta^2} \right. \\ \left. - \frac{\frac{\partial^2 g'(x_i)}{\partial \theta^2} - \left(\frac{\partial g'(x_i)}{\partial \theta} \right)^2}{(g'(x_i))^2} \right]$$

$$\frac{\partial^2 L(\xi, \theta, \lambda)}{\partial \theta \partial \lambda} = - \sum_{i=1}^n \left[- \exp(-g(x_i)) \frac{\partial g(x_i)}{\partial \theta} \frac{\partial g(x_i)}{\partial \lambda} - \exp(-g(x_i)) \frac{\partial^2 g(x_i)}{\partial \theta \partial \lambda} + \frac{\partial^2 g(x_i)}{\partial \theta \partial \lambda} \right. \\ \left. - \frac{\frac{\partial^2 g'(x_i)}{\partial \theta \partial \lambda} - \frac{\partial g'(x_i)}{\partial \theta} \frac{\partial g'(x_i)}{\partial \lambda}}{(g'(x_i))^2} \right]$$

$$\frac{\partial^2 L(\xi, \theta, \lambda)}{\partial \lambda^2} = - \sum_{i=1}^n \left[- \exp(-g(x_i)) \left(\frac{\partial g(x_i)}{\partial \lambda} \right)^2 - \exp(-g(x_i)) \frac{\partial^2 g(x_i)}{\partial \lambda^2} + \frac{\partial^2 g(x_i)}{\partial \lambda^2} \right. \\ \left. - \frac{\frac{\partial^2 g'(x_i)}{\partial \lambda^2} - \left(\frac{\partial g'(x_i)}{\partial \lambda} \right)^2}{(g'(x_i))^2} \right]$$

Aplicación a las Series de Vientos Máximos de Argentina

Estimación de Parámetros

Los valores de los parámetros del modelo Potencial-Gumbel estimados, en nudos, para 20 estaciones se presentan en las siguientes tablas, acompañados de sus respectivas desviaciones estándar e intervalos de confianza de 95%, aproximados.

Tabla 4.4.1: Valores estimados del parámetro de forma λ con sus desviaciones estándar e intervalos de 95% de confianza.

Estación	$\hat{\lambda}$	Desvío estándar	Intervalo de 95%
9 de Julio	1,20	0,48	(0,26, 2,14)
Aeroparque	1,13	0,50	(0,15, 2,11)
Bariloche	0,28	0,63	(-0,96, 1,52)
Concordia	0,96	0,56	(-0,13, 2,05)
Córdoba	1,13	0,64	(-0,13, 2,38)
Corrientes	1,62	0,47	(0,70, 2,54)
Esquel	1,40	0,44	(0,55, 2,26)
Ezeiza	0,52	0,59	(-0,64, 1,68)
Formosa	1,71	0,54	(0,66, 2,77)
Gualetgüaichú	1,28	0,50	(0,30, 2,26)
Laboulaye	1,28	0,58	(0,14, 2,43)
Lago Argentino	0,94	0,63	(-0,30, 2,18)
Malargüe	1,13	0,59	(-0,03, 2,29)
Mar del Plata	0,55	0,59	(-0,61, 1,71)
Monte Caseros	1,02	0,63	(-0,22, 2,26)
Paraná	0,47	0,52	(-0,55, 1,49)
Resistencia	1,75	0,55	(0,67, 2,82)
Salta	1,00	0,58	(-0,15, 2,14)
Santa Rosa	0,37	0,52	(-0,65, 1,38)
Villa Dolores	1,18	0,62	(-0,05, 2,40)

Tabla 4.4.2: Valores estimados de los parámetros ξ y θ con sus desviaciones estándar e intervalos de 95% de confianza.

Estación	ξ (nudos)	Desv. est(ξ)	Int. de 95%	θ (nudos)	Desv. est(θ)	Int. de 95%
9 de Julio	44,19	1,78	(40,70, 47,68)	7,70	1,19	(5,36, 10,03)
Aeroparque	51,50	1,86	(47,85, 55,15)	8,03	1,25	(5,58, 10,49)
Bariloche	61,35	2,33	(56,79, 65,91)	8,70	2,05	(4,69, 12,71)
Concordia	47,97	3,30	(41,50, 54,44)	14,07	2,33	(9,50, 18,64)
Córdoba	45,58	2,26	(41,15, 50,01)	9,37	1,53	(6,38, 12,36)
Corrientes	53,34	2,58	(48,29, 58,40)	11,30	1,79	(7,80, 14,80)
Esquel	85,89	2,00	(81,97, 89,80)	8,74	1,37	(6,06, 11,42)
Ezeiza	51,81	1,42	(49,02, 54,59)	5,86	1,14	(3,63, 8,10)
Formosa	48,94	1,62	(45,75, 52,12)	7,01	1,10	(4,85, 9,17)
Gualeduaichú	43,55	2,19	(39,27, 47,84)	9,47	1,48	(6,56, 12,38)
Laboulaye	54,09	1,88	(50,40, 57,78)	7,98	1,26	(5,51, 10,44)
Lago Argentino	48,76	1,99	(44,85, 52,67)	8,33	1,43	(5,52, 11,14)
Malargue	69,11	3,43	(62,39, 75,83)	14,44	2,32	(9,89, 18,99)
Mar del Plata	54,70	2,05	(50,69, 58,72)	8,50	1,62	(5,34, 11,67)
Monte Caseros	56,15	2,02	(52,19, 60,10)	8,45	1,43	(5,64, 11,26)
Paraná	58,14	1,94	(54,33, 61,95)	8,27	1,51	(5,31, 11,24)
Resistencia	55,17	1,71	(51,81, 58,52)	7,40	1,17	(5,10, 9,69)
Salta	41,83	1,29	(39,31, 44,36)	5,43	0,89	(3,69, 7,17)
Santa Rosa	49,65	3,00	(43,76, 55,54)	12,71	2,47	(7,88, 17,54)
Villa Dolores	49,70	1,36	(47,05, 52,36)	5,67	0,91	(3,88, 7,46)

Estimación de los Percentiles Superiores

Nuevamente, acudiendo a la normalidad asintótica de los estimadores de máxima verosimilitud se aplica el Método Delta para construir intervalos de confianza aproximados para los percentiles superiores de la distribución, que son funciones de los parámetros estimados.

Sea p tal que $0 \leq p \leq 1$, entonces el percentil de orden p , x_p cumple que:

$$P(X \leq x_p) = p,$$

luego

$$\exp\left(-\exp\left(-\frac{c}{\lambda}\left[\left(1 + \frac{x_p - \xi}{c\theta}\right)^\lambda - 1\right]\right)\right) = p$$

y así tenemos que:

$$x_p = \xi + c\theta \left[\left(1 - \frac{\lambda}{c} \ln(-\ln(p))\right)^{1/\lambda} - 1 \right]$$

Luego utilizando los valores estimados de los parámetros ξ y θ obtenemos los estimadores de los percentiles mediante la fórmula:

$$\hat{x}_p = \hat{\xi} + c\hat{\theta} \left[\left(1 - \frac{\hat{\lambda}}{c} \ln(-\ln(p))\right)^{1/\hat{\lambda}} - 1 \right]$$

Las velocidades de recurrencia correspondientes a un periodo de T años se obtienen mediante la siguiente expresión:

$$\hat{x}_p = \hat{\xi} + c\hat{\theta} \left[\left(1 - \frac{\hat{\lambda}}{c} \ln\left(-\ln\left(1 - \frac{1}{T+1}\right)\right)\right)^{1/\hat{\lambda}} - 1 \right]$$

Entonces si se aplica el Método Delta a la función:

$$\hat{x}_p = \xi + c \hat{\theta} \left[\left(1 - \frac{\hat{\lambda}}{c} \ln(-\ln(p)) \right)^{1/\lambda} - 1 \right]$$

El intervalo aproximado será

$$\left(\hat{x}_p - z_{\alpha} \hat{D}, \hat{x}_p + z_{\alpha} \hat{D} \right)$$

Donde

$$\hat{D}^2 = \nabla_{x_p}(\xi, \theta, \lambda)^t \Sigma \nabla_{x_p}(\xi, \theta, \lambda)$$

Y

$$\nabla_{x_p}(\xi, \theta, \lambda) = \left(\frac{\partial \hat{x}_p}{\partial \xi}, \frac{\partial \hat{x}_p}{\partial \theta}, \frac{\partial \hat{x}_p}{\partial \lambda} \right)$$

Finalmente las derivadas de las velocidades de recurrencia con respecto de los parámetros serán:

$$\begin{cases} \frac{\partial \hat{x}_p}{\partial \xi} = 1 \\ \frac{\partial \hat{x}_p}{\partial \theta} = c \left[\left(1 - \frac{\lambda}{c} \ln(-\ln(p)) \right)^{1/\lambda} - 1 \right] \\ \frac{\partial \hat{x}_p}{\partial \lambda} = \left(1 - \frac{\lambda}{c} \ln(-\ln(p)) \right)^{1/\lambda} \left\{ \frac{1}{\lambda^2} \ln \left(1 - \frac{\lambda}{c} \ln(-\ln(p)) \right) + \frac{\frac{\ln(-\ln(p))}{c}}{\lambda \left(1 - \frac{\lambda}{c} \ln(-\ln(p)) \right)} \right\} \end{cases}$$

Se aplicó a las series de velocidades máximas anuales del viento.

