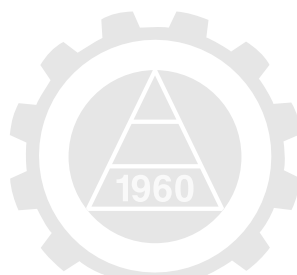


**SOBRE UNA METODOLOGÍA PARA ESTIMAR
ALTOS CUANTILES: UNA APLICACIÓN A LA
DESERCIÓN**



JUAN BAUTISTA ROJAS

**UNIVERSIDAD
EAFIT®**
Abierta al mundo

**ESCUELA DE CIENCIAS Y HUMANIDADES
DEPARTAMENTO DE CIENCIAS BÁSICAS
MAESTRÍA EN MATEMÁTICAS APLICADAS
MEDELLÍN
2012**

SOBRE UNA METODOLOGÍA PARA ESTIMAR ALTOS CUANTILES: UNA APLICACIÓN A LA DESERCIÓN

JUAN BAUTISTA ROJAS

Trabajo presentado como requisito parcial para optar al Título de
Magíster en Matemática Aplicada

**UNIVERSIDAD
EAFIT**

Director:

JOSÉ ANTONIO SOLANO ATEHORTÚA

Profesor Universidad de Antioquia

**ESCUELA DE CIENCIAS Y HUMANIDADES
DEPARTAMENTO DE CIENCIAS BÁSICAS
MAESTRÍA EN MATEMÁTICAS APLICADAS**

MEDELLÍN

2012

UNIVERSIDAD EAFIT
ESCUELA DE CIENCIAS BÁSICAS Y HUMANAS

Los jurados abajo firmantes certifican que han leído y que aprueban la tesis titulada **“SOBRE UNA METODOLOGÍA PARA ESTIMAR ALTOS CUANTILES: UNA APLICACIÓN A LA DESERCIÓN”**, presentada por **JUAN BAUTISTA ROJAS**.

Fecha: Febrero de 2012

Asesor: _____

José Antonio Solano Atehortúa

Jurado: _____

Jurado: _____

AGRADECIMIENTOS

Porque lograr un objetivo, fuera del esfuerzo individual, se requiere de un buen entorno, a mi familia: María del Carmen Rojas (Madre), Juan Rojas (Padre), a mis hermanos y a la indiscutible unión de mis tías, a doña Julia que me ha apoyado desde su lugar más anhelado, a Paulo Mora por la paciencia que me tuvo en cuanto a las correcciones, sugerencias y edición de este texto, a mi asesor José Solano ya que me dejó participar en este proceso, mis compañeros de trabajo y amigos. A todos ellos gracias por permitir que el camino fuera más llevadero.

RESUMEN

Uno de los mayores intereses en la teoría estadística de valores extremos es inferir las propiedades de la cola de las funciones de distribución en el dominio de atracción de una distribución de valores extremos y predecir eventos raros. La hidrología, los seguros y las finanzas son algunos campos de la ciencia donde esta teoría se aplica con gran frecuencia. En este trabajo se exhibe una metodología para estimar altos cuantiles de distribuciones cuyo dominio de atracción pertenece a una de las distribuciones de valores extremos. En el proceso de estimación del cuantil se emplea un método de optimización no lineal para hallar el máximo de una función.

Un ejemplo numérico es dado para validar la teoría y una aplicación a la deserción estudiantil es llevada a cabo usando una base de datos de estudiantes recopilada durante ocho semestres.

Palabras clave. Distribución de valores extremos, distribuciones generalizadas de Pareto, método de picos sobre el umbral.

ÍNDICE GENERAL

CAPÍTULO 1: INTRODUCCIÓN	19
1.1 Los Riesgos empresariales	22
1.2 Problema de estimación del riesgo	24
1.3 Distribuciones de Pareto Generalizadas GPD	28
CAPÍTULO 2: RESULTADOS TEÓRICOS	33
2.1 Teoría Básica de Valor Extremo	33
2.1.1 Formulación de la relación límite para el máximo y el mínimo	36
2.1.2 Formulaciones alternativas de la relación límite	39
2.1.3 Distribución Generalizada de Valores Extremos	53
2.1.4 Dominios de Atracción y Convergencia Débil	56
2.2 Métodos de umbral y Distribución Generalizada de Pareto	68
CAPÍTULO 3: ESTIMACIÓN DEL ÍNDICE DE VALOR EXTREMO Y PRUEBAS	81
3.1 Estimador de Hill $\zeta > 0$	84
3.2 Función de distribución de excesos de la muestra	98
3.3 Resumen	101
3.4 Procedimiento de Inferencia Estadística	101
3.5 Proceso general de la aplicación	102

3.6	Cálculo de u^{opt}	104
3.7	Estimando ζ y β	105
3.8	Cálculo de $F(u^{opt})$	106
CAPÍTULO 4: CASO DE APLICACIÓN		107
4.1	Prueba 1	108
4.2	Generando la muestra	109
4.3	Aplicación del método POT	112
4.4	Una aplicación a la deserción	115
CONCLUSIONES		121
APÉNDICE		123
A.1	123
A.2	123
A.3	124
BIBLIOGRAFÍA		125

ÍNDICE DE FIGURAS

CAPÍTULO 1: INTRODUCCIÓN	19
1.1 cola de la distribución $F(\cdot)$ a la derecha de la abscisa u , que es la parte del gráfico de $F(\cdot)$ a la derecha del punto $(u, F(u))$	26
1.2 Gráfica de (1.1): (ramas de hipérbolas horizontales)	27
1.3 Gráfica de (1.2) (la función de distribución exponencial)	27
1.4 Gráfica de (1.3): (ramas de parábolas)	27
1.5 Distribución de Pareto para el parámetro de escala $\beta = 1$, parámetro de localización $S = 0$ y varios valores positivos para el índice de cola ζ	30
1.6 Distribución de Pareto para el parámetro de escala $\beta = 1$, parámetro de localización $S = 0$ e índice de cola $\zeta = 0$	30
1.7 Distribución de Pareto para el parámetro de escala $\beta = 1$, parámetro de localización $S = 0$ y varios valores negativos para el índice de cola ζ	31
CAPÍTULO 2: RESULTADOS TEÓRICOS	33
2.1 $\max\{X_1\}$	38
2.2 $\max\{X_1, \dots, X_{70}\}$	39
2.3 $\max\{X_1, \dots, X_{700}\}$	39

2.4	Cuantiles de la distribución mínima de GEV, en forma canónica ($\tilde{a}' = 0$ y $\tilde{b}' = 1$). El parámetro toma los valores $\tilde{\gamma}' = -2, -1, 0, 1/2, 1, 2$ donde $\tilde{\gamma}'$ es como α dado en la definición 6	67
2.5	Relación entre una cola gruesa y una cola fina	79

CAPÍTULO 3: ESTIMACIÓN DEL ÍNDICE DE VALOR EXTREMO Y PRUEBAS **81**

3.1	Gráfica de la función $G_{\xi, \beta(u), 0}^T(\cdot)$ resultante de la transformación [(1), (2), (3), (4)] aplicada a $G_{\xi, \beta(u), 0}(\cdot)$, que por su parte, es la función proporcionada por el teorema 2.15 y que aproxima a $F_u(\cdot)$. La condición suficiente para que este procedimiento funcione es que la $F(\cdot)$ esté en el dominio de atracción de una de las D.V.E.	100
3.2	Datos $X_1, X_2, X_3, X_4, X_5, X_6, X_7$ y los correspondientes excesos $y_1, \dots, y_{N_{u^{opt}}}$ sobre el umbral u . En este caso $N_{u^{opt}} = 3$	102

CAPÍTULO 4: APLICACIÓN **107**

4.1	Gráfica de (4.1)	108
4.2	Transformación inversa	109
4.3	Datos obtenidos de la muestra, en el eje de las abscisas se representa el tiempo y el eje de las ordenadas la magnitud de cada dato. Inicialmente se pueden detectar puntos extremos como el punto 11, punto 71 y el punto 381.	110
4.4	Función de distribución muestral o empírica construida a partir de los datos suministrados X_1, \dots, X_{500}	110
4.5	Gráfico de probabilidad–probabilidad de los datos de $F^{prueba}(x)$	111
4.6	Gráfico cuantil–cuantil de los datos de $F^{prueba}(x)$	112
4.7	Gráfica de la función de distribución muestral o empírica	113
4.8	Ajuste de normalidad para Z^{Sc}	116
4.9	Gráfico de dispersión de Z^{Sc}	118

4.10	Gráfico de probabilidad–probabilidad de los datos de la Z^{Sc}	118
4.11	Gráfico cuantil–cuantil de los datos de la Z	119
4.12	Gráfica de la media de excesos de la muestra para 147 datos provenientes de los alumnos que desertaron el primer semestre del programa de Matemáticas de 2007-1 hasta 2010-2.	119
4.13	Polígono de frecuencia acumulada para la variable Z^{Sc}	120

CAPÍTULO 1

INTRODUCCIÓN

En varias áreas del conocimiento se hace necesario la previsión de eventos extremos y en la última década ésta metodología ha adquirido relevancia debido en gran parte a las exigencias impuestas por organismos de regulación en el ámbito financiero. La mayoría de las veces tales eventos se pueden caracterizar por valores muy pequeños o muy grandes que toma una variable aleatoria cuantitativa, los cuales son llamados valores extremos. El estudio respecto de tales valores pertenece a una rama de la estadística matemática llamada teoría de valores extremos (EVT del inglés *Extreme Value Theory*). Clásicamente el estudio de los valores extremos se ha dado por medio del estudio de la función de distribución $F(x) = P(X \leq x)$ y de la **función de sobrevivencia** $\bar{F}(x) := 1 - F(x)$ o del cuantil p

$$x_p = \inf \{x : F(x) \geq p\},$$

de alguna variable X de interés como puede ser: la temperatura diaria en una zona, el nivel del agua en las orillas, la pérdida diaria de un agente económico, entre otras. El interés de este trabajo es realizar un estudio del comportamiento del cuantil p , cuando p está muy próximo del número 1, esto debido a que el interés se inclina sobre el estudio de los altos cuantiles.

ALGUNOS USOS DE LA TEORÍA DE VALORES EXTREMOS

En esta se presentan ilustraciones y casos de aplicación de la teoría de los valores extremos en diversas áreas del conocimiento: seguros, hidrología y finanzas. Mas ejemplos pueden ser revisados en Coles S. (2004).

Según Beirlant *et al.* (2004), en seguros de vida, de incendios y de automóviles ocurren muchos siniestros. Los incendios industriales específicamente, causan varios efectos colaterales como la pérdida de la propiedad, desempleos temporales y pérdidas de contratos. Estos pueden venir en ocasiones acompañados de otros grandes siniestros, lo que sitúa en riesgo la solvencia de una parte sus-

tancial de la compañía. Por esto, la necesidad por parte de las compañías, de la previsión, de altos cuantiles en la distribución del valor de los siniestros, para hacer frente a las grandes pérdidas, motivando el desarrollo de técnicas apropiadas para la estimación de los mismos. Las empresas aseguradoras queriendo disminuir sus responsabilidades con la aceptación de un riesgo considerado como excesivo o peligroso, transfieren a las reaseguradoras una parte de la responsabilidad y del premio recibido. En un contrato de reaseguro de exceso de daños, la reaseguradora paga por el monto excedente, a un determinado límite de los siniestros. La distribución de la cola superior del valor de los siniestros es un factor clave en la determinación del premio de un contrato de reaseguro.

Precipitaciones, evaporización e inundaciones son cantidades comunes de estudio en la hidrología, las cuales cuando se estudian a través del tiempo, presentan variaciones periódicas generalmente anuales. Con el objeto de eliminar la dependencia de los datos y estudiar los valores extremos, los profesionales de esta área, frecuentemente usan los máximos anuales obteniendo así una serie de datos que representan estos máximos, por ejemplo, el nivel máximo de agua registrado por año durante 20 años, proporciona una serie con 20 datos. Desde este momento en adelante, se calcula el tiempo de ocurrencia de una medida hidrológica determinada a partir de la estimación de lo que se conoce como período de retorno de T años. El interés puede ser inclusive estudiar la magnitud de la cantidad para un período de T años dado, la cual es conocida como nivel del retorno de T años. Si se piensa en el máximo anual de una medida hidrológica representada por una variable X , tenemos que el período de retorno corresponde a,

$$T(x) = \frac{1}{P(X > x)}$$

o sea, el inverso de la probabilidad de sobrevivencia de una cantidad determinada X aplicada en una magnitud x . El nivel de retorno de T años, x_T , corresponde al $(1 - \frac{1}{T})$ -ésimo cuantil de X , con

$$x_T = \inf \left\{ x : F(x) \geq 1 - \frac{1}{T} \right\}$$

En términos generales se dice que x_T corresponde al nivel que será excedido en media, cada T años, o incluso, el nivel que será excedido con probabilidad $\frac{1}{T}$ en un año cualquiera. Es posible hallar una relación directa entre el período de retorno y el nivel de retorno. La mayor parte del tiempo se desea calcular el nivel de retorno de $T = 100$ años, x_{100} o sea, el 0.99-cuantil de los máximos, pues

$1 - \frac{1}{100} = 0.99$, aun que las estimaciones sean basadas en períodos de tiempo cortos, como ocurre en el conjunto de datos aplicados en Van Noortwijk J. M. (1999). Existen algunos trabajos que se centran en el uso de intervalos de confianza para niveles de retorno como por ejemplo Rust *et al* (2006), donde los métodos *bootstrap* son usados para calcular intervalos de confianza para niveles de retorno de inundaciones.

En la última área en la que vamos a proporcionar un ejemplo de aplicación de la teoría de valores extremos es en finanzas. En este caso se tienen series de precios de activos financieros. Uno de los objetivos de las finanzas es la estimación de los riesgos de un portafolio de activos financieros, los cuales en general se miden por medio de las variaciones de los precios de tales activos. Suponiendo que el intervalo de tiempo entre las observaciones no varía (las observaciones son tomadas por ejemplo semanalmente), podemos denotar el precio de un activo determinado en el instante t por P_t . Los principales tipos de variación de los precios son denominados retorno líquido simple, retorno bruto simple y log-retorno. El retorno líquido simple en el instante t se define como la variación relativa entre los precios consecutivos

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$$

y el retorno bruto simple corresponde a la cantidad $1 + R_t$. El log-retorno, que es el más usado, se define por

$$r_t = \log(1 + R_t)$$

Como se expreso anteriormente, el ánimo en este campo es estimar el riesgo financiero de mercado, ya que este riesgo está directamente relacionado con las pérdidas potenciales asociadas al comportamiento del mercado. Una de las medidas que más se usa para estimar este riesgo es conocida como el VaR (del inglés *Value at Risk*). Se puede pensar el VaR como el p -ésimo cuantil del retorno (un cuantil alto de esta distribución es de mucho interés para el mercado).

De acuerdo con Tsay R. S. (2002), desde el punto de vista de una entidad financiera el VaR puede ser definido como la pérdida máxima de una posición financiera durante un período de tiempo dado con un nivel de probabilidad también dado, tratándose como una medida de pérdida asociada a un evento raro sobre condiciones de mercado normales. Detalles adicionales sobre el valor en riesgo y las formas de calcularlo se pueden encontrar en Tsay R. S. (2002). Sin embargo, cuando los rendimientos son excesivos, un modelo más adecuado de los escena-

rios de los posibles rendimientos requiere del uso de las distribuciones de valores extremos.

Ahora, continuando con las ideas plantadas anteriormente, nos disponemos a exhibir una aplicación más de este modelo como lo es el de la deserción académica, en el cual estamos interesados en predecir si un estudiante está en alto riesgo de desertar a sus estudios, para poder intervenir rápidamente este hecho antes de que ocurra.

1.1 LOS RIESGOS EMPRESARIALES

La palabra riesgo proviene del italiano *risico* o *rischio*, y este del árabe clásico *rizq*, lo que depara la providencia. El riesgo es la probabilidad de un evento adverso y sus consecuencias.