En las Tablas siguientes se presentan los valores estimados de los percentiles 0,50; 0,98 y 0,995 correspondientes a las velocidades de vientos con periodo de recurrencia de 2, 50 y 200 años respectivamente.

Tabla 4.4.3: Valores estimados (y sus desviaciones estándar estimadas) de los percentiles 0,50 correspondientes a las velocidades de vientos, en nudos, con periodo de recurrencia de 2 años.

Estación	Velocidad (Nudos)	Desv. Est.	Int. 95%
9 de Julio	46,97	1,89	(43,27, 50,68)
Aeroparque	54,42	2,00	(50,50, 58,34)
Bariloche	64,68	2,85	(59,10, 70,27)
Concordia	53,14	3,65	(45,98, 60,29)
Córdoba	48,99	2,45	(44,19, 53,79)
Corrientes	57,34	2,62	(52,20, 62,47)
Esquel	89,02	2,07	(84,96, 93,07)
Ezeiza	54,02	1,68	(50,73, 57,31)
Formosa	51,40	1,63	(48,21, 54,60)
Gualectuaichú	46,97	2,31	(42,44, 51,49)
Laboulaye	56,96	1,99	(53,06, 60,86)
Lago Argentino	51,82	2,22	(47,47, 56,17)
Malargue	74,36	3,71	(67,09, 81,63)
Mar del Plata	57,91	2,40	(53,20, 62,61)
Monte Caseros	59,24	2,23	(54,87, 63,61)
Paraná	61,27	2,29	(56,78, 65,76)
Resistencia	57,76	1,72	(54,39, 61,14)
Salta	43,83	1,42	(41,04, 46,61)
Santa Rosa	54,49	3,60	(47,43, 61,54)
Villa Dolores	51,76	1,46	(48,90, 54,62)

Tabla 4.4.4: Valores estimados (y sus desviaciones estándar estimadas) de los percentiles 0,98 correspondientes a las velocidades de vientos, en nudos, con periodo de recurrencia de 50 años.

Estación	Velocidad (Nudos)	Desv. Est.	Int. 95%
9 de Julio	71,62	6,88	(58,13, 85,10)
Aeroparque	81,02	7,55	(66,22, 95,82)
Bariloche	114,38	25,55	(64,30, 164,46)
Concordia	103,97	15,89	(72,82, 135,11)
Córdoba	80,10	9,49	(61,50, 98,70)
Corrientes	87,68	7,23	(73,51, 101,85)
Esquel	114,65	6,51	(101,89, 127,41)
Ezeiza	81,76	11,00	(60,20, 103,32)
Formosa	69,59	4,47	(60,83, 78,35)
Gualectuachú	76,28	7,88	(60,83, 91,72)
Laboulaye	81,57	6,98	(67,89, 95,26)
Lago Argentino	82,18	10,02	(62,54, 101,82)
Malargue	122,25	14,29	(94,24, 150,26)
Mar del Plata	97,19	15,37	(67,06, 127,31)
Monte Caseros	88,75	9,38	(70,36, 107,13)
Paraná	101,76	15,93	(70,53, 132,98)
Resistencia	76,71	4,57	(67,75, 85,66)
Salta	63,07	5,99	(51,33, 74,81)
Santa Rosa	122,00	28,23	(66,67, 177,33)
Villa Dolores	70,12	5,45	(59,44, 80,80)

Tabla 4.4.5: Valores estimados (y sus desviaciones estándar estimadas) de los percentiles 0,995 correspondientes a las velocidades de vientos, en nudos, con periodo de recurrencia de 200 años.

Estación	Velocidad (Nudos)	Desv. Est.	Int. 95%
9 de Julio	80,61	10,39	(60,25, 100,98)
Aeroparque	90,99	11,67	(68,12, 113,86)
Bariloche	144,80	57,51	(32,08, 257,52)
Concordia	124,36	26,01	(73,38, 175,34)
Córdoba	91,78	15,24	(61,90, 121,65)
Corrientes	97,45	10,09	(77,67, 117,22)
Esquel	123,42	9,36	(105,08, 141,77)
Ezeiza	95,83	21,50	(53,69, 137,97)
Formosa	75,31	6,29	(62,99, 87,64)
Gualeguaiichú	86,72	11,77	(63,65, 109,79)
Laboulaye	90,32	10,67	(69,40, 111,23)
Lago Argentino	94,45	16,82	(61,48, 127,41)
Malargue	140,20	22,63	(95,85, 184,56)
Mar del Plata	116,68	29,49	(58,88, 174,48)
Monte Caseros	100,27	15,39	(70,10, 130,43)
Paraná	122,98	31,07	(62,08, 183,87)
Resistencia	82,62	6,40	(70,07, 95,16)
Salta	70,68	9,78	(51,51, 89,84)
Santa Rosa	160,36	58,53	(45,64, 275,08)
Villa Dolores	76,88	8,60	(60,02, 93,73)

Pruebas por Simulaciones

Se generó 1000 muestras de máximos de la siguiente manera:

Cada muestra se obtiene mediante el procedimiento:

- Se genera una matriz de 12x24 de una distribución Normal(50, 5²).
- Se toma el máximo de cada columna (12 datos), es decir, se obtiene una muestra de máximos de tamaño 24.
- Es igual al ejemplo de las series de vientos.

Para cada una de estas 1000 muestras

- Se calcularon los intervalos de confianza de 95% para los percentiles $x_{0,5}$, $x_{0,98}$ y $x_{0,995}$ utilizando el modelo propuesto Potencial Gumbel y el Modelo Gumbel.
- Se calcularon las longitudes de los intervalos.
- Se calculó el percentil exacto según el modelo que generó los datos y la función que toma el valor uno si este pertenece al intervalo calculado y cero si no pertenece.
- Se estimó la longitud media de los intervalos de confianza y los niveles de confianza para cada percentil y para ambos modelos.

Basado en 1000 muestras de tamaño 24.

	Estimaciones	Modelo Gumbel	Modelo Potencial-Gumbel
Percentil $x_{0,50}$	Amplitud media	2,291	2,263
	Nivel de confianza	0,912	0,928
Percentil $x_{0,98}$	Amplitud media	6,705	7,725
	Nivel de confianza	0,914	0,948
Percentil $x_{0,995}$	Amplitud media	8,690	12,341
	Nivel de confianza	0,814	0,958

Para el tamaño de muestra igual a 24 datos tenemos:

- Para el percentil 0,50 las amplitudes medias y los niveles de confianza son similares.
- Para el percentil 0,98 la amplitud media de los intervalos es menor para el modelo Gumbel y el nivel de confianza es mayor para el modelo Potencial-Gumbel.
- Para el percentil 0,995 la amplitud media de los intervalos es mucho menor para el modelo Gumbel, pero el nivel de confianza se mantiene cerca de 95 para el modelo Potencial-Gumbel.
- Es razonable que las amplitudes sean mayores para el modelo Potencial-Gumbel dado que se agrega un parámetro más para estimar y sólo son 24 datos. Sin embargo cuando estimo un percentil más elevado el modelo potencial Gumbel mantiene la confiabilidad.

Se realizó el mismo procedimiento para el caso de contar con 50 datos para ver el comportamiento de los modelos. (Siguen siendo muestras de máximos tomados sobre 12 datos)

Basado en 1000 muestras de tamaño 50.

	Estimaciones	Modelo Gumbel	Modelo Potencial-Gumbel
Percentil $x_{0,50}$	Amplitud media	1,62	1,60
	Nivel de confianza	0,896	0,936
Percentil $x_{0,98}$	Amplitud media	4,72	5,00
	Nivel de confianza	0,686	0,954
Percentil $x_{0,995}$	Amplitud media	6,11	7,52
	Nivel de confianza	0,450	0,956

Para este caso de muestras de tamaño 50 tenemos:

- Para los percentiles 0,50 y 0,95 las amplitudes son muy similares y para el percentil 0,995 la amplitud media es un poco mayor en el caso Potencial-Gumbel.
- En todos los casos, el modelo Potencial-Gumbel es muy superior en cuanto a la confiabilidad.
- Esto pasa por que al tener mayor cantidad de datos las amplitudes disminuyen pero en el caso Gumbel muchas estimaciones caen fuera del intervalo.