En el documento técnico *RiskMetrics*TM, elaborado por el banco J.P. Morgan (1996), riesgo es definido como el grado de incertidumbre con relación a retornos líquidos futuros y puede asumir diversas formas basadas en la fuente de incertidumbre:

- *Riesgo de mercado*: es la pérdida potencial por cambios en las variables financieras relevantes del mercado que inciden sobre la valuación o sobre los resultados esperados de las operaciones activas, pasivas o causantes de pasivo contingente, tales como tasas de interés, tipos de cambio, índices de precios, entre otros.
- *Riesgo de crédito*: es la pérdida potencial causada por el incumplimiento de las obligaciones contractuales por parte de la contraparte. Habitualmente se relaciona el riesgo de crédito con instituciones financieras, empresas y organismos de otros sectores.
- *Riesgo de operacional*: es el de pérdida que puede ocurrir por una falta de adecuación o de un fallo de los procesos, el personal o los sistemas internos, o bien como consecuencias de eventos externos. Esta definición incluye el riesgo legal, pero excluye el riesgo estratégico y el riesgo reputacional.
- *Riesgo de liquidez*: es la posibilidad de que una entidad incurra en la incapacidad de cumplir cabalmente y de manera oportuna las obligaciones de pago, este riesgo es reflejado por la inhabilidad de la firma de financiar los activos sin liquidez.

Para Jorion P. (1997), las corporaciones están expuestas a tres tipos de riesgo:

- *Riesgos de negocio*: es la posibilidad de que varíen los beneficios de una compañía antes de atender a los pagos de la deuda, la compañía asume estos riesgos con el propósito de crear una ventaja competitiva y agregar valor a los accionistas.
- *Riesgos estratégicos*: son aquellos que destruyen el valor de la empresa por efecto de cambios que no fueron previsto en el ambiente político y económico.
- *Riesgos financieros*: El riesgo financiero se refiere a la probabilidad de ocurrencia de un evento que tenga consecuencias financieras negativas para la organización

El mismo Jorion P. (1997) clasifica los riesgos financieros en cinco tipos: riesgo de mercado, riesgo de crédito, riesgo de liquidez, riesgo operacional y riesgo legal.

- *Riesgo legal*: surgen debido a que la contraparte no tenga autoridad legal o regulatoria para tomar parte en una transacción. Este riesgo está enmarcado en el riesgo operativo en el sentido en que se debe a la no exigencia de acuerdos contractuales a la contraparte.

Saunders A. (1996) clasifica los riesgos de intermediación financiera en nueve categorías: riesgo de tasa de interés (riesgo asumido por una entidad cuando sus activos y pasivos están descalzados), riesgo de mercado, riesgo de crédito, riesgo de las actividades fuera del balance (riesgo debido a actividades relacionadas a activos y pasivos de contingencia), riesgo tecnológico y operacional, riesgo de cambio (oscilaciones en la tasa de cambio que afectan los activos y pasivos localizados en exterior), riesgo soberano (posibilidad de que los deudores en el extranjero interrumpen sus pagos por intervención del gobierno), riesgo de liquidez y riesgo de insolvencia (riesgo de que la entidad no posea capital suficiente para pagar sus deudas en los plazos estipulados de vencimiento).

Consultando la bibliografía sobre este tema, se nota que algunos autores forjan conceptos de riesgo como algo asociado a la posibilidad o probabilidad de pérdida, mientras que otros utilizan el término para indicar un grado de incertidumbre tanto en relación a las pérdidas como en relación a las ganancias. En este trabajo, el término riesgo será utilizado en su forma más amplia, es decir, como la oscilación alrededor de un valor esperado.

En cuanto al término riesgo operacional, el concepto que se usará será el adoptado por el Comité de Supervisión Bancaria de Basilea, es decir, riesgo operacional es el riesgo de pérdida debido a la inadecuación o a fallos de los procesos, el personal y los sistemas internos o bien a causa de acontecimientos externos.

¿Cómo estimar el capital regulatorio por riesgo operacional? ¿Cómo validar modelos de riesgo operacional? ¿Cómo impacta el riesgo de operacional a las entidades financieras durante épocas de crisis? ¿Cómo ajustar distribuciones de severidad y frecuencia? ¿Qué utilizar ¿cópulas o correlaciones? ¿Cuáles son los beneficios de usar la Teoría del valor extremo? ¿Cómo aprovechar los datos externos para estimar el riesgo operacional?

Volvemos ahora a el problema central que estamos tratando, a saber, estimar altos cuantiles de distribuciones de colas pesadas, como se observó anteriormente esto proporciona medidas que pueden usarse como medidas de riesgo en varios campos de la ciencias.

1.2 PROBLEMA DE ESTIMACIÓN DEL RIESGO

Definición 1. (Variable aleatoria v.a.) Dado un espacio de probabilidad (Ω, \mathcal{A}, P) y un espacio medible (S, Σ) , una aplicación $X : \Omega \rightarrow S$ es una variable aleatoria si es una aplicación \mathcal{A}, Σ -medible.

En la mayoría de los casos se toma como espacio medible de llegada el formado por los números reales junto con la σ -álgebra de Borel (el generado por la topología usual de \mathbb{R}), quedando pues la definición de esta manera:

Dado un espacio de probabilidad (Ω, \mathcal{A}, P) una variable aleatoria real es cualquier función $\mathcal{A}/\mathcal{B}(\mathbb{R})$ -medible donde $\mathcal{B}(\mathbb{R})$ es la σ -álgebra boreliana.

Definición 2 (Función de distribución). La distribución de probabilidad de una v.a. X , también llamada función de distribución de X es la función $F_X(x)$, que asigna a cada evento definido sobre X una probabilidad dada por la siguiente expresión: $F_X(x) = P(X \leq x)$ y de manera que se cumplan las siguientes tres condiciones:

1. $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$
2. Es continua por la derecha.
3. Es monótona no decreciente.

La distribución de probabilidad de una v.a. describe teóricamente la forma en que varían los resultados de un experimento aleatorio. Intuitivamente se trataría de una lista de los resultados posibles de un experimento con las probabilidades que se esperarían ver asociadas con cada resultado.

Definición 3 (Función de densidad de probabilidad (FDP)). La función de densidad de probabilidad (FDP) o, simplemente, función de densidad, representada comúnmente como $f(x)$, se utiliza con el propósito de conocer cómo se distribuyen las probabilidades de un suceso o evento, en relación al resultado del suceso.

La FDP es la derivada (ordinaria o en el sentido de las distribuciones) de la función de distribución de probabilidad $F(x)$, o de manera inversa, la función de distribución es la integral de la función de densidad:

$$F(x) = \int_{-\infty}^x f(t) dt$$

La función de densidad de una v.a. determina la concentración de probabilidad alrededor de los valores de una variable aleatoria continua.

Definición 4 (Función cuantil). En probabilidad la función cuantil de una distribución de probabilidad es la inversa de la función de distribución. Dada una función de distribución continua y estrictamente monótona, $F : \mathbb{R} \rightarrow (0,1)$, la función cuantil, F^{-1} , devuelve un valor x tal que $P(X \leq x) = p$.

Si la distribución de probabilidad es discreta, en lugar de continua, entonces puede haber saltos entre los valores en el dominio de su función de distribución, mientras que si la función de distribución es monótona no estricta, puede haber "zonas llanas"(intervalos en los que el valor de la función se mantiene constante) en su rango. En cualquiera de los casos, la función no estaría bien definida, por lo que se establece la siguiente definición alternativa:

$$F^{-1}(p) = \inf \{x \in \mathbb{R} : p \leq F(x)\}$$

para una probabilidad $0 < p < 1$, devolviendo la función cuantil el valor mínimo de x para el cual se mantiene la probabilidad anterior.

Observación 1. De aquí en adelante escribiremos F^{\leftarrow} en lugar de F^{-1} para referirnos a la función cuantil.

El problema genérico en la estimación del riesgo en todos los casos antes mencionados, se pueden expresar en la forma siguiente:

Definición 5 (Problema de estimación de riesgo). *Dada una muestra x_1, x_2, \dots, x_n de variables aleatorias independientes e idénticamente distribuidas iid con función de distribución de probabilidad $F(\cdot)$ desconocida, construir una función que aproxime la cola de esta distribución a la derecha de una abscisa $u = X_{n-k}$ donde u es conocido como umbral y los valores que puede asumir dependerán de la selección de k , que determina la fracción muestral de la cola (tail sample fraction), $\frac{k}{n}$.*

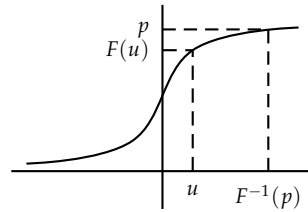


Figura 1.1. cola de la distribución $F(\cdot)$ a la derecha de la abscisa u , que es la parte del gráfico de $F(\cdot)$ a la derecha del punto $(u, F(u))$.

Objetivo: Encontrar x tal que $F(x) = p$ para p dada. La razón indica que la solución del problema sería resuelto tomando simplemente la función inversa $x = F^{-1}(p)$, pero la complicación es que la función $F(\cdot)$ es desconocida. Así que se debe seguir un método de solución compuesto por los siguientes pasos.

PROCEDIMIENTO PARA REALIZAR LA ESTIMACIÓN DEL CUANTIL

1. Escoger un valor u^{opt} , denominado umbral óptimo a la derecha del cual será estimada la cola de la función desconocida $F(\cdot)$. Esta no es tarea fácil, pues depende de la fracción muestral, algunas técnicas para la selección de la fracción muestral son usadas y comparadas en Beirlant *et al.* (2004) Beirlant J. *et al.* (2004) y Matthys *et al.* Matthys, G. *et al.* (2004). La selección de u^{opt} también depende de las propiedades de la función $F(\cdot)$ deducidas con base en la muestra.

2. Construir una aproximación para la cola de $F(\cdot)$ a la derecha de u^{opt} , usando una de las funciones siguientes, estas son las funciones generadoras de las distribuciones de Pareto generalizadas que veremos en la próxima sección:

$$1 - \frac{1}{(1+x)^\alpha}; \quad \text{para } \alpha > 0, \quad x \geq 0 \quad (1.1)$$

$$1 - e^{-x}; \quad \text{para } x \geq 0 \quad (1.2)$$

$$1 - (1-x)^\gamma; \quad \text{para } \gamma > 0, \quad x \in [0, 1] \quad (1.3)$$

Estas funciones proporcionan gráficas muy simples y con el fin de lograr un buen ajuste en un problema particular se puede aplicar una transformación como la dada por la traslación o, reducir la gráfica multiplicando por un factor η , con $0 < \eta < 1$, y ampliar la gráfica multiplicando por un factor δ , con $\delta > 1$. A continuación se muestran las funciones generadoras de las distribuciones generalizadas de Pareto en el plano para algunos valores de los parámetros α y γ .

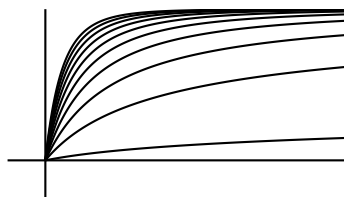


Figura 1.2. Gráfica de (1.1): (ramas de hipérbolas horizontales)

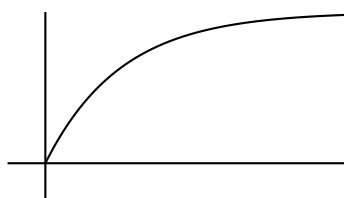


Figura 1.3. Gráfica de (1.2) (la función de distribución exponencial)

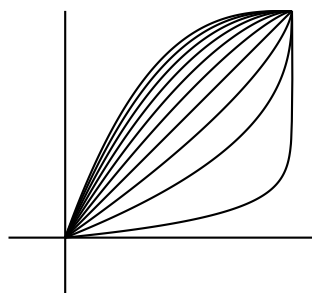


Figura 1.4. Gráfica de (1.3): (ramas de parábolas)

Usamos las familias de las funciones (1.1)-(1.3) para generar aproximaciones de las colas de funciones de distribución $F(\cdot)$ a partir de una abscisa que se ha llamado umbral óptimo. Note que, en primer lugar, para usar alguna de las aproximaciones en las ecuaciones (1.1)-(1.3) es necesario calcular los parámetros α o γ , y en segundo lugar esta forma de proceder debe garantizar que a partir de una abscisa T la diferencia entre la cola de la distribución $F(\cdot)$ con la del aproximador debe ser casi cero y que se debe tener condiciones sobre la $F(\cdot)$ para que tal límite exista. Esta situación se evaluará nuevamente en las próximas secciones pero se adelanta que no todas las funciones pueden ser bien aproximadas siguiendo

este procedimiento que se acaba de describir. Se desea cerrar esta sección introduciendo las distribuciones de valores extremos y en la siguiente sección las llamadas funciones generalizadas de Pareto (GPD del inglés *Generalized Pareto Distribution*),

Definición 6 (Distribuciones de valores extremos). *Las distribuciones de valores extremos están implícitamente en el teorema 2.2, en su forma más general como límites de sucesiones de distribuciones. Ahora, las distribuciones de valores extremos en su forma más sencilla vienen dadas por:*

a) La familia de Fréchet compuesta por las distribuciones $\Phi_\alpha(\cdot)$ con $\alpha > 0$ donde

$$\Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ e^{-x^{-\alpha}} & \text{si } x > 0 \end{cases} \quad (1.4)$$

b) La familia de Weibull son las distribuciones $\Psi_\alpha(\cdot)$ con $\alpha > 0$ donde

$$\Psi_\alpha(x) = \begin{cases} e^{-[(-x)^\alpha]} & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases} \quad (1.5)$$

c) La distribución de Gumbel se define por

$$\Lambda(x) = e^{-e^{-x}} \quad \text{para todo } x \in \mathbb{R} \quad (1.6)$$

1.3 DISTRIBUCIONES DE PARETO GENERALIZADAS GPD

Se acostumbra llamar distribución generalizada de Pareto a una función que tiene una de las formas de las ecuaciones (1.7), (1.8) o (1.9). Las funciones (1.7) y (1.8) fueron estudiadas por Pareto, y la función (1.9) se puede ver como una generalización de (1.7) y (1.8), de hay el término de distribuciones generalizadas de Pareto que se usa para referirse a cualquiera de estas tres familias de funciones.

Definición 7. *Sea ξ cualquier número real, $\xi \in \mathbb{R}$, S cualquier número real, $S \in \mathbb{R}$, y β un número que admite solo valores reales positivos. Las funciones generalizadas de Pareto se definen por medio de las siguientes tres ecuaciones*

$$\xi > 0 \longrightarrow G_{\xi,\beta,S}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x-S}{\beta}\right)^{-\frac{1}{\xi}} & x \geq S \\ 0 & x < S \end{cases} \quad (1.7)$$

$$\xi = 0 \longrightarrow G_{0,\beta,S}(x) = \begin{cases} 1 - e^{-\frac{x-S}{\beta}} & x \geq S \\ 0 & x < S \end{cases} \quad (1.8)$$

$$\xi < 0 \longrightarrow G_{\xi, \beta, S}(x) = \begin{cases} 0 & x < S \\ 1 - \left(1 + \xi \left(\frac{x-S}{\beta}\right)\right)^{-\frac{1}{\xi}} & x \in \left[S, S - \frac{\beta}{\xi}\right] \\ 1 & x > S - \frac{\beta}{\xi} \end{cases} \quad (1.9)$$

Note que la expresión de una GPD posee tres parámetros denotados por ξ , β y S . Estos parámetros son conocidos en la literatura de la siguiente manera

β es el parámetro de escala.

S es el parámetro de traslación o localización.

ξ es el parámetro de forma o índice.

Se advierte que cualquiera de estos tres tipos de funciones se pueden obtener a partir de las GPD con $\beta = 1$ y $S = 0$ (es decir, que tienen un solo parámetro ξ). Esta transformación suele seguir las ideas de traslación de curvas y expansiones de acuerdo al siguiente hecho:

Sea X una variable aleatoria cualquiera, sea $d > 0$ y $c > 0$ constantes reales arbitrarias

$$Y = cX + d$$

Sea:

$F_X(\cdot)$: Función de distribución de X

$F_Y(\cdot)$: Función de distribución de Y

Entonces

$$\begin{aligned} F_Y(x) &= P[Y \leq x] = P[cX + d \leq x] \\ &= P\left[X \leq \frac{x-d}{c}\right] = F_X\left(\frac{x-d}{c}\right) \quad \forall x \in \mathbb{R} \end{aligned}$$

esta relación sustenta que si la función de distribución de una variable aleatoria X se transforma mediante los dos pasos: primero estirar c veces el gráfico de F_X a lo largo del eje x ; y segundo trasladar d unidades a la derecha el gráfico obtenido en el paso anterior, el resultado será una nueva función de distribución. Además, si se designa por Y la variable aleatoria distribuida según esta nueva función de distribución, se tiene que Y y X se relacionan via $Y = cX + d$.