Se repite el mismo procedimiento para el caso de contar con 100 datos para ver el comportamiento de los modelos. (Siguen siendo muestras de máximos tomados sobre 12 datos)

Basado en 1000 muestras de tamaño 100.

	Estimaciones	Modelo Gumbel	Modelo Potencial-Gumbel'
Percentil $x_{0,50}$	Amplitud media	1,168	1,144
	Nivel de confianza	0,898	0,952
Percentil $x_{0,98}$	Amplitud media	3,381	3,410
	Nivel de confianza	0,322	0,976
Percentil $x_{0,995}$	Amplitud media	4,37	4,98
	Nivel de confianza	0,11	0,978

Nuevamente se observa que el Modelo Potencial-Gumbel es muy superior para estimar los percentiles.

Mas Simulaciones

Se generó 1000 muestras de máximos de la siguiente manera:

Cada muestra se obtiene mediante el procedimiento:

- Se genera una matriz de 12x25 de una distribución Normal(40, 10²). Ahora tienen mayor varianza.
- Se toma el máximo de cada columna (12 datos), es decir, se obtiene una muestra de máximos de tamaño 25.

Para cada una de estas 1000 muestras

- Se calcularon los intervalos de confianza de 95% para los percentiles $x_{0,5}$, $x_{0,98}$ y $x_{0,995}$ utilizando el modelo propuesto Potencial Gumbel y el Modelo Gumbel.
- Se calcularon las longitudes de los intervalos.
- Se calculó el percentil exacto según el modelo que generó los datos y la función que toma el valor uno si este pertenece al intervalo calculado y cero si no pertenece.
- Se estimó la longitud media de los intervalos de confianza y los niveles de confianza para cada percentil y para ambos modelos.

Basado en 1000 muestras de tamaño 25.

	Estimaciones	Modelo Gumbel	Modelo Potencial-Gumbel
Percentil $x_{0,50}$	Amplitud media	4,47	4,44
	Nivel de confianza	0,9238	0,9338
Percentil $x_{0,98}$	Amplitud media	13,10	15,16
	Nivel de confianza	0,9125	0,9363
Percentil $x_{0,995}$	Amplitud media	16,98	24,29
	Nivel de confianza	0,8025	0,9413

Para el tamaño de muestra igual a 24 datos tenemos:

- Para el percentil 0,50 las amplitudes medias y los niveles de confianza son similares.
- Para el percentil 0,98 la amplitud media de los intervalos es menor para el modelo Gumbel y el nivel de confianza es mayor para el modelo Potencial-Gumbel.
- Para el percentil 0,995 la amplitud media de los intervalos es mucho menor para el modelo Gumbel, pero el nivel de confianza se mantiene cerca de 95 para el modelo Potencial-Gumbel.
- Es razonable que las amplitudes sean mayores para el modelo Potencial-Gumbel dado que se agrega un parámetro más para estimar y sólo son 24 datos. Sin embargo cuando estimo un percentil más elevado el modelo potencial Gumbel mantiene la confiabilidad.

Se realizó el mismo procedimiento para el caso de contar con 50 datos para ver el comportamiento de los modelos. (Siguen siendo muestras de máximos tomados sobre 12 datos)

Basado en 1000 muestras de tamaño 50.

	Estimaciones	Modelo Gumbel	Modelo Potencial-Gumbel
Percentil $x_{0,50}$	Amplitud media	3,25	3,20
	Nivel de confianza	0,9087	0,9425
Percentil $x_{0,98}$	Amplitud media	9,47	10,01
	Nivel de confianza	0,6938	0,9487
Percentil $x_{0,995}$	Amplitud media	12,27	15,05
	Nivel de confianza	0,4313	0,9550

Para este caso de muestras de tamaño 50 tenemos:

- Para los percentiles 0,50 y 0,95 las amplitudes son muy similares y para el percentil 0,995 la amplitud media es un poco mayor en el caso Potencial-Gumbel.
- En todos los casos, el modelo Potencial-Gumbel es muy superior en cuanto a la confiabilidad.
- Esto pasa por que al tener mayor cantidad de datos las amplitudes disminuyen pero en el caso Gumbel muchas estimaciones caen fuera del intervalo.

Se repite el mismo procedimiento para el caso de contar con 100 datos para ver el comportamiento de los modelos. (Siguen siendo muestras de máximos tomados sobre 12 datos)

Basado en 1000 muestras de tamaño 100.

	Estimaciones	Modelo Gumbel	Modelo Potencial-Gumbel
Percentil $x_{0,50}$	Amplitud media	2,34	2,29
	Nivel de confianza	0,8975	0,9525
Percentil $x_{0,98}$	Amplitud media	6,78	6,84
	Nivel de confianza	0,315	0,9725
Percentil $x_{0,995}$	Amplitud media	8,78	10,03
	Nivel de confianza	0,0875	0,9775

Nuevamente se observa que el Modelo Potencial-Gumbel es muy superior para estimar los percentiles.

Capítulo 5

Mapa de Vientos de Argentina

Este Capítulo está dedicado a la construcción de las curvas iso-percentiles de las velocidades máximas de vientos en las diferentes estaciones meteorológicas de la República Argentina. Para la construcción del mapa es necesario considerar todos los datos disponibles, 72 estaciones meteorológicas, para cubrir lo mejor posible el territorio nacional. Se utiliza el modelo Gumbel, por ser el más sencillo, para estimar velocidades de vientos con periodos de recurrencia de 2, 50 y 200 años. Para verificar el modelo adoptado, se han representado las observaciones en función del tiempo de lo cual se deduce que hay varias estaciones con apartamientos de la estacionariedad y homogeneidad. Se han realizado los Q-QPlot para chequear el ajuste al modelo Gumbel. La construcción de los mapas con las isolíneas de las velocidades de recurrencia se realizó con el Software Surfer 6.0 (1995) (Surface Mapping System). Este software construye las isolíneas utilizando el método de interpolación lineal basado en una superficie poliédrica de caras triangulares que pasa por los puntos-datos. Sólo se muestran las figuras para el caso de velocidades del viento, en metros por segundo, con periodo de recurrencia de 50 años.

5.1: Introducción

En esta sección se presenta un trabajo previo a la aplicación de la transformada de Box-Cox para mejorar el ajuste del modelo Gumbel a los datos, el principal interés está en la presentación de la información estadística ubicada espacialmente en todo el territorio nacional. Esto se hace mediante un mapa de curvas de nivel, en otras palabras, el objetivo es la estimación de las distribuciones de probabilidades de las velocidades máximas anuales del viento para todo el territorio nacional y la presentación en la forma de mapas de curvas de nivel de los percentiles correspondientes a velocidades del viento con periodo de recurrencia de 50 años.

La información básica para el trazado de las curvas iso-percentiles consiste de los registros de velocidades máximas mensuales del viento correspondiente a 82 estaciones meteorológicas de la Rca. Argentina, provistas por el Servicio Meteorológico Nacional desde el año 1967 al año 1991. Estos registros contienen las velocidades máximas mensuales a partir de las cuales se calcularon las velocidades máximas anuales. Al igual que antes, se decidió tomar como dato básico al máximo anual para simplificar el problema de datos faltantes y para emplear una metodología con la cual los especialistas en el tema están más familiarizados.

Del conjunto de series, se eligieron aquellas cuyos registros fueron obtenidos en su mayoría con "anemocinemógrafo instantáneo" o "anemocinemógrafo totalizador", habiéndose seleccionado en total 72 estaciones.

El modelo Gumbel:

Sean X_1, \dots, X_n las velocidades máximas anuales correspondientes a n años observados en una estación dada. Se realizan las siguientes hipótesis:

- i) Las variables aleatorias X_1, \dots, X_n son idénticamente distribuidas, es decir la serie de observaciones es estacionaria.
- ii) Las variables X_1, \dots, X_n son independientes.
- iii) La distribución de probabilidad de cada X_i es Gumbel con parámetros que dependen de la estación considerada.

Es claro que estas hipótesis serán satisfechas sólo aproximadamente por las observaciones reales. En particular, el ciclo solar de una duración aproximada de 11 años implica que las hipótesis de estacionariedad e independencia se cumplirán sólo en un modo aproximado, lo cual puede observarse al representar las observaciones en función del tiempo.

La Figura 5.1.1. muestra las curvas correspondientes al máximo anual para algunas estaciones.