La transformación descrita en el párrafo anterior deja claro que cualquier GPD con parámetros ξ , β y S es obtenida de la GPD con el mismo valor de ξ , $\beta = 1$ y $S = 0$, donde los correspondiente c y d son $c = \beta$ y $d = S$, o equivalentemente, el

gráfico de la GPD con parámetros ξ , β y S es el gráfico de la GPD con el mismo valor de ξ , $\beta = 1$ y $S = 0$ después de haber sido estirado β veces a lo largo del eje x y trasladado S unidades horizontalmente.

Este hecho justifica los nombre dados a los parámetros β (parámetro de escala) y S (parámetro de localización). A seguir vamos a ilustrar las GPD con $\beta = 1$ y $S = 0$ para varios valores del parámetro de forma o índice de cola ξ .

Con $\xi > 0$

$$G_{\xi}(x) = \begin{cases} 0 & \text{para } x < 0 \\ 1 - (1 + \xi x)^{-\frac{1}{\xi}} & \text{para } x \geq 0 \end{cases}$$

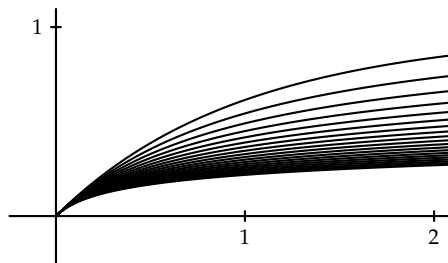


Figura 1.5. Distribución de Pareto para el parámetro de escala $\beta = 1$, parámetro de localización $S = 0$ y varios valores positivos para el índice de cola ξ

Con $\xi = 0$

$$G_0(x) = \begin{cases} 0 & \text{para } x < 0 \\ 1 - e^{-x} & \text{para } x \geq 0 \end{cases}$$

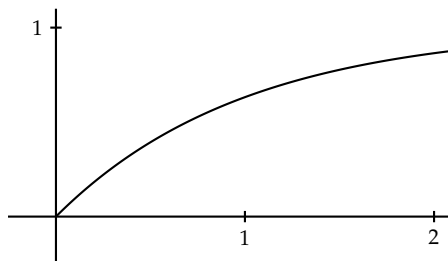


Figura 1.6. Distribución de Pareto para el parámetro de escala $\beta = 1$, parámetro de localización $S = 0$ e índice de cola $\xi = 0$

Con $\zeta < 0$

$$G_{\zeta}(x) = \begin{cases} 0 & \text{para } x < 0 \\ 1 - (1 + \zeta x)^{-\frac{1}{\zeta}} & \text{para } x \in \left[0, \frac{1}{\zeta}\right] \\ 1 & \text{para } x > \frac{1}{\zeta} \end{cases}$$

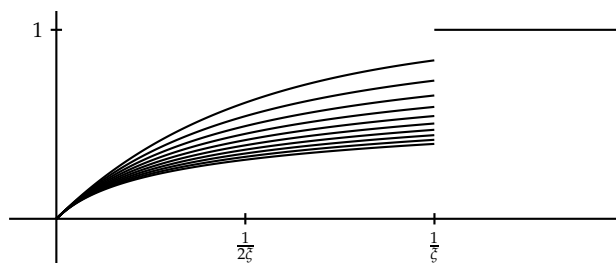


Figura 1.7. Distribución de Pareto para el parámetro de escala $\beta = 1$, parámetro de localización $S = 0$ y varios valores negativos para el índice de cola ζ

CAPÍTULO 2

RESULTADOS TEÓRICOS

2.1 TEORÍA BÁSICA DE VALOR EXTREMO

INTRODUCCIÓN

Inicialmente se presentan dos teoremas que dan varias relaciones de equivalencia, las cuales son la base para probar que una función de distribución de un conjunto de datos pertenece al dominio de atracción de una distribución límite: bien sea la Distribución de valor Extremo o Distribución Generalizada de Pareto.

El teorema de Fisher, Tippett y Gnedenko de los extremos de una muestra, es fundamental ya que a partir de él se reconoce la distribución límite de valores extremos con el parámetro de forma que es el esencial dentro de todo el proceso.

Posteriormente se presentan varios teoremas sobre las condiciones necesarias y suficientes para que una distribución determinada pertenezca al dominio de atracción de Gumbel, Frechet o Weibull.

El teorema de Von Mises presenta condiciones suficientes para que una distribución determinada pertenezca al dominio de atracción de una distribución de valores extremos y adicionalmente proporciona una forma de evadir las distribuciones de probabilidad degeneradas.

Más adelante se presentan teoremas sobre la distribución generalizada de Pareto que es fundamental en el ajuste de los valores de variables aleatorias que están por encima de un umbral alto.

Adicionalmente, a manera conceptual, se presentan algunos métodos para estimar el parámetro de forma de las distribuciones de valor extremo. Uno de estos es el denominado Método de Hill para cuando el parámetro de forma toma el valor positivo ($\xi > 0$). Se destaca este método porque en el área de la finanzas este

parámetro es un valor real positivo. Del parámetro de forma se analiza específicamente propiedades como consistencia débil y la normalidad asintótica, con el fin de tener elementos para formular pruebas de hipótesis e intervalos de confianza de este parámetro que es el que define el tipo de cola de la distribución.

Finalmente los extremos son estudiados considerando el proceso de excesos sobre un umbral, correspondientes al método de Picos Sobre un Umbral. En este contexto introducimos la Distribución Generalizada de Pareto (GPD por sus siglas en inglés) a la que hace referencia el teorema de Pikand, los parámetros de forma y escala serán estimados usando el método de optimización introducido en el capítulo tres para maximizar una función de dos variables no lineal. Con la Distribución Generalizada de Pareto obtenida a partir de la muestra se procede a realizar la estimación del cuantíl de interés.

El papel que juega la distribución generalizada de Pareto en el método de picos sobre el umbral es similar al papel que juega la distribución generalizada de valores extremos en el método de máximos por bloques. Como el objetivo de este trabajo es el estudio de los procesos de excesos sobre un umbral se centrará la atención en las funciones GPD.

La teoría sobre valor extremo descrita en esta tesis es bastante reciente: Solamente en 1980s los contornos de la teoría estadística tomaron forma. La teoría de valor extremo probabilística en una dimensión fue desarrollada por M Frechet (1927), R. Fisher y L. Tippett (1928) y R von Mises (1936) y culminó en el trabajo de B. Gnedenko (1943). La teoría estadística fue iniciada por J. Pickands III (1975).

La teoría asintótica de los extremos de la muestra ha sido desarrollada en paralelo con la teoría central del límite y en efecto las dos teorías tienen en común que mientras en la teoría central del límite la distribución a la cual converge la media es a una distribución normal estándar después de restarle la media de la población y dividir por la varianza de la media que es igual a $\frac{\sigma}{\sqrt{n}}$ valor que depende de n el máximo o el mínimo es decir los valores extremos.

Sea $X_1, X_2, X_3 \dots$ variables aleatorias independientes e idénticamente distribuidas. La teoría central del límite está concentrada con el comportamiento límite de las sumas parciales¹ $S_n = X_1 + X_2 + \dots + X_n$ cuando $n \rightarrow \infty$ mientras la teoría de los extremos de una muestra está concentrada con el comportamiento límite

¹Para el caso del teorema central del límite con respecto a \bar{X} , $a_n = \frac{\sigma}{\sqrt{n}}$ y $b_n = \mu$

del máximo y mínimo, extremos muestrales $\max(X_1, \dots, X_n)$ o $\min(X_1, \dots, X_n)$ cuando $n \rightarrow \infty$.

La teoría de valores extremos es el área de la estadística dedicada a desarrollar modelos y técnicas para estimar el comportamiento de eventos inusuales o raros. Los eventos raros son los que pertenecen a las colas de la distribución y son los que están alejados del grueso de la distribución (de la media o mediana). Otra manera de definir los eventos extremos son aquellos valores que están por encima de un valor denominado umbral, pero también se puede definir como el máximo o el mínimo de una variable en un cierto período.

Uno de los principales teoremas basados en la teoría asintótica un tipo de teorema central de límite versión valores extremos derivado en principio por Fisher y Tippet (1928) de manera heurística y luego desde un punto de vista más riguroso por Genedenko (1943) establecieron que hay solamente tres tipos de distribuciones (Distribución tipo Gumbel, Tipo Frechet y tipo Weibull) que se alcanzan como distribuciones límite de los máximos y mínimos de algún proceso aleatorio sobre intervalos de tiempo fijos.

Durante los años cincuenta Von Misses (1954) y Jenkisson (1955) unificaron las tres distribuciones en una sola denominada distribución generalizada de valores extremos (GEV) la cual depende de tres parámetros. Esta unificación de las tres distribuciones en una sola permitió simplificar los procedimientos de modelaje.

Otro desarrollo que se inició a partir de la década del setenta por Pickands (1975) y que se continuo en las décadas de los ochenta y noventa se basó en los métodos basados en los excesos por encima de un umbral en lugar del máximo o el mínimo de un periodo pero acá se origina otro tipo de distribución conocida como la distribución generalizada de Pareto la cual fue introducida por Pickands (1975), su aplicación se hace en el análisis de eventos extremos en el modelamiento de grandes derechos en seguros, en distribuciones de tiempos de falla en estudios de confiabilidad y en cualquier situación en la cual se requiera debido a la presencia de colas pesadas o colas ligeras.

Otro de los avances en teoría de Valor Extremo que se ha producido en los últimos años es extender esta teoría a procesos no estacionarios y dependientes además de desarrollar la inferencia para extremos multivariados.

2.1.1 Formulación de la relación límite para el máximo y el mínimo

Las distribuciones de valores extremos surgen formalmente como distribuciones límite para el máximo y el mínimo de una sucesión de variables aleatorias. Los conceptos fundamentales en los que se basa la teoría de valores extremos son los modos de convergencia: Convergencia en distribución, convergencia en probabilidad, convergencia casi segura y convergencia de funciones inversas generalizadas. Por ahora definamos lo que es convergencia en probabilidad.

Definición 8 (Convergencia en probabilidad). *Se dice que una sucesión $\{X_n\}$ de v.a. converge en probabilidad a la variable aleatoria X si para $\forall \varepsilon > 0$*

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$$

lo cual equivale a que

$$\forall \varepsilon > 0 \quad P(|X_n - X| > \varepsilon) \rightarrow 0,$$

cuando $n \rightarrow \infty$ la convergencia en probabilidad se denota por $X_n \xrightarrow{P} X$.

Supongamos que son variables aleatorias independientes e idénticamente distribuidas con función de densidad de probabilidad

$$p_{x_j}(x) = f(x), \quad j = 1, 2, \dots, n$$

Sea F la función de distribución acumulada para la v.a. X y X^* su punto final superior esto es $x^* = \{x : F(x) < 1\}$ el cual puede ser infinito entonces

$$\max(X_1, \dots, X_n) \xrightarrow{P} x^*, \quad n \rightarrow \infty$$

donde

$$F(x) = P(X_j \leq x) \quad \forall_{j,x} \quad (2.1)$$

Luego para el máximo

$$X_n = M_n = \max\{X_1, X_2, \dots\} \quad (2.2)$$

La función de distribución acumulada, en teoría, viene dada por

$$\begin{aligned} P(M_n \leq x) &= P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\ &= [P(X_i \leq x)]^n = F^n(x) \end{aligned} \quad (2.3)$$

Para resolver el problema del desconocimiento de la función de distribución acumulada se pueden implementar dos soluciones la primera, es ajustar una distribución con los datos observados y así estimar aproximadamente el valor de F pero si hay algún error este se hará muy grande ya que F esta elevado a la n .

La otra solución es aceptar que F es desconocido y tratar de obtener la distribución aproximada que puede tener $F^n(x)$ y que solo se puede lograr a partir de datos extremos con una teoría análoga al teorema central del límite.

Pero hay que tener en cuenta que se puede tener el siguiente caso: en el límite la distribución puede ser degenerada.

Definición 9 (Distribución Degenerada). Una variable aleatoria X se dice que tiene una distribución degenerada en un punto h si su función de densidad de probabilidad es

$$p_x(x) = \begin{cases} 1 & \text{si } x = h \\ 0 & \text{si } x < h \end{cases}$$

La función de distribución de una variable aleatoria degenerada es:

$$F(x) = \begin{cases} 0 & \text{si } x < h \\ 1 & \text{si } x \geq h \end{cases}$$

Por ejemplo para cualquier $x < x^*$ donde x^* es el punto superior de F , es decir x^* es el valor más pequeño de x tal que se cumple que

$$\lim_{n \rightarrow \infty} F^n(x) = 0$$

por lo que la función de distribución de M_n degenera en un punto de masas en x^* .

Cuando n tiende a infinito se tiene que para cualquier valor fijo de x

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

El cual es una función degenerada. Para evitar esta dificultad, se normaliza y centraliza la variable M_n :

$$M_n^* = \frac{M_n - b_n}{a_n} \tag{2.4}$$

El máximo de la variable aleatoria puede ser estabilizado a través de las sucesiones $\{a_n\}$ y $\{b_n\}$ con $a_n > 0$ y $b_n \in R$ que corresponden a los parámetros de escala y de localización tales que

$$\begin{aligned} P\left(\frac{M_n - b_n}{a_n} \leq x\right) &= P(M_n \leq a_n x + b_n) \\ &= F^n(a_n x + b_n) \xrightarrow{n \rightarrow \infty} G(x) \Leftrightarrow \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x) \end{aligned} \tag{2.5}$$

La escogencia apropiada de $\{a_n\}$ y $\{b_n\}$ estabiliza la escala y la localización de M_n^* cuando se incrementa n , y así se evita la degeneración de M_n . Como se verá más adelante, de esta manera M_n^* converge en distribución a alguna de las familias, Gumbel, Frechet o Weibull (ver definición 6).

Ejemplo 2.1. *Para explicar las razones del surgimiento de las constantes a_n y b_n , dados en (2.4): Considere una sucesión de variables aleatorias X_1, X_2, \dots donde cada X_i posee una distribución lognormal. Esto significa que cada X_i es distribuida como e^Z , Z es la variable aleatoria normal estándar. A continuación se desea simular 10000 realizaciones de la variable aleatoria $\max\{X_1, \dots, X_{70}\}$ y se construye el histograma de estas realizaciones de esta variable aleatoria con base en esas 10000 valores. Este histograma se presenta en la figura 2.2. Debido a los argumentos de la estadística, este histograma se puede considerar como una aproximación de la función de densidad de la distribución de la variable aleatoria $\max\{X_1, \dots, X_{70}\}$.*

Debe observarse también los histogramas presentados en las figuras, 2.1 y 2.3 donde se esbozan las densidades de las variables aleatorias

$$\max\{X_1\} \text{ y } \max\{X_1, \dots, X_{700}\}$$

se percibe a partir de la secuencia de las figuras que los histogramas se separan a la derecha y quedan cada vez más dispersos. Justamente lo que se pretende con las constantes b_n y a_n es “empujar” la distribución de las variables aleatorias,

$$\max\{X_1\} \text{ y } \max\{X_1, \dots, X_n\}$$

a la izquierda cuando $b_n \geq 0$, mientras que la constante a_n la estira a_n veces.