La adopción de la distribución Gumbel como modelo de probabilidad está basada en la teoría de extremos, (Capítulo1) la cual indica que las únicas distribuciones límites posibles para el máximo de variables aleatorias independientes son: Gumbel, Fréchet y Weibull. Tanto la distribución Gumbel como la Fréchet se usan como modelos en los estudios de vientos.

La Norma CIRSOC 102: Acción del viento sobre Estructuras ha adoptado la distribución Fréchet, Sin embargo la nueva norma adopta la distribución Gumbel. En el trabajo *Estimación de velocidades de recurrencia de vientos basada en velocidades máximas mensuales* de Viollaz y otros (1995) se encuentra que los datos de nuestro país tienen un buen ajuste al modelo Gumbel.

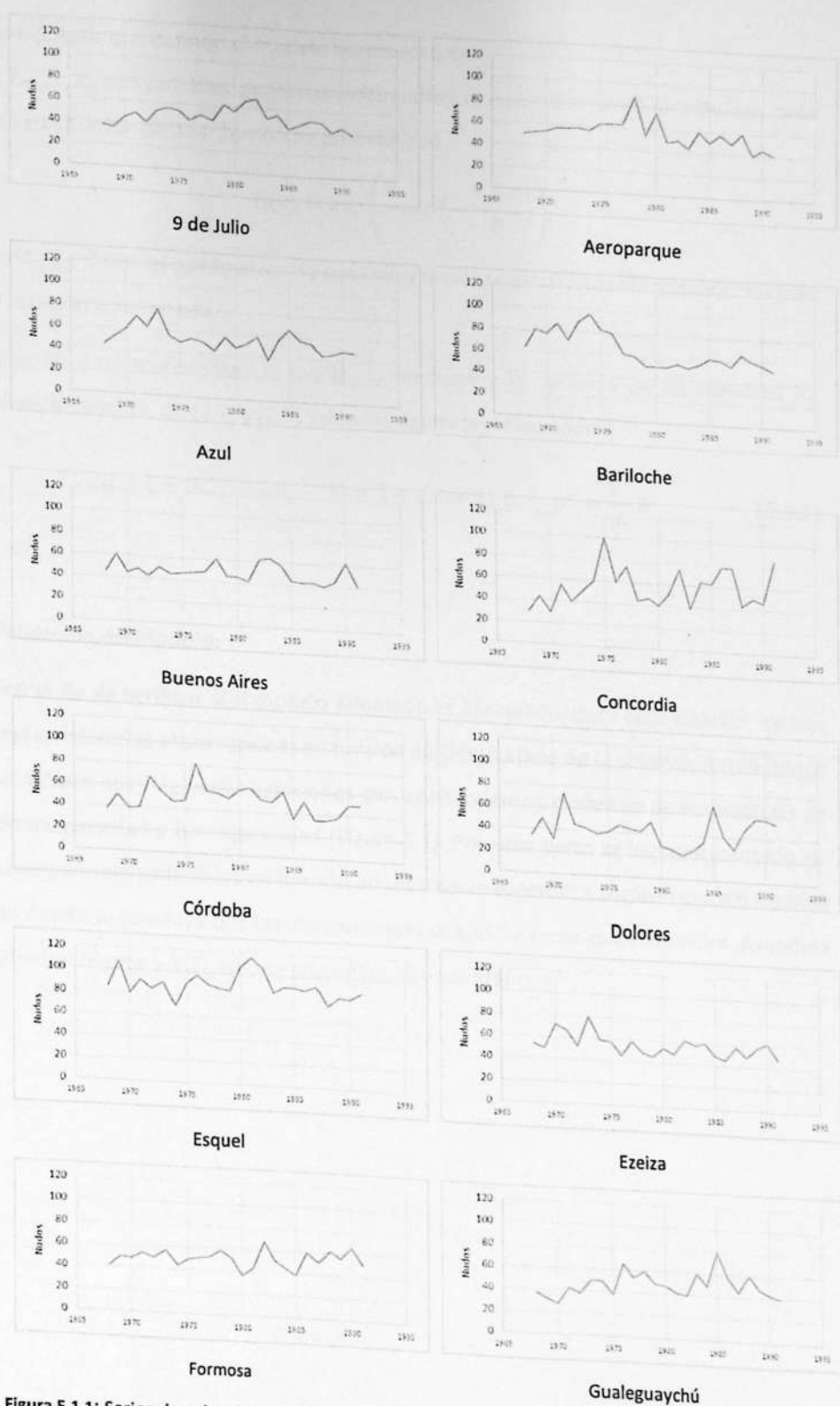


Figura 5.1.1: Series de velocidades máximas anuales del viento

Las hipótesis que definen el modelo se resumen en:

Si X_1, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas, cada una con función de distribución de probabilidad

$$G(x) = \exp\left(-\exp\left(-\frac{x-\xi}{\theta}\right)\right),$$

donde ξ y θ son los parámetros de posición y escala respectivamente que dependen de la estación considerada.

Si μ , M y σ^2 representan la media, la mediana y la varianza de las variables X_i , respectivamente, de (2.2) y (2.3) valen las siguientes relaciones:

$$\mu = \xi + 0.57722 \theta, \quad M = \xi + 0.36651 \theta, \quad \sigma^2 = \frac{\pi^2}{6} \theta^2 \quad (5.1.1)$$

Validación del Modelo:

Con el fin de verificar si el modelo adoptado es apropiado, para cada estación se han representado las observaciones en función del tiempo (año de la observación) de donde se deduce que hay varias estaciones con apartamientos evidentes de las hipótesis de estacionariedad y homogeneidad (Figura 5.1). Por otra parte se han representado las observaciones ordenadas en función de sus valores esperados según el modelo Gumbel de donde se concluye que las observaciones se ajustan razonablemente bien al modelo Gumbel (Figura 5.1.2), sólo se presentan algunos gráficos.

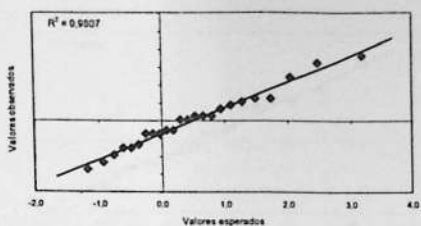


Figura 2.1. Prueba de la distribución Gumbel, estación 9 de Julio

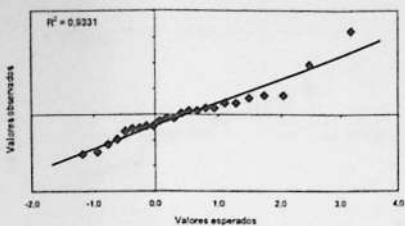


Figura 2.2. Prueba de la distribución Gumbel, estación Aeroparque

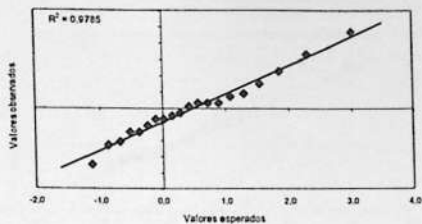


Figura 2.3. Prueba de la distribución Gumbel, estación Azul

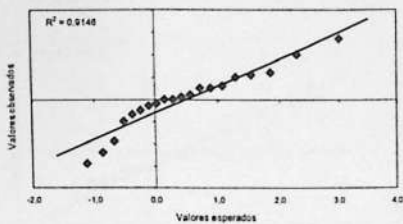


Figura 2.4. Prueba de la distribución Gumbel, estación Bahía Blanca

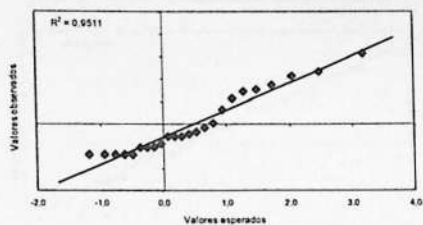


Figura 2.5. Prueba de la distribución Gumbel, estación Bariloche

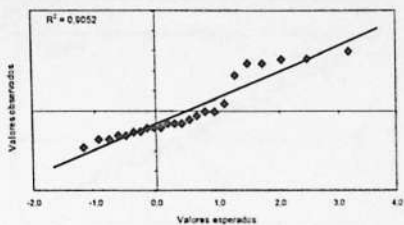


Figura 2.6. Prueba de la distribución Gumbel, estación Buenos Aires

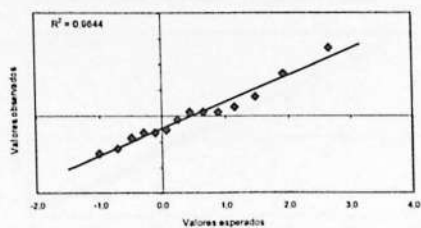


Figura 2.7. Prueba de la distribución Gumbel, estación Catamarca

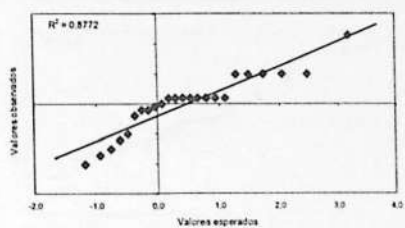


Figura 2.8. Prueba de la distribución Gumbel, estación Ceres

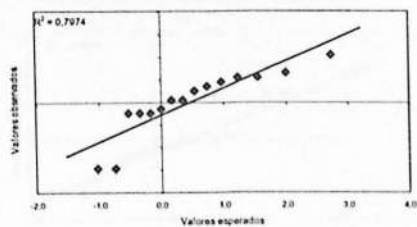


Figura 2.9. Prueba de la distribución Gumbel, estación Cipolletti

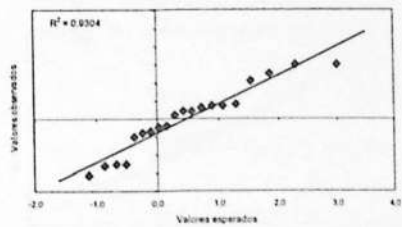


Figura 2.10. Prueba de la distribución Gumbel, estación Comodoro Rivadavia

Figura 5.1.2: Q-QPlot para Chequear el ajuste a la distribución Gumbel.

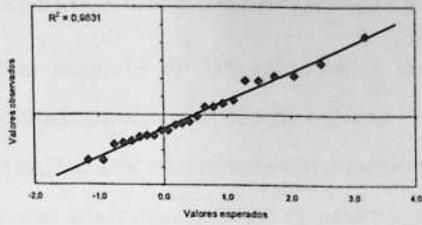


Figura 2.11: Prueba de la distribución Gumbel, estación Concordia

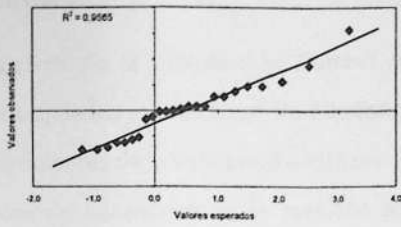


Figura 2.12: Prueba de la distribución Gumbel, estación Córdoba aero

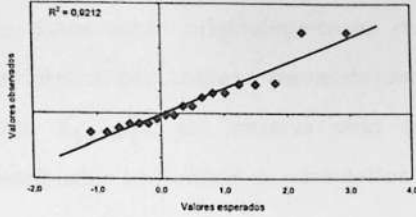


Figura 2.13: Prueba de la distribución Gumbel, estación Córdoba Obs

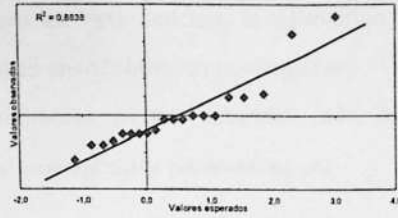


Figura 2.14: Prueba de la distribución Gumbel, estación Coronel Suárez

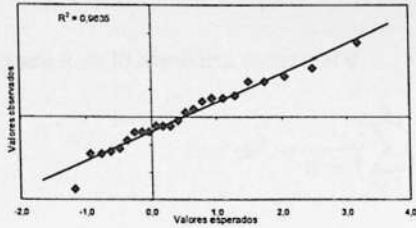


Figura 2.15: Prueba de la distribución Gumbel, estación Corrientes

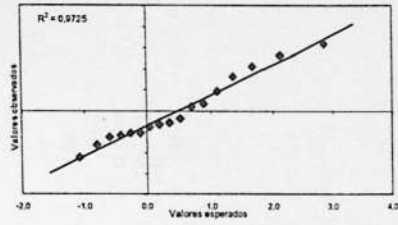


Figura 2.16: Prueba de la distribución Gumbel, estación Cristo Redentor

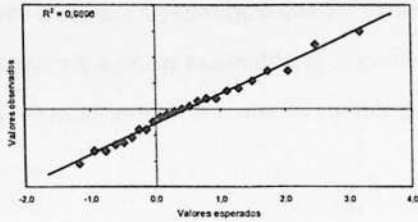


Figura 2.17: Prueba de la distribución Gumbel, estación Dolores

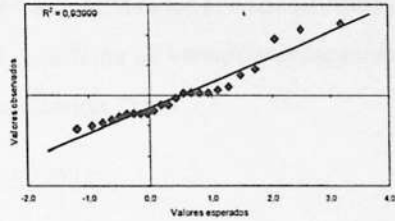


Figura 2.18: Prueba de la distribución Gumbel, estación El Palomar

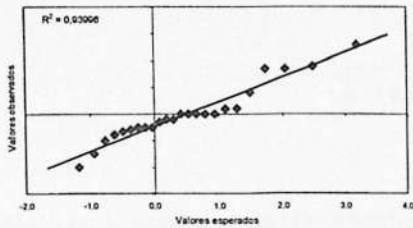


Figura 2.19: Prueba de la distribución Gumbel, estación Esquel

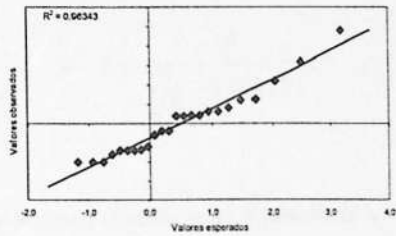


Figura 2.20: Prueba de la distribución Gumbel, estación Ezeiza

Figura 5.1.2 (continuación): Q-QPlot para Chequear el ajuste a la distribución Gumbel.

5.2: Estimación de Parámetros y Percentiles

En la mayoría de las aplicaciones los parámetros de la distribución Gumbel son estimados por el método de máxima verosimilitud, de los momentos o de estadísticos de orden, sobre la base de observaciones independientes de la variable. En esta sección hemos preferido estimar el parámetro de posición basándose en la mediana para obtener un estimador robusto.

Los datos están originalmente en nudos pero en esta sección, la estimación de parámetros percentiles y curvas de nivel se realizó en m/s (metros por segundo).

Sean X_1, \dots, X_n las observaciones correspondientes a una estación dada, cuya distribución es Gumbel de parámetros ξ y θ . Estimamos estos parámetros por

$$\hat{\xi} = \bar{x} - 0.36651\hat{\theta}, \quad \hat{\theta}^2 = \frac{6}{\pi^2} s^2$$

donde \bar{x} es la mediana muestral y

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Los parámetros ξ y θ en general varían de una estación a otra. Sin embargo en estudios sobre vientos se encontró que los parámetros están relacionados y en consecuencia una hipótesis a priori razonable es suponer que el coeficiente de variación de las variables X_i se mantiene más o menos estable para cada estación. O sea

$$\frac{\pi}{\sqrt{6}} \theta = k(\xi + 0.57722 \theta),$$

de donde se obtienen las relaciones

$$\theta = \frac{k}{\frac{\pi}{\sqrt{6}} - k 0.57722} \xi \quad \text{y} \quad \theta = \frac{M}{\frac{\pi}{\sqrt{6}} - k 0.2171} \xi$$

Por lo tanto suponer que el coeficiente de variación es constante es equivalente a suponer que los cocientes $\frac{\theta}{\xi}$ y $\frac{\theta}{M}$ son constantes.

Para de verificar la validez de la relación se reemplazan los parámetros por sus estimadores. Denotando por s_k e \tilde{y}_k el desvío estándar y la mediana respectivamente de la estación k , respectivamente, se tiene la siguiente ecuación

$$s_k = a\tilde{y}_k + e_k \quad k = 1, 2, \dots, 64. \quad (5.2.1)$$

donde e_k es una variable aleatoria que surge al reemplazar los parámetros por sus estimadores, la Figura 3 muestra que la validez de la hipótesis es aceptable.