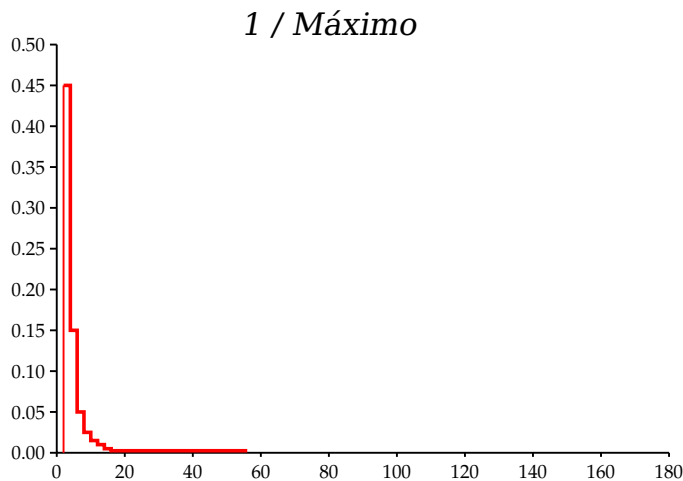


Figura 2.1. $\max\{X_1\}$

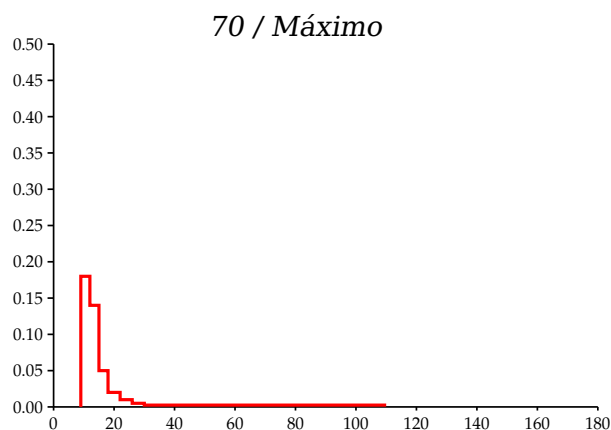


Figura 2.2. $\text{máx}\{X_1, \dots, X_{70}\}$

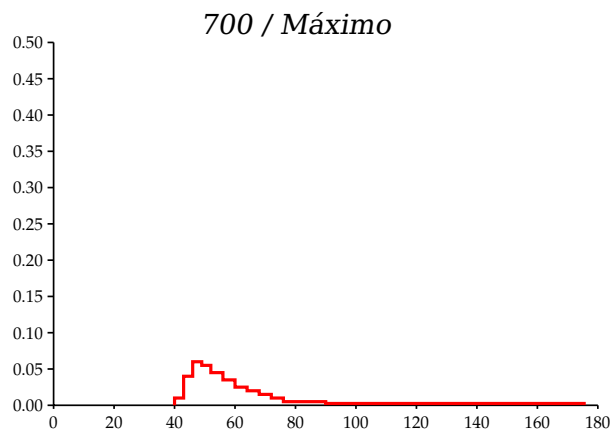


Figura 2.3. $\text{máx}\{X_1, \dots, X_{700}\}$

Las figuras 2.1, 2.2 y 2.3 ilustran las densidades de las distribuciones de las variables aleatorias para $\text{máx}\{X_1\}$, $\text{máx}\{X_1, \dots, X_{70}\}$ y $\text{máx}\{X_1, \dots, X_{700}\}$ respectivamente, las cuales fueron formadas por los histogramas de las muestras correspondientes a estas variables. Las muestras fueron obtenidas por simulación y la distribución de cada X de este ejemplo es lognormal.

2.1.2 Formulaciones alternativas de la relación límite

Se pueden obtener las siguientes relaciones equivalentes a (2.5), que servirán como base para posteriores demostraciones.

Se quiere mostrar que

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

es equivalente a

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

Demostración. Partiendo de (2.5)

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

Tomando logaritmo a ambos lados

$$n [\ln(F(a_n x + b_n))] = \ln[G(x)]$$

Sustituyendo $\bar{F} = 1 - F$ se obtiene que:

$$\begin{aligned} n[\ln(1 - \bar{F}(a_n x + b_n))] &= \ln[G(x)] \\ -n[-\ln(1 - \bar{F}(a_n x + b_n))] &= \ln[G(x)] \end{aligned}$$

Del cálculo elemental se sabe que $-\ln(1 - x) \sim x$ cuando $x \rightarrow 0$, luego usando este hecho,

$$-\ln(1 - \bar{F}(a_n x + b_n)) \sim \bar{F}(a_n x + b_n)$$

porque

$$\lim_{n \rightarrow \infty} \bar{F}(a_n x + b_n) = 0$$

por lo tanto

$$\lim_{n \rightarrow \infty} \bar{F}(a_n x + b_n) = \lim_{n \rightarrow \infty} (1 - F(a_n x + b_n)) \quad (2.6)$$

Aplicando la ecuación (2.5) a la expresión (2.6), se obtiene que:

$$\begin{aligned} \lim_{n \rightarrow \infty} \bar{F}(a_n x + b_n) &= \lim_{n \rightarrow \infty} \left(1 - (G(x))^{\frac{1}{n}}\right) \\ &= 1 - (G(x))^0 = 1 - 1 = 0 \end{aligned}$$

Entonces se tiene que

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

o lo que es lo mismo

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

ó

$$\lim_{n \rightarrow \infty} \frac{1}{n[1 - F(a_n x + b_n)]} = \frac{1}{-\ln(G(x))} \quad (2.7)$$

□

Teorema 2.1 (Equivalencia de enunciados). Sean $\{a_n\}$ y $\{b_n\}$ sucesiones de números reales, y G una función de distribución no degenerada, entonces,

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

2. $\forall x \in G(x)$, continuo y $0 < G(x) < 1$; $a(t) := a_{\llbracket t \rrbracket}$ (con $\llbracket t \rrbracket$ la parte entera de t)

$$\lim_{t \rightarrow \infty} t(1 - F(a(t)x + b(t))) = -\log G(x) \quad (2.8)$$

3. $\forall x > 0$ punto de continuidad de $D(x) = G^{\leftarrow}(e^{-\frac{1}{x}})$, $a(t) := a_{\llbracket t \rrbracket}$, y $b(t) := b_{\llbracket t \rrbracket}$

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x), \quad (2.9)$$

entonces 1. \Leftrightarrow 2, 2. \Leftrightarrow 3.

Teorema 2.2 (Fisher y Tippett (1928), Gnedenko (1943)). El teorema de Fisher, Tippett y Gnedenko es un resultado importante por muchos aspectos, es el teorema fundamental de la Teoría de Valor Extremo. Muestra que la función de distribución límite forma una familia de un parámetro de forma simple y explicito aparte de los parámetros de localización y escala.

El teorema que se probará a continuación identifica la clase de distribuciones no degeneradas que puede ocurrir como límite de la relación (2.5).

Sean $\{a_n\}$ y $\{b_n\}$ sucesiones de números reales y $M_n = \max\{X_1, X_2, \dots\}$, si

$$\begin{aligned} P\left(\frac{M_n - b_n}{a_n} \leq x\right) &= P(M_n \leq a_n x + b_n) \\ &= F^n(a_n x + b_n) \xrightarrow[n \rightarrow \infty]{} G(x) \Leftrightarrow \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x) \end{aligned}$$

Entonces la distribución de valor extremo es

$$G_{\xi}(x) = G(x, \xi) = \begin{cases} \exp\left(-\left(1 + \xi \frac{x - D(1)}{H'(0)}\right)^{-\frac{1}{\xi}}\right), & 1 + \xi x > 0 \\ \exp\left(-e^{-\frac{x - D(1)}{H'(0)}}\right) & x \in \mathbb{R}, \xi = 0 \end{cases} \quad (2.10)$$

Donde ξ es como α en la definición 6.

Demostración. Consideremos la distribución límite:

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

Si 1 es un punto de continuidad de la función $D(x)$ se obtiene

$$D(1) = \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = 0$$

Sea $E(x) = D(x) - D(1)$ entonces para los puntos donde hay continuidad $x > 0$

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D(x) - D(1) =: E(x) \quad (2.11)$$

$y > 0$

$$\frac{U(txy) - U(t)}{a(t)} = \frac{U(txy) - U(ty)}{a(ty)} \frac{a(ty)}{at} + \frac{U(ty) - U(t)}{a(t)} \quad (2.12)$$

Afirmamos que $\lim_{n \rightarrow \infty} \frac{U(ty) - U(t)}{a(t)}$ y $\lim_{n \rightarrow \infty} \frac{a(ty)}{a(t)}$ existen.

Supongamos que no.

Entonces hay A_1, A_2, B_1, B_2 con $A_1 \neq A_2$ o $B_1 \neq B_2$, donde B_i son puntos límites de $\frac{U(ty) - U(t)}{a(t)}$ y A_i son puntos de $\lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)}$ cuando $t \rightarrow \infty$

Se encuentra de (2.12) y de las definiciones de A_i y B_i que

$$E(xy) = E(x)A_i + B_i \quad i = 1, 2 \quad (2.13)$$

Para $i = 1$ $E(xy) = E(x)A_1 + B_1$

Para $i = 2$ $E(xy) = E(x)A_2 + B_2$

Restando y factorizando se tiene

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

ya que E no puede ser constante (G no puede ser degenerada). Se puede concluir que $A_1 = A_2$ y $B_1 = B_2$, por lo tanto $A(y) := \lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)}$ existe para $\forall x > 0$ y para $\forall x, y > 0$

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

por lo tanto para: $s = \log x$, $t = \log y$ ($x, y \neq 1$), como $H(x) = E(e^x)$

$$H(t+s) = E(e^{t+s}) = E(e^s e^t) = E(e^s)A(e^t) + E(e^t)$$

$E(x)(A_1 - A_2) = B_1 - B_2$ para todo $x > 0$. Ya que E no puede ser constante (recordar que G es no degenerada) nosotros debemos tener que $A_1 = A_2$ y por lo tanto también $B_1 = B_2$. Conclusión:

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

existe para $y > 0$, y para $x, y > 0$

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

Por lo tanto para $s := \log x$, $t = \log y$ ($x, y \neq 1$), y $H(x) := E(e^x)$, se tiene

$$\begin{aligned} H(t+s) &= E(e^{t+s}) = E(e^t e^s) = E(e^s e^t) = E(e^s)A(e^t) + Ee^t \\ &= H(s)A(e^t) + H(t) \end{aligned} \tag{2.14}$$

Lo cual puede ser escrito como (ya que $H(0) = 0$)

$$\frac{H(t+s) - H(t)}{s} = \frac{H(s) - H(0)}{s} A(e^t) \tag{2.15}$$

Hay ciertamente una t en la cual H es diferenciable (ya que H es monótona) por lo tanto (2.15) H es diferenciable en todo valor y

$$\begin{aligned} \lim_{s \rightarrow \infty} \frac{H(t+s) - H(t)}{s} &= \lim_{s \rightarrow \infty} \frac{H(s) - H(0)}{s} A(e^t) \\ H'(t) &= H'(0)A(e^t) \end{aligned} \tag{2.16}$$

La ecuación (2.15) puede ser escrita como:

$$H(t+s) = H(s)A(e^t) + H(t) \tag{2.17}$$

(ya que $H(0) = 0$). Definamos: $Q(t) = \frac{H(t)}{H'(0)}$. Note que $H'(0)$ no puede ser cero.

Por (2.16) se tiene $H'(t) = H'(0)A(e^t)$ para $t = 0$ se tiene

$$H'(0) = H'(0)A(e^0) \rightarrow 1 = A(1)$$

H no puede ser constante ya que G es no degenerada. Luego

$$Q(0) = 0 \quad , \quad Q'(0) = 1 \quad , \quad Q(t+s) - Q(t) = Q(s)A(e^t),$$

aplicando (2.14) y por (2.16) Sustrayendo la misma expresión con t y s intercambiados se consigue

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

Por lo tanto

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

ya que

$$\lim_{s \rightarrow 0} \frac{Q(s)}{s} = \lim_{s \rightarrow 0} \frac{Q(s) - Q(0)}{s} = Q'(0) = 0$$

ya que

$$Q(t) = \frac{H(t)}{H'(t)} \Rightarrow Q'(t) = \frac{[H'(t)]^2 - H(t)H''(t)}{[H'(t)]^2} \Rightarrow Q'(0) = 1$$

Derivando nuevamente se obtiene:

$$Q'(t)Q''(0) + Q(t)Q'''(0) = Q''(t) \Rightarrow Q''(0)Q(t) = Q''(t)$$

Pero como $(\log Q')'(t) = Q''(0) = \xi \in R$ para todo t . Se sigue que

$$Q(t) = \int_0^t e^{t\xi} ds \quad \text{ya que } Q(0) = 1$$

Esto significa que

$$H(t) = H'(0) \frac{e^{\xi t} - 1}{\xi} \quad \text{y} \quad D(t) = D(1) + H'(0) \frac{t^\xi - 1}{\xi}$$

Por lo tanto

$$D^{\leftarrow}(x) = \left(1 + \xi \frac{x - D(1)}{H'(0)} \right)^{\frac{1}{\xi}} \quad (2.18)$$

Ahora $D(x) = G^{\leftarrow} \left(e^{-\frac{1}{x}} \right)$, y por lo tanto

$$D^{\leftarrow}(x) = \frac{1}{-\log G(x)} \quad (2.19)$$

Combinando (2.18) y (2.19) se obtiene la prueba del teorema.

$$\frac{1}{-\log G(x)} = \left(1 + \xi \frac{x - D(1)}{H'(0)} \right)^{\frac{1}{\xi}}$$

invirtiendo se obtiene:

$$-\log G(x) = \left(1 + \xi \frac{x - D(1)}{H'(0)} \right)^{-\frac{1}{\xi}}$$

multiplicando por -1 a ambos lados y aplicando la función exponencial, se tiene:

$$G(x, \xi) = \exp \left(- \left(1 + \xi \frac{x - D(1)}{H'(0)} \right)^{-\frac{1}{\xi}} \right)$$

Definición 10 (Cola de una distribución). *Se dice que una variable aleatoria X tiene una distribución de colas pesadas si la misma posee el coeficiente de curtosis superior al de la distribución normal, es decir, para μ_X la media de X y σ_X la desviación estándar de X , se cumple que*

$$E \left[\left(\frac{X - \mu_X}{\sigma_X} \right)^4 \right] > 3$$

Esta definición es un poco incompleta en el sentido que si una variable aleatoria no posee cuarto momento nada podría decirse con respecto a si tiene cola pesada o no, por esta razón se utiliza la definición de Sigman K. (1999), esta es, dada una variable aleatoria X , con función de distribución $F(x) = P(X \leq x)$ y función de sobrevivencia $\bar{F}(x) = 1 - F(x)$, se dice que X posee cola pesada si $\bar{F}(x) \geq 0 \forall x \geq 0$ y

$$\lim_{x \rightarrow \infty} P(X > x + y | X > x) = \lim_{x \rightarrow \infty} \frac{\bar{F}(x + y)}{\bar{F}(x)} = 1, \quad y \geq 0$$

De este modo, se puede verificar con base en la última igualdad que toda distribución de cola pesada posee función generadora de momentos infinita, es decir,

$$E(e^{tX}) = \infty$$

Se denotará por L a la colección de todas las distribuciones de probabilidad que poseen cola pesada, es decir,

$$L = \{F \mid F \text{ es una distribución de cola pesada.}\}$$

Dentro de esta clase de distribuciones existen dos subclases:

a) La distribución subexponencial (poseen colas con caídas más lentas que la cola de cualquier distribución exponencial con media $1/\lambda > 0$) denotadas por $S = \{F \mid F \text{ es una distribución subexponencial}\}$.

b) La distribución de variación regular R , donde

$$R = \{F \mid F \text{ es una distribución de variación regular}\}$$

Definición 11 (cola subexponencial). *Una variable aleatoria positiva X se dice que tiene una distribución subexponencial, si para todo $n \geq 2$*

$$\lim_{x \rightarrow \infty} \frac{P(X_1 + X_2 + \dots + X_n)}{\bar{F}(x)} = n$$

Afirmamos de nuevo que el nombre subexponencial se da por el hecho que la cola de cualquier distribución de este tipo decae en forma más lenta que cualquier distribución exponencial con media $1/\lambda > 0$, es decir,

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x)}{e^{-\lambda x}} = \lim_{x \rightarrow \infty} e^{\lambda x} \bar{F}(x) = \infty$$

La distribución Weibull, que fue introducida en la ecuación (1.5) del texto, posee cola pesada y pertenece a la clase de distribuciones subexponencial cuando su parámetro de forma es inferior a uno. La función de sobrevivencia de Weibull es $\bar{F}(x) = e^{-(\frac{x}{\alpha})^\beta}$, $x > 0$, donde α ($\alpha > 0$) es conocido como el parámetro de escala y β ($\beta > 0$) como el parámetro de forma. Suponiendo que $0 < \beta < 1$

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x)}{e^{-\lambda x}} = \lim_{x \rightarrow \infty} e^{\lambda x} e^{-(\frac{x}{\alpha})^\beta} = \lim_{x \rightarrow \infty} e^{-\lambda x - (\frac{1}{\alpha})^\beta x^\beta} \quad (2.20)$$

$$= \exp \left\{ \lim_{x \rightarrow \infty} \left[x \left(\lambda - \frac{1}{\alpha^\beta} x^{\beta-1} \right) \right] \right\} \quad (2.21)$$

$$= \exp \left\{ \left[\lim_{x \rightarrow \infty} x \right] \left[\lim_{x \rightarrow \infty} \left(\lambda - \frac{1}{\alpha^\beta} x^{\beta-1} \right) \right] \right\} \quad (2.22)$$

observe que como se está bajo el supuesto de que $0 < \beta < 1$, tenemos en este caso que $\lim_{x \rightarrow \infty} \left(\lambda - \frac{1}{\alpha^\beta} x^{\beta-1} \right) = \lambda$. Por tanto,

$$\lim_{x \rightarrow \infty} e^{\lambda x} e^{-(\frac{x}{\alpha})^\beta} = \exp \left\{ \lim_{x \rightarrow \infty} x \lambda \right\} = \infty.$$

La otra subclase de la colección de todas las distribuciones de probabilidad con cola pesada que se va a ser referencia es la clase de la distribuciones con cola de variación regular a la cual pertenecen las funciones de variación regular y las funciones de variación lenta.