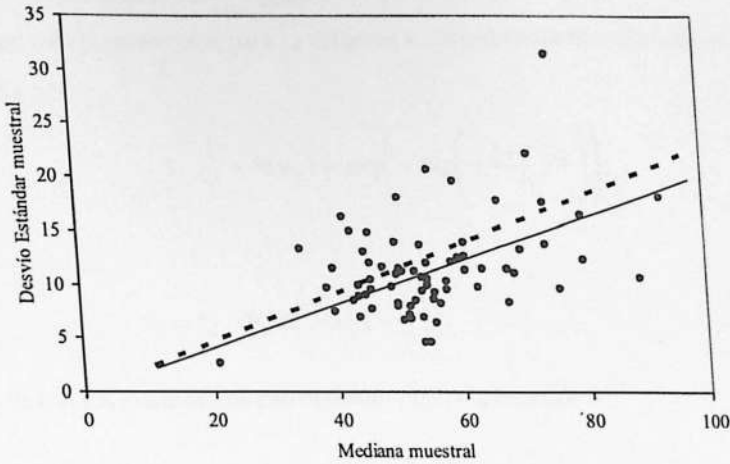


Figura 5.2.3: Desviaciones estándar en función de las medianas. La línea llena es la recta ajustada por mínimos cuadrados. La línea punteada con pendiente $\alpha^* = 0.2278$ es la recta adoptada en este trabajo para estimar el parámetro de escala.

El estimador por mínimos cuadrados de α da: $\hat{\alpha} = 0.2012$, aceptar este valor de $\hat{\alpha}$ podría implicar la subestimación de los parámetros θ_k para ciertas estaciones. Obviamente esto sería riesgoso pues conduciría a una subestimación de los percentiles superiores de la velocidad máxima del viento, por ello adoptamos el valor $\alpha^* = 0.2278$, escogido de manera que el 75% de los puntos caigan debajo de la recta con esta pendiente (Figura 5.2.3).

Estimación de los percentiles:

Los cálculos de riesgos de estructuras civiles frecuentemente están basados en las velocidades de recurrencia de la velocidad máxima anual del viento. Sin embargo a nuestro criterio los cálculos de las probabilidades de riesgo deberían basarse en la distribución de probabilidad de la velocidad máxima del viento en el periodo igual a la vida útil de diseño de la estructura. En esta sección se calculan las velocidades de recurrencia del viento con periodo T igual a 2, 50 y 200 años.

La velocidad de recurrencia v_t para la estación k correspondiente a un periodo de T años, se define por

$$1 - \frac{1}{T} = F(v_T) = \exp\left(-\exp\left(-\frac{v_T - \xi_k}{\theta_k}\right)\right),$$

luego

$$v_T = \xi_k - \theta_k \ln\left(-\ln\left(1 - \frac{1}{T}\right)\right), \quad (5.2.2)$$

donde ξ_k y θ_k son los parámetros correspondientes a la estación k.

De las ecuaciones (5.1.1) y (5.2.2) se obtiene

$$\hat{\theta}_k = \frac{\sqrt{6}}{\pi} \alpha^* \tilde{y}_k = 0.1776 \tilde{y}_k$$

y de (5.1) y (5.3) se obtiene

$$v_T = \bar{x} \left[1 - \frac{\sqrt{6}}{\pi} \alpha^* \left(0.36651 + \ln\left(-\ln\left(1 - \frac{1}{T}\right)\right)\right) \right]$$

Particularizando para periodos T = 2, 50 y 200 años se tiene:

$$v_2 = 1.0000 x_k, \quad v_{50} = 1.6279 x_k, \quad v_{200} = 1.8755 x_k.$$

En la Tabla 1 se dan las estimaciones de los parámetros ξ_k y θ_k , y las velocidades del viento con periodo de recurrencia T igual a 2, 50 y 200 años.

Tabla 5.2.1: Velocidades máximas anuales del viento con periodos de recurrencia de 2, 50 y 200 años.

Estaciones	θ [m/s]	ξ [m/s]	Periodo de recurrencia de 2 años	Periodo de recurrencia de 50 años [m/s]	Periodo de recurrencia de 200 años [m/s]
9 de Julio	3.27	24.52	25.72	41.82	48.16
Aeroparque	4.16	26.77	28.29	46.00	52.98
Azul	3.73	26.15	27.78	45.16	52.02
Bahía Blanca	5.48	35.80	38.07	61.89	71.28
Bariloche	5.53	29.10	31.38	51.02	58.76
Catamarca	4.46	25.37	27.27	44.33	51.05
Ceres	3.96	21.13	22.64	36.80	42.38
Comodoro Rivadavia	7.27	45.18	47.84	77.78	89.58
Concordia	7.20	23.81	26.24	42.65	49.13
Córdoba Obs.	2.62	25.28	26.24	42.65	49.13
Coronel Suarez	3.84	18.64	20.06	32.62	37.57
Cristo Redentor	12.61	34.47	39.10	63.56	73.21
Dolores	5.18	19.96	23.15	37.63	43.35
Esquel	4.27	44.22	45.79	74.43	85.73
Formosa	3.14	25.58	26.75	43.49	50.09
Gobernador Gregores	3.32	33.76	34.98	56.87	65.50
Gualedguaiichú	4.63	23.03	24.69	40.14	46.24
Iguazú	5.27	13.50	18.01	29.27	33.71
Jujuy	3.38	20.88	22.12	35.96	41.42
Junín	2.76	26.77	27.78	45.16	52.02
Laboulaye	3.79	28.41	29.84	48.51	55.87
La Plata Aero	5.44	25.79	27.78	45.16	52.02
La Rioja	3.54	21.34	22.64	36.80	42.38
Lago Argentino	4.19	24.70	28.29	46.00	52.98
Las Lomitas	3.78	22.28	23.66	38.47	44.31
López Juárez	3.15	24.31	25.72	41.82	48.16
Malargüe	7.06	35.48	38.07	61.89	71.28
Mar de Plata Aero	4.83	28.58	30.35	49.34	56.83
Marcos Juárez	5.54	23.69	25.72	41.82	48.16
Mendoza Aero	4.55	23.80	25.72	41.82	48.16
Monte Caseros	4.95	29.05	30.87	50.18	57.79
Neuquén	3.88	30.99	32.41	52.69	60.68
Orán	5.94	19.43	22.12	35.96	41.42
Paraná	5.02	29.55	31.38	51.02	58.76
Paso de Indios	4.42	33.99	35.50	57.71	66.46
Paso de Los Libres	4.26	26.22	27.78	45.16	52.02
Pehuajó	3.39	25.77	27.27	44.33	51.05
Pigüé	2.94	19.50	20.58	33.45	38.53
Posadas	3.65	27.47	28.81	46.83	53.94
Puelches	1.82	27.63	28.29	46.00	52.98
Puerto Deseado	4.56	33.31	34.98	56.87	65.50
Punta Indio	4.74	21.93	23.66	38.47	44.31
Resistencia	3.28	28.34	29.32	47.67	54.91
Río Gallegos	6.59	38.74	41.16	66.91	77.06
Río Grande	4.08	28.34	29.84	48.51	55.87
Rosario Aero	4.50	29.73	31.38	51.02	58.76
Salta	2.76	21.62	22.64	36.80	42.38
San Antonio Oeste	3.94	26.85	28.29	46.00	52.98
San Juan	2.52	27.89	28.81	46.83	53.94
San Julián	4.91	39.36	41.16	66.91	77.06
San Luis	1.80	27.12	27.78	45.16	52.02
Santa Cruz	5.30	34.07	36.01	58.54	67.43
Santa Rosa	8.22	25.79	28.81	46.83	53.94
Santiago del Estero	6.50	18.20	21.61	35.13	40.46
Tandil	7.08	31.87	34.47	56.03	64.54
Trelew	3.81	37.90	39.10	63.56	73.21
Tres Arroyos	5.91	21.24	23.66	38.47	44.31
Tucumán	4.15	22.14	23.66	38.47	44.31
Ushuaia B N	4.58	31.25	32.92	53.53	61.65
Viedma Aero	8.88	33.79	37.04	60.22	69.35
Villa Dolores	2.71	25.50	26.75	43.49	50.09
Villa Reynolds	7.81	28.00	30.87	50.18	57.79

5.3: Trazado de las Isolíneas:

La construcción de los mapas con las isolíneas de las velocidades de recurrencia (isopercentiles) se realizó con el Software *Surfer 6.0* (1995) (*Surface Mapping System*). Este software construye las isolíneas utilizando un método de interpolación lineal basado en una superficie poliédrica de caras triangulares que pasa por los puntos-datos.

Sólo se muestran las figuras para el caso de velocidades del viento con periodo de recurrencia de 50 años. La Figura 5.3.1 muestra el mapa de las isolíneas y los puntos-datos y la Figura 5.3.2 el mapa coloreado según la magnitud de la velocidad del viento.

En la construcción de los mapas se usó sólo información estadística la cual en algunas partes del territorio nacional es más bien escasa. Creemos que se podrían mejorar estos mapas buscando métodos que combinen la información estadística con el conocimiento de los expertos en el tema.

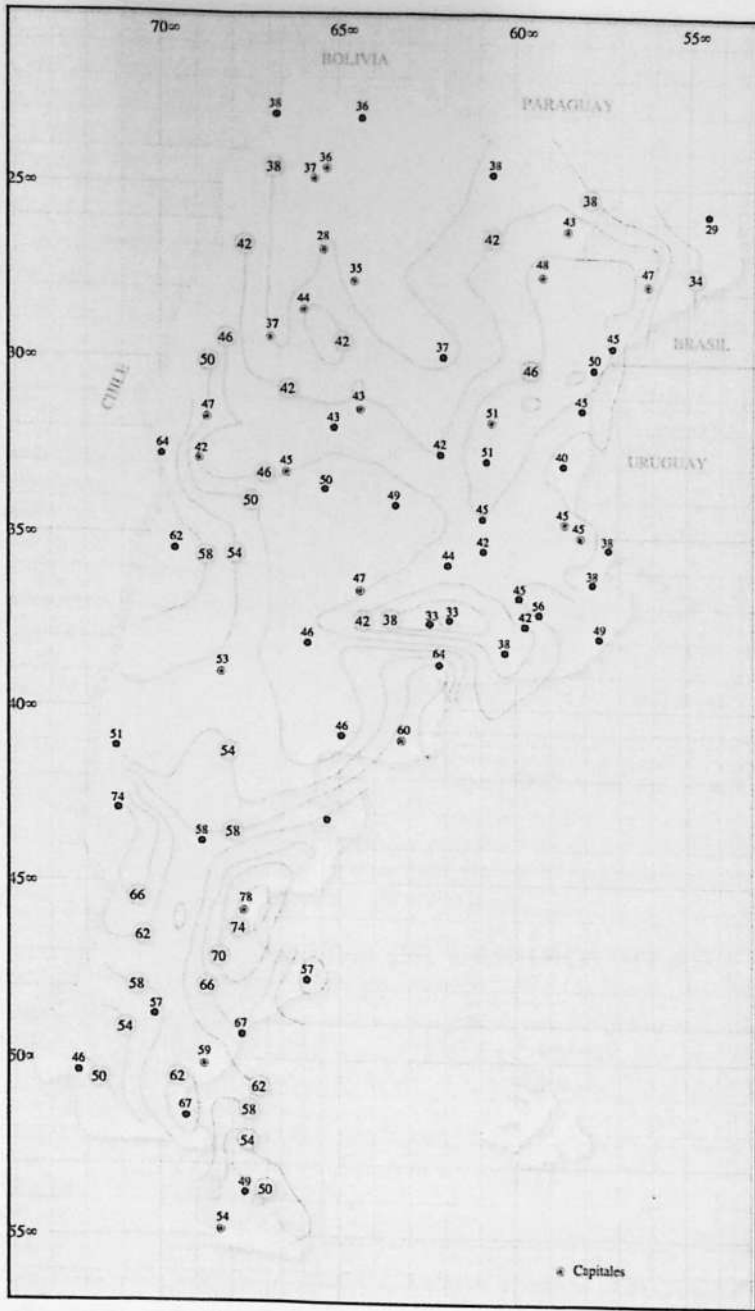


Figura 5.3.1: Mapa de curvas de velocidades del viento con periodos de recurrencia de 50 años (percentil 98).

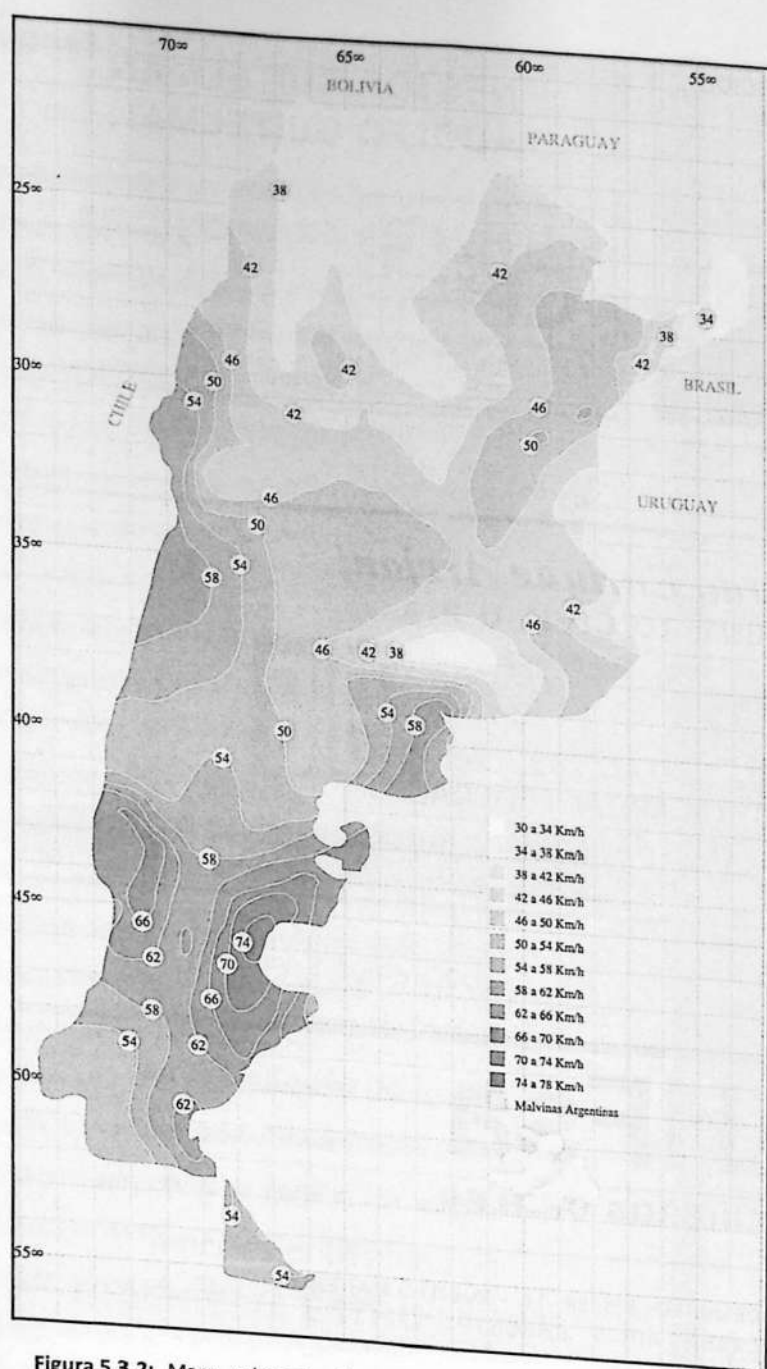


Figura 5.3.2: Mapa coloreado de velocidades del viento con periodos de recurrencia de 50 años (percentil 98).

Conclusiones

En la búsqueda de modelos para la distribución del máximo la principal preocupación está en obtener una buena estimación de la cola derecha de la distribución, es decir, estimar con la mejor precisión posible las probabilidades de ocurrencia de sucesos extremos en cada caso de interés para realizar las acciones preventivas necesarias. Este problema está presente en muchísimos casos reales en Ciencias ambientales, en economía y finanzas, empresas de seguros, etc.

Dada su extensa aplicabilidad existirán casos de variables de interés con diferentes distribuciones de probabilidad, en el presente trabajo se plantea resolver casos dentro del dominio de atracción de la distribución Gumbel que incluye una gran variedad de distribuciones de interés práctico.

La contribución principal de este trabajo radica en la postulación del modelo Potencial-Gumbel, el cual surge de la aplicación de la transformación Box-Cox modificada, que es más amplio que la propia distribución Gumbel y que está contenido en el dominio de atracción de la distribución Gumbel. El modelo paramétrico Potencial-Gumbel recorre el dominio Gumbel en el sentido que según varía el parámetro de forma λ la correspondiente distribución varía el peso de la cola y entonces permite discriminar las distribuciones dentro del dominio Gumbel y con ello obtener mayor precisión que estimar con la propia asíntota.

En muchos de los trabajos más destacados sobre este tema los estudios se vuelven hacia la distribución límite de extremos Fréchet la cual contiene a las distribuciones conocidas como "distribuciones de colas pesadas", la principal razón de esto es que estas distribuciones son de algún modo más seguras en el sentido que tienden a sobreestimar los percentiles superiores. Pero también hay que tener en cuenta que esta ganancia en seguridad representa una pérdida de precisión en las estimaciones, y en algunos casos aplicados podría traducirse en pérdidas económicas importantes.