Definición 12. Se dice que una función $G : (0, \infty) \rightarrow (0, a)$
 $x \mapsto G(x) = l_G(x)x^v$; con
 $a \in \mathbb{R}$ es una función de variación regular si $\lim_{t \rightarrow \infty} G(tx)G(t) = x^v$, y $v \in \mathbb{R}$
donde v se conoce como el índice de variación regular (index of regular variation).
La notación para indicar que G es una función de variación regular con índice v será dada por $G \in RV_v$.

Si la última ecuación es cierta, se podría representar G como

$$G(x) = l_G(x)x^v, \quad \forall x > 0 \quad \text{y} \quad v \in \mathbb{R}$$

donde l_G se conoce como *función de variación lenta* (l_G positiva) y satisface

$$\lim_{t \rightarrow \infty} \frac{l_G(tx)}{l_G(t)} = 1, \quad \forall x > 0.$$

Para el caso de funciones de variación lenta se puede usar la notación $l_G \in RV_0$, ya que con esto se quiere expresar que $v = 0$.

Definición 13. Una variable aleatoria positiva X , se dice que tiene una distribución con cola de variación regular (con $X \in R$), si su función de sobrevivencia $\bar{F}(x)$ puede ser escrita en la forma

$$1 - F(x) = \bar{F}(x) = l_F(x)x^{-\gamma} \quad , \quad x > 0$$

donde $F(\cdot)$ corresponde a la función de distribución de la variable aleatoria y γ (por determinar) es llamado el índice de la cola (tail index) y $l_F(x)$ es una función de variación lenta.

Se puede probar que si $X \in R$, entonces $X \in S$, como se puede verificar en Goldie y Kuppelberg Goldie C. & Kluppelberg C. (1997) y también en Embrechts *et al.* Embrecht P. *et al.* (1997), donde además se encuentran más detalles sobre distribuciones de colas pesadas.

La distribución de Fréchet dada en (1.4), es un ejemplo de distribución con cola pesada de variación lenta y la distribución Weibull (1.5) con parámetro de forma menor que uno, es un ejemplo de distribución subexponencial que no pertenece a la clase de distribuciones con colas de variación regular.

El máximo dominio de atracción de una distribución de valores extremos se puede caracterizar por medio de variación regular y sus extensiones. La clave es que las distribuciones de valores extremos se relacionan con las distribuciones de Pareto generalizadas.

Definición 14 (Índice de Valor Extremo). El parámetro ξ dado en (2.10) es llamado el índice de valor extremo como se muestra en Embrecht P. *et al* (1997).

La parametrización del teorema 2.2 es debido a Von Mises (1936) y Jenkinson (1955). Los diferentes tipos de distribución están relacionados según el valor del parámetro ξ como se ve a continuación:

- a) Para $\xi > 0$, $G(x; \xi) < 1$ para todo x , es decir el punto final superior de la distribución es infinito. Además, cuando $x \rightarrow \infty$, $1 - G(x) \sim \xi^{-\frac{1}{\xi}} x^{-\frac{1}{\xi}}$, es decir la distribución tiene una cola pesada a la derecha. Y los momentos mayores o iguales a $\frac{1}{\xi}$ no existen.
- b) Para $\xi = 0$ el punto final superior de la distribución es infinito. La distribución es además de colas ligeras: $1 - G_0(x) e^{-x}$ cuando $x \rightarrow \infty$, y todos los

momentos existen.

- c) Para $\zeta < 0$ el punto final de la distribución es $-\frac{1}{\zeta}$ así que tiene una cola corta que verifica $1 - G_{\zeta}(-\zeta^{-1} - x) \sim (-\zeta x)^{-\frac{1}{\zeta}}$, cuando $x \rightarrow 0$.

Una parametrización alternativa es la siguiente:

- a) Para $\zeta > 0$ usando $G_{\zeta}\left(\frac{x-1}{\zeta}\right)$ y con $\alpha = \frac{1}{\zeta} > 0$

$$\begin{aligned} G\left(\frac{x-1}{\zeta}, \zeta\right) &= \exp\left(-\left(1 + \zeta \frac{x-1}{\zeta}\right)^{-\frac{1}{\zeta}}\right) \\ &= \exp\left(-\left(1 + x - 1\right)^{-\frac{1}{\zeta}}\right) = \exp\left(-x^{-\alpha}\right) \\ G_2(x) &:= \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0 \end{cases} \end{aligned}$$

Este tipo es llamada la distribución del tipo Frechet (Frechet (1927)).

- b) La función de distribución con $\zeta = 0$

$$G_1(x) = \exp(-e^{-x}),$$

Para todo x real, es llamada la doble exponencial o distribución de Gumbel.

- c) Para $\zeta < 0$ se usa $G_{\zeta}\left(-\frac{1+x}{\zeta}\right)$ y con $\alpha = -\frac{1}{\zeta} > 0$

$$\begin{aligned} G_3(x) &:= \begin{cases} \exp(-(-x)^{\alpha}), & x < 0 \\ 1 & x \geq 0 \end{cases} \\ G\left(-\frac{1+x}{\zeta}, \zeta\right) &= \exp\left(-\left(1 + \zeta \frac{-1-x}{\zeta}\right)^{-\frac{1}{\zeta}}\right) \\ &= \exp\left(-(-x)^{-\frac{1}{\zeta}}\right) = \exp(-(-x)^{\alpha}) \end{aligned}$$

Este tipo es llamada la distribución de Weibull inversa.

Si la relación (2.5) se cumple con $G = G_{\zeta}$ para algún $\zeta \in \mathbb{R}$ se dice que la función de distribución F esta en el dominio de atracción de G_{ζ} y se denota: $F \in D(G_{\zeta})$

El teorema 2.2 conduce a la siguiente reformulación del teorema 2.1 y como en el siguiente teorema se requiere la inversa generalizada se definirá previamente:

Definición 15 (Ínfimo de un conjunto). *Se dice que un conjunto A está acotado inferiormente si y solo si existe un número real c que es menor o igual a cualquier elemento de A . La mayor de las colas inferiores se denomina el ínfimo de A , y se denota $\inf(A)$*

Definición 16 (Inversa Generalizada de una función monótona). *Supongamos que h es una función no decreciente en los reales. La inversa generalizada de h está definida como:*

$$h^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : h(x) \geq t\}$$

Usaremos la convención de que ínfimo de un conjunto vacío es ∞ .

Definición 17 (Función cuantil). *La inversa generalizada de la función de distribución*

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

Es llamada la función cuantil de la función de distribución. El cuantil $x_p = F^{\leftarrow}(p)$ define el cuantil p de F . Entre los cuantiles se tienen los percentiles.

Lema 2.1. *Suponga que f_n es una sucesión de funciones no decrecientes y g es una función no decreciente. Suponga que para cada x en un intervalo abierto (a, b) hay un punto de continuidad de g ,*

$$\lim_{n \rightarrow \infty} f_n(x) = g(x) \tag{2.23}$$

Sean $f_n^{\leftarrow}, g^{\leftarrow}$ la inversa continua a izquierda de f_n y g . Entonces, para cada x en el intervalo $(g(a), g(b))$ es un punto de continuidad de g^{\leftarrow} tenemos que:

$$\lim_{n \rightarrow \infty} f_n^{\leftarrow}(x) = g^{\leftarrow}(x) \tag{2.24}$$

Demostración. Sea x un punto de continuidad de g^{\leftarrow} . Para $\varepsilon > 0$, Debemos demostrar que para $n, n_0 \in \mathbb{N}, n \geq n_0$,

$$f_n^{\leftarrow}(x) - \varepsilon \leq g^{\leftarrow}(x) \leq f_n^{\leftarrow}(x) + \varepsilon \tag{2.25}$$

Vamos a probar la desigualdad de la derecha; La prueba de la demostración del lado izquierdo es similar.

Se escoge $0 < \varepsilon_1 < \varepsilon$ tal que $g^{\leftarrow}(x) - \varepsilon_1$ es un punto de continuidad de g . Esto es posible ya que los puntos de continuidad de g forman un conjunto denso. Como g^{\leftarrow} es continua en x , $g^{\leftarrow}(x)$ es un punto de crecimiento de g ; por lo tanto $g(g^{\leftarrow}(x) - \varepsilon_1) < 1$. Se toma $\delta < x - g(g^{\leftarrow}(x) - \varepsilon_1)$. Dado que $g^{\leftarrow}(x) - \varepsilon_1$ es un punto de continuidad de g , existe n_0 tal que $f_n(g^{\leftarrow}(x) - \varepsilon_1) < g(g^{\leftarrow}(x) - \varepsilon_1) + \delta < x$ para $n \geq n_0$. De la definición 4 se tiene que $g^{\leftarrow}(x) - \varepsilon_1 \leq f_n^{\leftarrow}(x)$. \square

Teorema 2.3 (Reformulaciones de las equivalencias del teorema 2.1). Para todo $\zeta \in \mathbb{R}$ las siguientes afirmaciones son equivalentes:

1. Existen constantes reales $a_n > 0$ y b_n , tal que

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_{\zeta}(x) = \exp\left(- (1 + \zeta x)^{-\frac{1}{\zeta}}\right) \quad (2.26)$$

Para todo x con $1 + \zeta x > 0$

2. Existe una función positiva $D_{\zeta}(x)$ tal que para todo $x > 0$

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D_{\zeta}(x) = \frac{x^{\zeta} - 1}{\zeta} \quad (2.27)$$

Donde para $\zeta = 0$ el lado derecho es interpretado como $\log x$

3. Existe una función positiva F tal que

$$\lim_{t \rightarrow \infty} t(1 - F(a(t)x + U(t))) = (1 + \zeta x)^{-\frac{1}{\zeta}} \quad (2.28)$$

para todo x con $1 + \zeta x > 0$

4. Existe una función positiva f tal que

$$\lim_{t \rightarrow x'} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \zeta x)^{-\frac{1}{\zeta}} \quad (2.29)$$

para todo x para el cual $1 + \zeta x > 0$, donde $x^* = \sup\{x : F(x) < 1\}$

Demostración. La equivalencia de 2 y 3 se demuestra en el lema 2.1 (ver De Haan, L. y Ferreira, A. 2006) y se aplica este mismo lema a la relación (2.7). Sea la función U la inversa continua a la izquierda de $\frac{1}{1-F}$. Se debe tener en cuenta que $U(t)$ está definida para $t > 1$. Se sigue (2.7) es equivalente a

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

para cada x positivo. que es más simple que (2.7).

Así que es suficiente probar que la relación anterior implica 3. Sea x un punto de continuidad de

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

Para $t \geq 1$,

$$\frac{U(\llbracket t \rrbracket x - b_{\llbracket t \rrbracket})}{a_{\llbracket t \rrbracket}} \leq \frac{U(tx) - b_{\llbracket t \rrbracket}}{a_{\llbracket t \rrbracket}} \leq \frac{U\left(\llbracket t \rrbracket x \left(1 + \frac{1}{\llbracket t \rrbracket}\right)\right) - b_{\llbracket t \rrbracket}}{a_{\llbracket t \rrbracket}}$$

el lado derecho es eventualmente menor $D(x')$ para cualquier punto continuo $x' > x$ con $D(x') > D(x)$. Ya que D es continuo en x se tiene

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

esto es 3 del teorema (2.8).

Se probará ahora que 2 implica 4.

Para $\varepsilon > 0$, aplicando g a ambos lados de (2.25) se tiene

$$g(h^{\leftarrow}(t) - \varepsilon) \leq t \leq g(h^{\leftarrow}(t) + \varepsilon),$$

donde g es una función no decreciente y h^{\leftarrow} es la inversa continua a la derecha. Se sigue que

$$\begin{aligned} \frac{(1 - \varepsilon)^{\xi} - 1}{\xi} &\leftarrow \frac{U\left(\frac{1 - \varepsilon}{1 - F(t)}\right) - U\left(\frac{1}{1 - F(t)}\right)}{a\left(\frac{1}{1 - F(t)}\right)} < \frac{t - U\left(\frac{1}{1 - F(t)}\right)}{a\left(\frac{1}{1 - F(t)}\right)} \\ &< \frac{U\left(\frac{1 + \varepsilon}{1 - F(t)}\right) - U\left(\frac{1}{1 - F(t)}\right)}{a\left(\frac{1}{1 - F(t)}\right)} \rightarrow \frac{(1 + \varepsilon)^{\xi}}{\xi} \end{aligned}$$

$t \rightarrow x^*$ Luego,

$$\lim_{t \rightarrow x^*} \frac{t - U\left(\frac{1}{1 - F(t)}\right)}{a\left(\frac{1}{1 - F(t)}\right)} = 0$$

Por lo tanto por 2 para todo $x > 0$

$$\lim_{t \rightarrow x^*} \frac{t - U\left(\frac{1}{1 - F(t)}\right)}{a\left(\frac{1}{1 - F(t)}\right)} = \frac{x^{\xi} - 1}{\xi}$$

y por el lema 1.1.1 (Texto Extreme Value Theory An Introduction, de Haan y Ferreira)

$$\lim_{t \rightarrow x^*} \frac{1 - F(t)}{1 - F\left(t + xa \left(\frac{1}{1 - F(t)}\right)\right)} = (1 + \zeta x)^{\frac{1}{\zeta}}$$

Es decir 4 se cumple. Lo mismo 4 implica 2. El rango completo de posibles distribuciones límite para M_n^* esta dado por el teorema de los tipos de Extremos (ver teorema 2.2).

Definición 18 (Distribución Max-estable). Sea X_1, X_2, \dots variables aleatorias independientes e idénticamente distribuidas con una función de distribución F . La función de distribución F es llamada max-estable si para alguna selección de constantes $a_n > 0$, y b_n real

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = P(X_1 \leq x)$$

Para toda x y $n = 1, 2, \dots$. La clase de distribuciones max estables o estable en los máximos es la misma clase de distribuciones de valor extremo.

Una distribución para la cual el máximo muestral re-escalado, posee la misma distribución que la variable aleatoria inicial.

Teorema 2.4 (Distribución Max-Estable). Una distribución es max-estable si y solo si es una de las distribuciones de valores extremos.

Demostración.

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x) \Leftrightarrow \lim_{n \rightarrow \infty} P\left(\frac{M_{nk} - b_{nk}}{a_{nk}} \leq x\right) = G(x) \quad (2.30)$$

$$\Leftrightarrow P\left(\frac{M_{nk} - b_{nk}}{a_{nk}} \leq x\right) = \left[P\left(\frac{M_n - b_n}{a_n} \leq x\right)\right]^k \quad (2.31)$$

En consecuencia de (2.30) y (2.31)

$$\begin{aligned} \Leftrightarrow P(M_{nk} \leq x) &\approx G\left(\frac{x - b_{nk}}{a_{nk}}\right) \\ \Leftrightarrow P(M_{nk} \leq x) &\approx G^k\left(\frac{x - b_{nk}}{a_{nk}}\right) \end{aligned}$$

Por lo tanto

$$\Leftrightarrow G(x) = G^k\left(\tilde{a}_k x + \tilde{b}_k\right)$$

2.1.3 Distribución Generalizada de Valores Extremos

Von Mises (1954) y Jenkinson (1955) derivaron cada uno por su lado, la distribución generalizada de extremos que unificó las tres distribuciones límites de valores extremos establecidas en el teorema 2.2.