En resumen se puede concluir que:

- 1- Aunque no se haya probado explícitamente, se puede inducir informalmente que la distribución Potencial-Gumbel tiene cierta densidad dentro del dominio

Gumbel en el siguiente sentido: Conciba el dominio de atracción de la distribución Gumbel ordenado según la cola derecha de las distribuciones, estas van desde moderadamente pesadas hasta Muy livianas. Ahora suponga que se hace variar el parámetro de forma λ de manera continua desde 0 hasta infinito, entonces la distribución Potencial-Gumbel también tiene distribuciones de cola derecha que van desde moderadamente pesadas hasta Muy livianas. La cola derecha puede hacerse tan liviana como se quiera haciendo tender a infinito el parámetro de forma. Por otra parte si se toma $\lambda \rightarrow 0$ se tiene el caso de colas más pesadas del dominio Gumbel que correspondería a distribuciones muy cercanas a la distribución Fréchet.

- 2- El modelo Potencial-Gumbel no presenta problemas en la estimación de sus parámetros dado que se pueden utilizar diferentes métodos de estimación convencionales como ser Máxima Verosimilitud, estadísticos de orden y método de los momentos ponderados por probabilidad. Similar al caso del modelo Gumbel no existen expresiones algebraicas explícitas para los estimadores de los parámetros pero se pueden resolver numéricamente sin complicaciones mayores.
- 3- El método utilizado en este trabajo permite visualizar la bondad del ajuste mediante el Q-QPlot, en el cual se toma el valor del parámetro de forma óptimo. En estos gráficos se puede además, visualizar claramente la influencia de los valores extremos de la muestra, principalmente del máximo muestral, y en algunos casos el valor máximo de la muestra puede ser subestimado (por alguna razón externa) y entonces en el Q-QPlot se verá claramente como este dato no está en concordancia con el resto de la muestra. En estos casos no queda más remedio que buscar una estimación más segura en el sentido de preferir equivocarse por arriba en lugar de subestimar la probabilidad de ocurrencia de valores grandes.
- 4- En el ajuste del modelo Potencial-Gumbel a los datos de series de vientos máximos mensuales de Argentina puede verse que en cada estación meteorológica (ubicadas geográficamente en diferentes zonas del País) hay un comportamiento diferente en cuanto al peso en las colas derechas de las distribuciones que es captado por el modelo. Esto permite estimar con mayor

precisión los percentiles superiores de cada estación, y en caso que se quiera uniformizar el modelo asignando un único parámetro de forma a todas las estaciones, se puede adoptar un criterio conservador eligiendo de entre los valores estimados, un valor mínimo para λ , que en este caso sería de 0,45. En la tabla 4.4.7 se puede evaluar la pérdida de precisión que se da por obtener estimaciones más conservadoras de los percentiles. En la columna 3 se presentan las estimaciones de los percentiles tomando un caso ya muy extremo de estimaciones conservadoras. Este último caso sería equivalente a utilizar como modelo de ajuste una distribución muy próxima a la distribución de Fréchet.

5- En definitiva, el modelo Potencial-Gumbel propuesto aquí tiene la capacidad de discriminar distribuciones de diferentes colas dentro del dominio de atracción de la distribución Gumbel, lo que permitirá estimaciones más precisas de los percentiles superiores de la distribución, y dada la riqueza en distribuciones del dominio Gumbel el modelo propuesto aquí deberá servir para resolver un sinfín de casos aplicados.

Referencias

- [1] Mentz R. , Viollaz A. y Martinez C. (2003) Confidence Bands for Time Series Problems in the Time and Frecuency Domains. Brazilian Journal of Statistics, Vol.17, p179-211.
- [2] Riera J. D. and. Reimundín J. C (1970) Sobre la distribución de velocidades máximas de viento en la República Argentina, Simposio sobre Acciones en Estructuras, UNT.
- [3] Viollaz A. y Lazarte V.(2011) Parametrization of the domain of maximal attraction of the Gumbel distribution. Publicado en actas del III Congreso de Matemática Aplicada, Computacional e Industrial. Bahía Blanca.
- [4] Viollaz A. y Lazarte V.(2005) Transformación de tipo Box-Cox en Estadística de Extremos. Actas del XXXII Coloquio Argentino de Estadística, Córdoba.
- [5] Viollaz A. J. y Lazarte V. F.(2002) Mapa de Velocidades de Vientos de La Argentina.. Revista Ingeniería Estructural, No. 2, Vol. 25, p.34-36.
- [6] Viollaz A. J. y Lazarte V. F. (1999) Distribución del Máximo de una Serie Cronológica con Residuos Gumbel. IV Congreso Latinoamericano de Sociedades de Estadística, Mendoza.
- [7] Viollaz A. J. y otros (1995), Estimación de velocidades de recurrencia de vientos basada en velocidades máximas mensuales, Instituto Nacional de Tecnología Industrial, CIRSOC.
- [8] Viollaz, A, Lopez, E. y Lazarte, F.(1995).Un Modelo Lineal con Residuos Gumbel con Aplicaciones a Extremos. Actas XXIII Coloquio Argentino de Estadística. Córdoba,
- [9] Viollaz, A. (1993) Estimación del Máximo de una Serie de Observaciones con Función de Medias Periódica. Segundo Congreso Latinoamericano de Sociedades Estadísticas.
- [10] Viollaz, A. (1992). Estimación del Máximo de un Proceso Estocástico no Estacionario. Primer Congreso Iberoamericano de Estadística E Investigación Operativa, España.
- [11] Viollaz, A. y Altinger, M. (1990) Datos de Tornados y su Aplicación al Cálculo de Riesgos en la Argentina. Seminario Sobre Aspectos de Seguridad en Relación con el Emplazamiento de Centrales Nucleares, International Atomic Energy Agency, Viena, Austria.

- [12] Viollaz, A. (1982) Distribución de Probabilidad de la Velocidad Máxima Anual del Viento en la Rca. Argentina. Actas del Coloquio Argentino de Estadística.
- [13] Viollaz, A. y otros. (1977) Some Recent Results on Probabbilistics Models for Extreme Wind Speeds, Journal of Industrial Aerodynamics, Vol. 2, No. . 3, p.28-47..
- [14] Viollaz, A. J.(1975) Estudio de la Distribución de la Velocidad Máxima del Viento en la Rca, Argentina. Informe I-75-1, Laboratorio De Estructuras, Universidad Nacional de Tucumán.
- [15] Viollaz, A. y Altinger M. Evaluación del Riesgos de Vientos Máximos y Tornados sobre Líneas de Alta Tensión de Yaciretá. Informe Técnico por Contrato.

Bibliografía

- [1] David, H. A. (1981). *Order Statistics*, 2nd. Ed. J. Wiley and Sons.
- [2] Embrechts, P., Kluppelberg, C. y Mikosch, T. (1997). *Modeling Extreme Events*. Springer Verlag.
- [3] Fearn D.H. and Nebenzahl E. (1995). Using power transformations when approximating quantiles *Comun. Statist. Theory Meth.* 24(4).
- [4] Fisher, R. A. and Tippett, L. H. C. (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Ph, Soc*24, 180-190
- [5] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*. Roberts. Krieger publishing Company Malabar, Florida.
- [6] Kakade C. S., Shirke D. T. and Kundu D. (2007) Inference for $P(Y < X)$ in Exponentiated Gumbel Distribution. *Journal of Statistics and Applications*. In press.
- [7] Kendall and Stuart. Edition (2005). *Advanced Theory of Statistics*. Volume I. y Volume 2A. Oxford University Press Inc., New York.
- [8] Leadbetter, M., Lindgren, G. and Rootzen (1983), H. *Extremes and Related Properties of Random Sequences and Processes*.
- [9] Nadarajah S. (2006). The exponentiated Gumbel distribution with climate application. *Environmetrics*; 17: 13-23.
- [10] Parzen E. (2004). Quantile Probability and Statistical Data Modeling. *Statistical Science*, Vol. 19 No. 4, 652-662.
- [11] Persson K. and Rydén J. (2010) Exponentiated Gumbel Distribution for Estimation of Return Levels of Significant Wave Height. *Journal of Environmental Statistics*. 2010, Volume 1, Issue 3
- [12] Pearson K. and Rydén L. (2007). Exponentiated Gumbel distribution for estimation of return levels of significant wave height. U.U.D.M. Report 2007:19 ISSN 1101-3591.
- [13] Resnik, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag New York Inc.
- [14] Teugels J. L. and Vanroelen G. (2004). Box-Cox transformations and heavy-tailed distributions. *Journal of Applied Probability*. 41(2004), 213-227.
- [15] Vanroelen G. (2003) The effect of transformations on second-order regular variation. Doctoral Thesis, Katholieke Univ. Leuven.

- [16] Von Mises (1936). La distribución de la plus grande de n valeurs. Selected papers II.
Am. Math. Soc. 271-294.