Estas tres distribuciones límites tienen comportamientos diferentes que están relacionados con las distintas formas que pueda tener la cola de la distribución F de los X_i , y esto se puede precisar más considerando el comportamiento de la distribución límite G en el punto final superior x^* . Así para la distribución Weibull, x^* es finito, mientras que la distribución Frechet y Gumbel es infinito, $x^* = \infty^+$.

Sin embargo, la función de densidad para G decae exponencialmente para el caso de la distribución de Gumbel, y en forma polinomial para la distribución Frechet.

En las primeras aplicaciones de la teoría de valores extremos, era usual adoptar una de estas tres familias y luego estimar los parámetros relevantes de la distribución escogida. Pero esta metodología tiene dos debilidades, primero se necesita una técnica adecuada para escoger cual familia de distribución es la más apropiada para manejar los datos; segundo, una vez tomada la decisión, la inferencia asume que no hay incertidumbre en la selección de la familia de distribución más apropiada.

Los problemas anteriores se resuelven al combinar en una sola distribución las tres distribuciones de valores extremos conocida como la distribución generalizada de valores extremos (GEV) dada por la siguiente expresión:

$$G(x, \mu, \xi, \sigma) = \exp \left\{ - \left[\xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \quad (2.32)$$

Definida sobre el conjunto

$$\left\{ x : 1 + \xi \left(\frac{x - \mu}{\sigma} \right) > 0 \right\}$$

Porque de lo contrario, G toma el valor de 0 o 1 y entonces es una distribución degenerada, este hecho puede ser visto en Salvadori G., De Michele C. y Otros. Como se puede ver en Von Mises (1954), los parámetros satisfacen las condiciones

- $-\infty < \mu < \infty$

- $\sigma > 0$
- $-\infty < \xi < \infty$

Donde μ es el parámetro de localización, σ es el parámetro de escala, y ξ es el parámetro de forma, que es el más importante porque determina el comportamiento de la cola de distribución de los datos X_i .

1. Si $\xi > 0$ la cola de la distribución tiene un comportamiento tipo Frechet con $\alpha = \frac{1}{\xi}$
2. Si $\xi < 0$ la cola de la distribución tiene un comportamiento del tipo Weibull con $\alpha = -\frac{1}{\xi}$
3. Si $\xi \rightarrow 0$ la cola de la distribución tiene un comportamiento del tipo de Gumbel, con parámetros de localización y escala, arbitrarios.

La unificación de las tres familias de distribuciones de valores extremos en una sola, implica una gran simplificación en las implementaciones estadísticas; además, la inferencia de ξ permite que sean los mismos datos que determinen cual es el modelo más apropiado para el comportamiento de la cola de la distribución y así no hay necesidad de introducir en la inferencia incertidumbre debido a la escogencia del modelo más adecuado.

El siguiente teorema unifica en una sola distribución las tres distribuciones anteriores.

Teorema 2.5 (Distribución unificada de Valor Extremo²). Sean $\{a_n\}$ y $\{b_n\}$ sucesiones de números reales, $M_n = \{X_1, X_2, \dots, X_n\}$ y si

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{n \rightarrow \infty} G(x) \Leftrightarrow \lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x) \quad (2.33)$$

Para alguna distribución no degenerada G entonces G es un miembro de la familia GEV con

$$G(x, \sigma, \xi, \mu) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \quad (2.34)$$

Definida en el conjunto

$$\left\{x : 1 + \xi\left(\frac{x - \mu}{\sigma}\right) > 0\right\} \quad (2.35)$$

²Para un algoritmo para la estimación de los parámetros de la distribución de valor extremo generalizada ver el artículo titulado Algorithm AS 215: Maximum-Likelihood of the parameters of Generalized Extreme-Value Distribution escrito por J. R. M. Hosking.

donde $-\infty < \mu < \infty$, $\sigma > 0$ y $-\infty < \xi < \infty$

Las siguientes son propiedades importantes de la distribución generalizada de valores extremos.

a. La media existe si $\xi < 1$ y esta dada por:

$$\mu_1 = E(X) = \mu + \frac{\sigma}{\xi}[\sigma(1 - \xi) - 1] \quad (2.36)$$

b. La varianza existe si $\xi < \frac{1}{2}$ y está dada por:

$$\mu_2 = E(X - \mu)^2 = \frac{\sigma^2}{\xi^2}[\Gamma(1 - 2\xi) - \Gamma(1 - \xi)] \quad (2.37)$$

Para el caso limite $\xi \rightarrow 0$, estas ecuaciones se reducen a

$$\lim_{\xi \rightarrow 0} \mu_1 = \lim_{\xi \rightarrow 0} \left\{ \mu + \frac{\sigma}{\xi}[\sigma(1 - \xi) - 1] \right\} = \mu + \sigma\gamma \quad (2.38)$$

$$\lim_{\xi \rightarrow 0} \mu_2 = \lim_{\xi \rightarrow 0} \frac{\sigma^2}{\xi^2} [\Gamma(1 - 2\xi) - \Gamma(1 - \xi)] = \frac{\sigma^2 \pi^2}{6} \quad (2.39)$$

En donde $\gamma = 0.5772$ es la constante de Euler.

Otra característica importante a destacar de la distribución generalizada de extremos, es la estimación de los cuantiles extremos los cuales se pueden obtener invirtiendo la ecuación (2.32):

Si $\xi \neq 0$, como la función de distribución acumulada de la Generalizada de Pareto es:

$$G(x, \xi, \mu, \sigma) = \exp \left\{ - \left[\xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}$$

Pero para hallar el percentil $1 - p$ igualamos:

$$\begin{aligned} 1 - p &= \exp \left(- \left[\xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right) \Rightarrow -\log(1 - p) = \left[\xi \frac{x - \mu}{\sigma} \right]^{-\frac{1}{\xi}} \\ &\Rightarrow \{-\log(1 - p)\}^{-\xi} = \xi \frac{x - \mu}{\sigma} \\ &\Rightarrow x_p = \mu + \frac{\sigma}{\xi} \{-\log(1 - p)\}^{-\xi} \end{aligned}$$

Si $\xi = 0$ la función de distribución acumulada es:

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

Pero para hallar el percentil $1 - p$ igualamos:

$$\begin{aligned} 1 - p &= \exp\left(-e^{\frac{x-\mu}{\sigma}}\right) \Leftrightarrow \log(1 - p) = -e^{\frac{x-\mu}{\sigma}} \Leftrightarrow -\log(1 - p) = e^{\frac{x-\mu}{\sigma}} \\ \log\{-\log(1 - p)\} &= \frac{x - \mu}{\sigma} \Rightarrow \sigma \log\{-\log(1 - p)\} = x - \mu \\ &\Rightarrow x_p = \mu + \sigma \log\{-\log(1 - p)\} \end{aligned}$$

Si definimos $x_p = -\log(1 - p)$ entonces

$$x_p = \begin{cases} \mu + \frac{\sigma}{\xi} \left[1 - \{-\log(1 - p)\}^{-\xi}\right] & \text{para } \xi \neq 0 \\ \mu + \sigma \log\{-\log(1 - p)\} & \text{para } \xi = 0 \end{cases} \quad (2.40)$$

Un concepto que se requerirá en seguida es la convergencia débil el cual pasamos a definir:

Definición 19 (Convergencia en distribución o convergencia débil). *Se dice que $\{X_n\}$ converge en distribución o converge débilmente a la variable aleatoria X si para toda función f continua y acotada $\lim_{n \rightarrow \infty} E(f(X_n)) = E(f(X))$ lo cual se anota así $X_n \xrightarrow{d} X$*

2.1.4 Dominios de Atracción y Convergencia Débil

El interés principal de los investigadores en la teoría de valores extremos, es tratar de relacionar la función de distribución G con la función de distribución F , es decir dado F , ¿Podemos encontrar G ? La respuesta a esta pregunta está directamente relacionada con el concepto de dominio de atracción.

En este sentido, se dice que una función de distribución F , discreta o absolutamente continua, pertenece al dominio de máxima atracción de una función de distribución G no degenerada, si existen sucesiones $\{a_n\}$, $a_n > 0$ y $\{b_n\}$ tales que

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

Para todos los x , puntos de continuidad de G ; y se denota por $F \in D(G)$.

El dominio de atracción caracteriza un conjunto de distribuciones para las cuales sus máximos normalizados M_n^* , tienen distribución límite G , donde G pertenece a la familia de las distribuciones generalizadas de extremos.

Para poder caracterizar los dominios de atracción, es necesario establecer un conjunto de condiciones necesarias y suficientes sobre F que garanticen que para

$i = 1, 2, 3$; o en otras palabras, que condiciones debe cumplir F que garanticen la existencia de sucesiones $\{a_n\}$, $a_n > 0$ y $\{b_n\}$ tales que

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_i(x) \quad , \quad i = 1, 2, 3 \quad (2.41)$$

Este tipo de convergencia se llama convergencia débil de funciones de distribuciones o de variables aleatorias.

Definición 20 (Máximo dominio de atracción (MDA)). *Se dice que la distribución F pertenece al máximo dominio de atracción de la distribución G si existen secuencias de constantes $\{a_n\}$ y $\{b_n > 0\}$, tal que se cumple,*

$$G(x) = \lim_{n \rightarrow \infty} F^n(a_n + b_n x), \quad x \in \mathbb{R}, \quad (2.42)$$

Corolario 2.1 (Base para pruebas de condiciones necesarias). Sean $\{a_n\}$ y $\{b_n\}$ sucesiones de números reales tales que, si para $\forall x$

$$\lim_{n \rightarrow \infty} n(1 - F(a_n x + b_n)) = \mu(x)$$

entonces

$$\lim_{n \rightarrow \infty} P(M_n < a_n x + b_n) = \exp(-\mu(x))$$

A partir del corolario 2.1 se puede establecer los teoremas que determinan que condiciones debe cumplir F para que existan sucesiones normalizadoras $\{a_n\}$, $a_n > 0$ y $\{b_n\}$ que garanticen que M_n^* converge débilmente a una función de distribución no degenerada G .

En los teoremas siguientes $*x$ es el punto final inferior de la distribución F , definida como

$$*x = \alpha(F) = \inf\{x : F(x) > 0\} \quad (2.43)$$

y x^* es el punto final superior de la distribución F donde

$$x^* = W(F) = \sup\{x : F(x) < 1\} \quad (2.44)$$

Teorema 2.6 (Condición necesaria Distribución de Frechet). Sea $w(F) = \infty^+$ y $a \in \mathbb{R}$, $\alpha > 0$ tal que, si $\forall x > 0$

$$\lim_{n \rightarrow \infty^+} \frac{1 - F(x)}{1 - F(t)} = x^{-\alpha} \quad (2.45)$$

entonces existe una sucesión $\{a_n\}$ tal que

$$\lim_{n \rightarrow \infty} P(M_n < a_n x) = G_2(x, \alpha) \quad (2.46)$$

la sucesión normalizada $\{a_n\}$ se puede escoger como

$$a_n = \inf \left\{ x : 1 - F(x) \leq \frac{1}{n} \right\} \quad (2.47)$$

Demostración. Para nuestro propósito se escoge $b_n = 0$ y a_n como aparece en la ecuación (2.47). El supuesto de que $w(F) = \infty^+$ implica que $\{a_n\}$ tiende a ∞^+ , por lo tanto

1. Para $x < 0$ tenemos que

$$a_n x \xrightarrow{n \rightarrow \infty^+} \infty^- \Rightarrow 1 - F(a_n x) \xrightarrow{n \rightarrow \infty^+} 1 \Rightarrow \lim_{n \rightarrow \infty^+} n(1 - F(a_n x)) = \infty^+$$

Y por el corolario 2.1 se cumple

$$\lim_{n \rightarrow \infty^+} P(M_n \leq a_n x) = e^{-\infty^+} = 0$$

En consecuencia, el teorema 2.6

2. Para $x > 0$ aplicamos la hipótesis (2.45) con $t = a_n$, entonces obtenemos para $a_n \rightarrow \infty^+$

$$\lim_{n \rightarrow \infty} \frac{1 - F(a_n)}{1 - F(a_n)} = x^{-\alpha} \quad (2.48)$$

Para $\alpha > 0$ y $x > 0$, en consecuencia

$$\begin{aligned} \lim_{n \rightarrow \infty^+} n(1 - F(a_n x)) &= \lim_{n \rightarrow \infty^+} n \frac{(1 - F(a_n))(1 - F(a_n x))}{1 - F(a_n)} \\ &= \lim_{n \rightarrow \infty^+} n(1 - F(a_n))x^{-\alpha} \quad \text{de acuerdo a (2.48)} \\ &= x^{-\alpha} \lim_{n \rightarrow \infty^+} n(1 - F(a_n)) \end{aligned} \quad (2.49)$$

Ahora en vista de la definición de a_n en (2.47), se tiene que

$$1 - F(a_n + 0) \leq \frac{1}{n} \leq 1 - F(a_n)$$

O equivalentemente

$$1 \leq n(1 - F(a_n)) \leq \frac{1 - F(a_n)}{1 - F(a_n + 0)} \quad (2.50)$$

y como $1 - F$ decrece, entonces, $1 - F(a_n + 0) \geq 1 - F(a_n x)$ para $x > 1$, aplicando esta última desigualdad en (2.50) y usando de nuevo (2.45) obtenemos

$$1 \leq n(1 - F(a_n)) \leq \frac{1 - F(a_n)}{1 - F(b_n + 0)} \leq (1 - \varepsilon)x^\alpha \quad (2.51)$$

En donde $\varepsilon \rightarrow 0$ cuando $n \rightarrow \infty^+$. Como $x > 1$ es arbitrario, (2.51) implica

$$\lim_{n \rightarrow \infty^+} n(1 - F(a_n)) = 1$$

En consecuencia, en (2.49) se tiene

$$\lim_{n \rightarrow \infty^+} n(1 - F(a_n x)) = x^{-\alpha} \lim_{n \rightarrow \infty^+} n(1 - F(a_n)) = x^{-\alpha}$$

Luego por el corolario 2.1

$$\lim_{n \rightarrow \infty^+} P(M_n \leq x) = \begin{cases} e^{-x^{-\alpha}} & \text{si } x > 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Teorema 2.7 (Condición necesaria Distribución Weibull). Sea $w(F)$ finito y supongamos que la función de distribución

$$F^*(x) = F(wF(x) - 1) \quad \text{con } x > 0$$

satisface la ecuación (2.49). Entonces, existen sucesiones $\{a_n\}$ y $\{b_n\}$ tal que

$$\lim_{n \rightarrow \infty} P(M_n < a_n x + b_n) = G_3(x, \alpha)$$

Las sucesiones normalizadoras, $\{a_n\}$ y $\{b_n\}$, se pueden escoger como

$$b_n = w(F)$$

$$a_n = w(F) - \inf \left\{ x : 1 - F(x) \leq \frac{1}{n} \right\}$$

Demostración. Como $w(F^*) = \infty^+$ y $F^*(x) = (w(F(x)) - 1)$ con $w(F)$ finito, entonces como (2.49) se satisface para F^* , la conclusión del teorema 2.6 establece que para cualquier $x > 0$

$$\begin{aligned} \lim_{n \rightarrow \infty^+} (F^*)^n(a_n^* x) &= \lim_{n \rightarrow \infty^*} F^n \left(w(F) - \frac{1}{a_n^* x} \right) \\ &= G_2(x; \alpha) \end{aligned} \tag{2.52}$$

Donde

$$\begin{aligned} a_n^* x &= \inf \left\{ x : 1 - F^*(x) \leq \frac{1}{n} \right\} = \inf \left\{ x : 1 - F \left(w(F) - \frac{1}{x} \right) \leq \frac{1}{n} \right\} \\ &= \left(w(F) - \inf \left\{ x : 1 - F(x) \leq \frac{1}{n} \right\} \right)^{-1} \end{aligned}$$

En consecuencia, cuando se escoge $b_n = w(F)$ y $a_n = \frac{1}{a_n^*}$, la ecuación (2.52) se transforma en

$$\begin{aligned} \lim_{n \rightarrow \infty^*} F\left(b_n - \frac{a_n}{x}\right) &= G_2(x; \alpha) \quad \text{si } x > 0 \\ \lim_{n \rightarrow \infty^*} F(b_n - a_n x) &= G_2\left(-\frac{1}{x}; \alpha\right) \quad \text{si } x < 0 \end{aligned} \quad (2.53)$$

pero para $x < 0$

$$G_2\left(-\frac{1}{x}; \alpha\right) = \exp(-(-x)^\alpha) = G_3(x; \alpha)$$

Para completar la prueba, se debe notar que como $b_n = w(F)$ y $a_n > 0$

Teorema 2.8 (Condición necesaria Distribución Gumbel). Se asume que para algún γ finito, se cumple que

$$\int_{\gamma}^{w(F)} (1 - F(x)) dx < \infty^+$$

y para $\alpha(F) < t < w(F)$, definamos

$$R(t) = (1 - F(t))^{-1} \int_t^{w(F)} (1 - F(x)) dx \quad (2.54)$$

si para $\forall x \in R$

$$\lim_{t \rightarrow w(F)} \frac{1 - F(t + xR(t))}{1 - F(t)} = \exp(-x) \quad (2.55)$$

entonces existen sucesiones $\{a_n\}$ y $\{b_n\}$ tal que

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

Las constantes normalizadoras se pueden escoger como

$$b_n = \inf \left\{ x : 1 - F(x) \leq \frac{1}{n} \right\} \quad (2.56)$$

$$a_n = R(b_n) \quad (2.57)$$

Demostración. Para la prueba se escogerá $\{b_n\}$ y $\{a_n\}$ como lo indica (2.56) y (2.57) respectivamente. Se observa que $\lim_{n \rightarrow \infty} b_n = w(F)$. Luego para $t = b_n$ y $R(t) = R(b_n) = a_n$ según (2.57), y siguiendo (2.55), tenemos:

$$\lim_{n \rightarrow \infty^+} \frac{1 - F(b_n + a_n x)}{1 - F(b_n)} = \exp(-x) \quad \forall x \quad (2.58)$$

Por lo tanto para un x arbitrario

$$\begin{aligned} \lim_{n \rightarrow \infty^+} n(1 - F(a_n x + b_n)) &= \lim_{n \rightarrow \infty^+} n \frac{(1 - F(b_n))(1 - F(a_n x + b_n))}{1 - F(b_n)} \\ &= \exp(-x) \lim_{n \rightarrow \infty^+} n(1 - F(b_n)) \end{aligned} \quad (2.59)$$

pero por (2.56)

$$1 - F(b_n + 0) \leq \frac{1}{n} \leq 1 - F(b_n)$$

Por otra parte, para cualquier $\varepsilon > 0$

$$1 - F(b_n + \varepsilon a_n) \leq 1 - F(b_n + 0)$$

Estas dos últimas dos desigualdades se pueden combinar y producir el siguiente resultado

$$1 \leq n(1 - F(b_n)) \leq \frac{1 - F(b_n)}{1 - F(b_n + \varepsilon a_n)} \xrightarrow{n \rightarrow \infty^+} \exp(\varepsilon)$$

Usando el límite (2.58). Pero como $\varepsilon > 0$ es arbitrario, entonces

$$\lim_{n \rightarrow \infty} n(1 - F(b_n)) = 1$$

En consecuencia, volviendo al límite en (2.59)

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

Luego usando el corolario (2.1) se tiene que

$$\lim_{n \rightarrow \infty^+} P(M_n \leq b_n + a_n x) = \exp(-\exp(-x)) \quad \text{para } -\infty < x < \infty$$

Así que bajo las condiciones de uno de estos teoremas, la función de distribución F pertenece al dominio de atracción de la función de distribución Frechet o de la función de distribución Weibull o de la función de distribución Gumbel respectivamente. Además, estos resultados destacan el hecho de que la escogencia de las constantes normalizadoras no son únicas. También es importante destacar que las condiciones necesarias y suficientes dadas por los teoremas 2.6, 2.7 y 2.8 son equivalentes a las condiciones originales usadas por Gnedenko (1943). Sin embargo las condiciones necesarias y suficientes de estos teoremas probados son por lo general muy difíciles de verificar. Por esta razón, se utiliza un resultado alternativo que es el teorema de Von Mises (1936), cuyas condiciones de suficiencia son más fáciles de probar y no son tan restrictivas, de hecho son aplicables solo a funciones de distribución F absolutamente continuas. A continuación se derivan

condiciones suficientes de la función de distribución F para asegurar que hay una sucesión de constantes $a_n > 0$ y b_n tales que $\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\xi(x)$ para algún ξ y todo x . Estas condiciones requieren la existencia de la primera y segunda derivada de F .

Todo teorema de condición suficiente tiene su corolario adicional en donde esta condición suficiente se expresa en términos de la función inversa U en vez de la función de distribución F , que es una manera más sencilla de expresarla.

Teorema 2.9 (Condición suficiente). Sea F una función de distribución y x^* su extremo derecho. Supongamos que $F''(x)$ existe y $F'(x) > 0$ para todo x en alguna vecindad a la izquierda de x^* . Si

$$\lim_{t \rightarrow x^*} \left(\frac{1-F}{F'} \right)'(t) = \xi \quad (2.60)$$

o equivalentemente

$$\lim_{t \rightarrow x^*} \frac{(1-F(t))F''(t)}{(F'(t))^2} = -\xi - 1 \quad (2.61)$$

Entonces F esta en el dominio de atracción de G_ξ .

Demostración. Consideremos $\frac{1}{1-F(U(t))} = t$ por lo tanto

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

Luego

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

de modo de

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

La relación probada es equivalente a

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\xi - 1}{\xi} \quad (2.62)$$

Para todo $x > 0$. Así que se necesita demostrar que

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

Implica (2.62) para la misma ξ , de modo que $1 < x_0 < x$

$$\log U'(x) - \log U'(x_0) = \int_{x_0}^x \frac{U''(s)}{U'(s)} ds$$

Se tiene que para $x > 0$, $tx > 1$

$$\log U'(tx) - \log U'(t) = \int_1^x A(ts) \frac{ds}{s} \quad \text{con} \quad A(t) := \frac{tU''(t)}{U'(t)}.$$

Se sigue que para $0 < a < b < \infty$,

$$\limsup_{t \rightarrow \infty} \left| \log \frac{U'(tx)}{U'(t)} - \log x^{\xi-1} \right| = 0$$

Por lo tanto también, ya que $|e^s - e^t| < c|s - t|$ en un intervalo compacto para una constante positiva c ,

$$\limsup_{t \rightarrow \infty} \left| \frac{U'(tx)}{U'^t} - x^{\xi-1} \right| = 0$$

Esto implica que

$$\frac{U(tx) - U(t)}{tU(t)} - \frac{x^{\xi} - 1}{\xi} = \int_1^x \left(\frac{U'(ts)}{U'(t)} - s^{\xi-1} \right) ds$$

Converge a cero. Por lo tanto por la parte b del teorema 2.3 se concluye que F pertenece al dominio de atracción de $G_{\xi}(F \in D(G_{\xi}))$.

Corolario 2.2. La condición (2.60) es equivalente a:

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

Lo cual implica

$$\lim_{t \rightarrow \infty} \frac{U'(tx)}{U'(t)} = x^{\xi-1}$$

Localmente uniforme en $(0, \infty)$ y finalmente

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{tU'(t)} = \frac{x^{\xi} - 1}{\xi}$$

Así que por el teorema 2.2 se obtiene $F \in D(G_{\xi})$. Condiciones más simples son posibles para $\xi \neq 0$

Cuando $\xi > 0$ se está en el dominio de atracción de la distribución límite de Frechet. Algunas funciones que caen en esta clase son: Distribución Burr, Distribución Generalizada de Pareto, la distribución log-gamma, la distribución de Cauchy, la distribución F y la t .

Teorema 2.10 (Condición suficiente. Distribución de Frechet). ($\xi > 0$) Si se supone $x^* = \infty$ y F' Si

$$\lim_{t \rightarrow \infty} \frac{tF'(t)}{1 - F(t)} = \frac{1}{\xi} \quad (2.63)$$

para algún $\xi > 0$, entonces F esta en el dominio de atracción de G_{ξ} .

Demostración. Como en la prueba del teorema 2.9 se ve que la condición (2.63) es equivalente a

$$\lim_{t \rightarrow \infty} \frac{tU'(t)}{U(t)} = \xi \quad (2.64)$$

Además

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

por lo tanto (2.64) para $x > 0$

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^{\xi} \quad (2.65)$$

o

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

El cuál es la condición (2.26) del teorema 2.3.

Corolario 2.3. ($\xi > 0$) Condición (2.63) es equivalente a,

$$\lim_{t \rightarrow \infty} \frac{tU'(t)}{U(t)} = \xi$$

Que implica (2.61) en vista (2.65).

Cuando $\xi < 0$ se está en el dominio de atracción de la distribución límite de Weibull. Algunas distribuciones que caen en esta clase son: Uniforme, Beta, la Weibull de valor Extremo.

Teorema 2.11 (Condición suficiente. Distribución Weibull). ($\xi < 0$) Se supone $x^* < \infty$ y $F'(x)$ existe para $x < x^*$. Si

$$\lim_{t \rightarrow \infty} \frac{(x^* - t)F'(t)}{1 - F(t)} = -\frac{1}{\xi} \quad (2.66)$$

para algún ξ negativo, entonces F es el dominio de atracción de G_{ξ}

Demostración. Como antes, se puede ver que la condición (2.66) es equivalente a

$$\lim_{t \rightarrow \infty} \frac{tU'(t)}{U(\infty) - U(t)} = -\zeta \tag{2.67}$$

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

La relación (2.67) implica

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

esto es para $x > 0$

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

Corolario 2.4. ($\zeta < 0$) la condición (2.66) es equivalente a

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

que implica (2.67)

Las distribuciones de valores extremos se obtienen como distribuciones límite de los valores más grandes o más pequeños de una muestra aleatoria de tamaño creciente. Para obtener una distribución límite no degenerada es necesario reducir el valor más grande real aplicando una transformación lineal con coeficientes que dependen del tamaño de la muestra. Este proceso es análogo a la estandarización aunque no se restringe para esta sucesión particular de transformaciones lineales. Si hay una distribución límite de interés se debe encontrar como la distribución límite de alguna sucesión de valores reducidos transformados, donde a_n, b_n pueden depender de n pero no de x .

Para distinguir el límite de la distribución acumulada del valor más grande reducido de $F(x)$ será denotado por $G(x)$ esta debe satisfacer la ecuación

$$|G(x)|^n = G(a_n x + b_n)$$

Esta ecuación fue obtenida por Frechet (1927) y también por Fisher y Tippett (1928). Es algunas veces llamado *el postulado de estabilidad*.

El teorema siguiente presenta los valores de a_n y b_n para las tres familias de distribuciones límite de valores extremos: la distribución tipo Gumbel, la tipo Frechet y la tipo Weibull.

Teorema 2.12 (Constantes Normalizadas).

Hipótesis:

- a) Si $G = G_2(x; \alpha) = \begin{cases} \exp(-x^{-\alpha}) & x > 0 \\ 0 & x \leq 0 \end{cases}$
- b) Si $G = G_3(x; \alpha) = \begin{cases} \exp(-(-x^\alpha)) & x < 0 \\ 0 & x \geq 0 \end{cases}$
- c) Si $G = G_1(x; \alpha) = \exp(-e^{-x})$ para todo $x \in R$

Tesis: Se escoge

- a) $b_n = 0$ y $a_n = F^{-1}\left(1 - \frac{1}{n}\right)$ para la distribución tipo 2 (Distribución tipo Frechet).
- b) $b_n = F^{-1}(1)$ y $a_n = F^{-1}(1) - F^{-1}\left(1 - \frac{1}{n}\right)$ para una distribución tipo 3 (Distribución tipo weibull).
- c) $a_n = F^{-1}\left(1 - \frac{1}{n}\right) - F^{-1}\left(1 - \frac{1}{n^2}\right)$, $b_n = F^{-1}\left(1 - \frac{1}{n}\right)$ Para una distribución tipo 1 (Distribución tipo Gumbel).

Demostración.

- a) Distribución tipo 1 (Distribución tipo Gumbel)

$$[G(x)]^n = G(a_n x + b_n) = G(x + b_n), \text{ para un } a_n = 1$$

$$\begin{aligned} x = a_n x + b_n &\Rightarrow x - a_n x = b_n \Rightarrow (1 - a_n)x = b_n \\ &\Rightarrow x = \frac{b_n}{1 - a_n} \Rightarrow x = b_n(1 - a_n)^{-1} \end{aligned}$$

Por lo tanto, $G(x) = G(b_n(1 - a_n)^{-1})$

$$[G(x)]^{nm} = [G(x + b_n)]^m = G(x + b_n + b_m) = G(x + b_{nm})$$

por lo tanto, $b_n + b_m = b_{nm}$

$$[G(x)]^n = G(x + b_n)$$

$$n \log G(x) = \log G(x + b_n)$$

$$\log n + \log [-\log G(x)] = \log [-\log(x + \sigma \log n)]$$

$$\begin{aligned}
 h(x) &= \log [-\log G(x)] \\
 h(x) &= h(0) \frac{x}{\sigma} \Rightarrow h(x) = \frac{\sigma h(0) - x}{\sigma} \\
 -\frac{x - \sigma h(0)}{\sigma} &= -\log \{-\log G(x)\} \\
 e^{-\frac{x - \sigma h(0)}{\sigma}} &= -\log G(x) \Rightarrow -e^{-\frac{x - \sigma h(0)}{\sigma}} = \log G(x) \\
 G(x) &= \exp \left\{ -e^{-\frac{x - \mu}{\sigma}} \right\}
 \end{aligned}$$

dado que $\mu = \sigma \log(-\log G(0))$

Damos a continuación un ejemplo que dará a entender la diferencia entre el método de bloques para representar el comportamiento extremo de un fenómeno y el método de picos sobre el umbral el cual se desarrollará en la siguiente sección.

Ejemplo 2.2. *En general, los datos geofísicos son recolectados a través de un diario o sub-diario (por ejemplo, a una hora, un minuto, . . .) la resolución temporal, por ejemplo, consideremos una muestra de máximo (o mínimo, o promedio) las temperaturas diarias, o una muestra de la precipitación total diaria.*

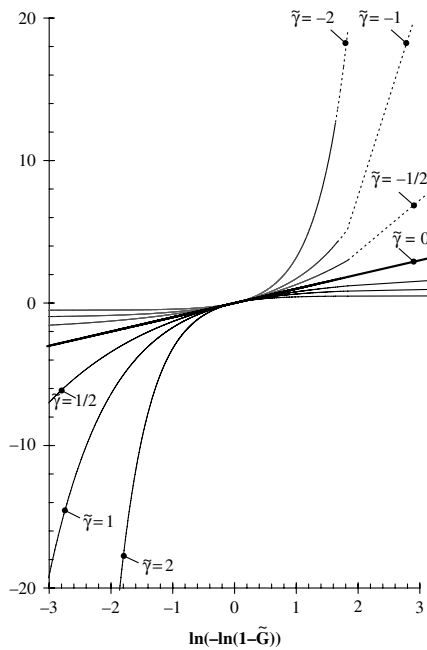


Figura 2.4. Cuantiles de la distribución mínima de GEV, en forma canónica ($\tilde{a}' = 0$ y $\tilde{b}' = 1$). El parámetro toma los valores $\tilde{\gamma}' = -2, -1, 0, 1/2, 1, 2$ donde $\tilde{\gamma}'$ es como α dado en la definición 6

En las aplicaciones prácticas, el interés es a menudo centrado en el máximos anuales. En consecuencia, las observaciones se puede dividir en k bloques consecutivos e independientes, donde k es el número de años observados, y cada bloque contiene $n = 365$ observaciones independientes, si la escala temporal es cada día (o $n = 8760$ en el caso de las muestras por hora, o $n = 525.600$, si la escala temporal es un minuto, y así sucesivamente). Entonces, la observación máxima se calcula para cada bloque: a su vez, se puede tomar una muestra de los máximos de tamaño k . Esta muestra se utiliza para estimar los parámetros de la distribución de los máximos.

El método descrito anteriormente representa el procedimiento estándar para recopilar y analizar, por ejemplo, máximos anuales (o mínimos). Sin embargo, tiene algunos inconvenientes, porque podría descartar la información importante de la muestra. De hecho, si el comportamiento extremo de un fenómeno persiste durante varios días en un año determinado, sólo la observación correspondiente al día más intenso va a generar el máximo anual. Como consecuencia de ello, todo el resto de la información relativa a la dinámica extrema desarrollada durante los días anteriores y posteriores serán descartados.

Esto motiva al estudio de un método alternativo para el análisis de comportamientos extremos, el cual se introduce en la siguiente sección.

2.2 MÉTODOS DE UMBRAL Y DISTRIBUCIÓN GENERALIZADA DE PARETO

INTRODUCCIÓN

Como se anticipó en la nota 2.2, a veces el análisis de los extremos a través del método de bloques proporciona una escasa representación del comportamiento extremo de un fenómeno.

Como alternativa, los máximos pueden ser analizados a través del método Peaks-Over-Threshold (POT). Consideremos una sucesión de v.a.i.i.d. X_1, X_2, \dots , que constituye una muestra de una distribución común F . El comportamiento extremo de estas v.a. pueden ser estudiadas considerando los eventos $\{X > u\}$, donde u es un umbral arbitrario, es decir, investigando excesos sobre un umbral dado.

En este caso, el comportamiento extremo se describe usualmente a través de la probabilidad condicional

$$P\{X > u + x \mid X > u\} = \frac{1 - F(u + x)}{1 - F(u)}, \quad x > 0 \quad (2.68)$$

Si la distribución subyacente F es conocida, también lo es la probabilidad de la ecuación. (2.68). Sin embargo, en las aplicaciones prácticas F no es conocida en general. Por lo tanto, una manera natural de proceder es buscar una aproximación para la distribución condicional de F . El teorema 2.14 es fundamental en el análisis de máximos por el método POT, y define la manera implícita generalizada de Pareto (GPD) de la distribución.

Ajuste de excesos sobre un umbral

Como se ha mencionado la teoría de valores extremos, primero que todo, trata con el comportamiento estocástico del máximo y el mínimo de variables aleatorias iid. Las propiedades de la distribución de extremos (máximo y mínimo), orden estadístico intermedio, de extremos, y de excesos sobre un umbral, son determinados por la cola de la distribución subyacente.

Recíprocamente, la cola de la distribución subyacente puede ser estudiada por medio de procedimientos estadísticos basados en el orden estadístico intermedio o extremo o en excesos sobre un umbral.

Definición 21 (Función de distribución de excesos). Sean X_1, \dots, X_n i.i.d. con una $F \in D(G_{\xi})$ para algún $\xi \in R$

Primero, seleccionamos un umbral alto u y se denota por

$$N_u = \text{card}\{i : i = 1, \dots, n, X_i > u\}$$

El número de excesos de u de X_{1, \dots, X_n} se denota los correspondientes excesos por Y_1, \dots, Y_n . La función de distribución de los excesos de X está dada por

$$\begin{aligned} F_u(y) &= P(X - u \leq y / X > u) \\ &= P(Y \leq y / X > u) \quad , \quad y \geq 0 \end{aligned}$$

Es la función de distribución de los excesos sobre el umbral u . La función anterior se puede escribir también como: $\bar{F}(u + y) = \bar{F}(u)\bar{F}_u(y)$. La función del promedio de los excesos juega un papel muy importante en muchos campos. Diferentes nombres surgen de aplicaciones específicas. Por ejemplo, F_u es conocida como la

distribución de probabilidad del exceso de vida o tiempo de vida residual en la teoría de la confiabilidad y en estadísticas médicas. En el contexto de seguros, F_u es usualmente referida como la función de distribución del exceso de pérdida, por lo tanto, el siguiente teorema permite caracterizar el modelo asintótico de los excesos sobre un umbral.

A continuación se demuestran cinco propiedades de la Distribución de Pareto Generalizada. La primera propiedad a su vez consta de tres numerales.

Definición 22 (Función del promedio de los excesos). La función

$$e(u) = E(X - u / X > x) \quad (2.69)$$

Es llamada la función del promedio de los excesos de X .

Teorema 2.13 (Propiedades de la Distribución de Pareto Generalizada GPD). Supongamos que X tiene una Distribución Generalizada de Pareto con parámetro ζ y β . Entonces $E(X) < \infty$ sí y solo si $\zeta < 1$.

Se demuestra con el enunciado anterior los siguientes tres literales: a, b y c.

a) Dado el supuesto (a) demostrar que: Si $r = -\frac{1}{\zeta}$ entonces

$$E \left[\left(1 + \frac{\zeta}{\beta} x \right)^{-r} \right] = \frac{1}{1 + \zeta r} \quad (2.70)$$

Demostración.

$$\begin{aligned} E \left[\left(1 + \frac{\zeta}{\beta} x \right)^{-r} \right] &= \int_0^\infty \left(1 + \frac{\zeta}{\beta} x \right)^{-r} \frac{1}{\beta} \left(1 + \frac{\zeta}{\beta} x \right)^{-\frac{1}{\zeta}-1} dx \\ &= \frac{1}{\beta} \lim_{b \rightarrow \infty} \int_0^b \left(1 + \frac{\zeta}{\beta} x \right)^{-r-\frac{1}{\zeta}-1} dx \end{aligned}$$

Haciendo la siguiente sustitución: $u = 1 + \frac{\zeta}{\beta} x \Rightarrow \frac{\beta}{\zeta} du = dx$

$$\begin{aligned} &= \frac{1}{\zeta} \lim_{b \rightarrow \infty} \int_1^{u_1} u^{-r-\frac{1}{\zeta}-1} du \\ &= \frac{1}{\zeta} \lim_{b \rightarrow \infty} \frac{\left(1 + \frac{\zeta}{\beta} x \right)^{\left(-r-\frac{1}{\zeta}\right)}}{-r-\frac{1}{\zeta}} \Big|_0^b \\ &= \frac{1}{1 + \zeta r} \end{aligned}$$

siempre y cuando $r > -\frac{1}{\zeta}$

b) Dado el supuesto (a) Demostrar que si $k \in \mathbb{N}$ entonces

$$E \left[\left(\ln \left(1 + \frac{\xi}{\beta} x \right) \right)^r \right] = \xi^k k!, \quad (2.71)$$

c) Dado el supuesto (a) Demostrar que si $\xi < \frac{1}{r}$ con $r \in \mathbb{R}$, entonces

$$\begin{aligned} E \left[\left(X \left(\overline{G}(X; \xi, \beta) \right) \right)^r \right] &= \frac{\beta}{(r+1-\xi)(r+1)} \frac{r+1}{|\xi|} > 0 \quad (2.72) \\ E \left[\left(X \left(\overline{G}_{\xi, \beta(x)}(x) \right) \right)^r \right] &= \int_0^\infty x \left(1 + \xi \frac{x}{\beta} \right)^{-\frac{r}{\xi}} \frac{1}{\beta} \left(1 + \xi \frac{x}{\beta} \right)^{-\frac{1}{\xi}-1} dx \\ &= \frac{1}{\beta} \int_0^\infty x \left(1 + \xi \frac{x}{\beta} \right)^{-\frac{r}{\xi}-\frac{1}{\xi}-1} dx \end{aligned}$$

$$u = 1 + \xi \frac{x}{\beta} \Rightarrow dx = \frac{\beta}{\xi} du \quad x = \frac{\beta}{\xi} (u - 1)$$

Por lo tanto queda la siguiente integral:

$$\begin{aligned} &= \frac{\beta}{\xi^2} \lim_{b \rightarrow \infty} \int_0^b (u-1) u^{-\frac{r}{\xi}-\frac{1}{\xi}-1} du \\ &= \frac{\beta}{\xi^2} \lim_{b \rightarrow \infty} \int_0^b \left(u^{-\frac{r}{\xi}-\frac{1}{\xi}} - u^{-\frac{r}{\xi}-\frac{1}{\xi}-1} \right) du \\ &= \frac{\beta}{(r+1-\xi)(r+1)}, \end{aligned}$$

siempre y cuando $\frac{r+1}{|\xi|} > 0$

$$E(X^r) = \frac{\beta^r \Gamma(\xi^{-1} - r)}{\xi^{r+1} \Gamma(1 + \xi^{-1})} r! \quad (2.73)$$

1) Para todo $\xi \in \mathbb{R}$, $F \in D(G_\xi)$ si y solo si

$$\lim_{u \rightarrow x^*} \sup \left| F_u(x) - G_{\xi, \beta(u)}(x) \right| = 0$$

para alguna función β positiva.

2) Si suponemos que $x_i \in D(\xi, \beta)$, $i = 1, 2$, entonces

$$\frac{\overline{G}(x_1 + x_2; \xi, \beta)}{\overline{G}(x_1, \beta, \beta)} = \overline{G}(x_2; \xi, \beta + \xi x_1)$$

Demostración.

$$\frac{\overline{G}(x_1 + x_2; \xi, \beta)}{\overline{G}(x_1; \xi, \beta)} = \frac{\left(1 + \xi \frac{x_1 + x_2}{\beta} \right)^{-\frac{1}{\xi}}}{\left(1 + \xi \frac{x_1}{\beta} \right)} = \left(\frac{1 + \xi \frac{x_1 + x_2}{\beta}}{1 + \xi \frac{x_1}{\beta}} \right)^{-\frac{1}{\xi}} = \left(\frac{1 + \xi \frac{x_1}{\beta} + \xi \frac{x_2}{\beta}}{1 + \xi \frac{x_1}{\beta}} \right)^{-\frac{1}{\xi}}$$

$$\begin{aligned}
&= \left(1 + \frac{\xi x_2}{1 + \frac{\xi x_1}{\beta}}\right)^{-\frac{1}{\xi}} = \left(1 + \frac{\xi x_2}{\beta + \xi x_1}\right)^{-\frac{1}{\xi}} = \left(1 + \frac{\xi}{\beta + \xi x_1} x_2\right)^{-\frac{1}{\xi}} \\
&= \overline{G}(x_2; \xi, \beta + \xi x_1)
\end{aligned}$$

- 3) Si se asume que n es Poi (λ) independiente de una sucesión $\{X_n\}$ i. i. d. con una Distribución Generalizada de Pareto con parámetros ξ, β . Escrito $M_n = \max(X_1, \dots, X_n)$. Entonces

$$P(M_n \leq x) = \exp \left\{ -\lambda \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}} \right\} = G(x; \xi, \mu, \beta)$$

donde $\mu = \beta \xi^{-1} (\lambda^\xi - 1)$ y $\sigma = \beta \lambda^\xi$

Demostración.

$$\begin{aligned}
P(M_n \leq x) &= \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} G^n(x; \xi, \beta) = \exp \{ -\lambda \overline{G}(x; \xi, \beta) \} \\
&= \exp \left\{ -\lambda \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}} \right\} \\
&= \exp \left\{ - \left(1 + \xi \frac{x - \xi^{-1} \beta (\lambda^\xi - 1)}{\beta \lambda^\xi}\right)^{-\frac{1}{\xi}} \right\}, \quad \xi \neq 0
\end{aligned}$$

Para el caso $\xi = 0$ se reduce a $P(M_n \leq x) = \exp \left\{ -e^{-\frac{x - \beta \ln \lambda}{\beta}} \right\}$

- 4) Supongamos que x tiene Distribución Generalizada de Pareto con parámetro $\xi < 1$ y β , entonces para $u < x^*$

$$e(u) = E(X - u / X > u) = \frac{\beta + \xi u}{1 - \xi}, \quad \beta + u\xi > 0$$

Demostración.

$$e(u)E(X - u / X > u) = \int_u^\infty (x - u) \frac{dF(x)}{\overline{F}(u)} = \frac{1}{\overline{F}(u)} \int_u^\infty \overline{F}(x) dx$$

$\overline{F}(x) = \overline{G}(x; \xi, \beta) = \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}}$ reemplazando se tiene

$$e(u) = \frac{1}{\left(1 + \xi \frac{u}{\beta}\right)^{-\frac{1}{\xi}}} \int_u^\infty \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}} dx$$

$$v = 1 + \zeta \frac{x}{\beta} \Rightarrow dv = \frac{\zeta}{\beta} dx \Rightarrow dx = \frac{\beta}{\zeta} dv$$

$$e(u) = \frac{1}{\left(1 + \zeta \frac{u}{\beta}\right)^{-\frac{1}{\beta}}} \lim_{b \rightarrow \infty} \int_u^b v^{-\frac{1}{\zeta}} \frac{\beta}{\zeta} dv = \frac{\beta}{\zeta \left(1 + \zeta \frac{u}{\beta}\right)^{-\frac{1}{\beta}}} \lim_{b \rightarrow \infty} \left[v^{-\frac{1}{\zeta} + 1} \right]$$

$$e(u) = \frac{\beta + \zeta u}{1 - \zeta},$$

siempre y cuando $\beta + \zeta u > 0$.

La propiedad (3) anterior es a menudo reformulada como sigue: *la clase de GPDs es cerrada con respecto al cambio del umbral*. Es más las variables aleatorias en la ecuación de (3) es la probabilidad condicional que, dado de que las variables aleatorias excedan a x_1 , es también que excedan el umbral $x_1 + x_2$. La regla establece que esta probabilidad es otra vez de tipo Pareto generalizada. Esta propiedad cerrada es importante en reaseguros, donde las GDP son básicas cuando se tratan con contratos con excesos de pérdidas.

La propiedad (2) anterior sugiere una GPD como una aproximación apropiada de los excesos F_u para u grandes. Este resultado se remota a Pickands y es a menudo formulado como sigue. Para alguna función β que sea estimada de los datos,

$$\bar{F}(x; u) = P(X - u > x / X > u) \approx \bar{G}(x; \zeta, \beta(u)) \quad , \quad x > 0$$

Alternativamente uno considera para $x > u$,

$$P(X > x / X > u) \approx \bar{G}(x; \zeta, u, \beta(u)) \tag{2.74}$$

En ambos casos u debe ser tomado lo suficientemente grande. Juntos (2) y (5) dan una técnica gráfica maravillosa para seleccionar u tan alto que una aproximación de la función de los excesos a una GPD está justificada: dada una muestra aleatoria X_1, \dots, X_n i. i. d. construir la función del promedio de los excesos empírica $e(u)$. De (5) se sabe que la función del promedio de los excesos de una GPD es lineal, por lo tanto chequear para una región u donde el gráfico de $e_n(u)$ llegue a ser rigurosamente lineal. Para tal u una aproximación de F_u por GPD parece razonable.

De las observaciones anteriores sugerimos el siguiente modelo aproximado para los excesos de una muestra i. i. d:

- El número de excesos de un umbral alto sigue un proceso de Poisson.

- Los excesos sobre un umbral alto pueden ser modelados por una Distribución Generalizada de Pareto (GPD).
- Un valor apropiado para el umbral alto puede ser encontrado graficando la función del promedio de los excesos empírica.
- La distribución del máximo de un número de Poisson de excesos iid sobre un umbral alto es una Distribución de valor extremo generalizada.

La propiedad (2) del teorema (2.14) da un resultado límite para $\bar{F}(y; u)$ como

$$\lim_{u \rightarrow x^*} \sup |F(x; u) - G(x; \xi, \beta(u))| = 0$$

Para una función positiva β . Basada en esta relación, para un u grande la siguiente aproximación se sugiere:

$$\bar{F}(y; u) \approx \bar{G}(y; \xi, \beta(u)) \quad (2.75)$$

Es importante resaltar que β es una función del umbral u . En la práctica u deberá tomar un valor suficientemente grande. Dado un valor de u , ξ , y $\beta = \beta(u)$ son estimados a partir de los datos de los excesos, por lo tanto los datos estimados dependerán de u .

La ecuación (2.75) luego sugiere un método para estimar el extremo de la cola de F estimado $\bar{F}_u(y)$ y $\bar{F}(u)$ separadamente. Un estimador natural para $\bar{F}(u)$ es dado por la función de distribución empírica

$$(\bar{F}(u)) = \bar{F}_n = \frac{1}{n} \sum_{i=1}^n I_{(x_i > u)} = \frac{N_u}{n}$$

Por otro lado la aproximación de la generalizada de Pareto motiva un estimador de la forma $(\bar{F}(u))(\bar{F}_u(y)) = \bar{G}(y; \hat{\xi}, \beta)$ Para un $\hat{\xi} = \hat{\xi}_{N_u}$ y $\beta = \beta_{N_u}$. Un estimador para la cola para cuando $y > 0$ tiene la forma:

$$(\bar{F}(u + y)) = \frac{N_u}{n} \left(1 + \hat{\xi} \frac{y}{\beta} \right)^{-\frac{1}{\hat{\xi}}} \quad (2.76)$$

En el caso de Frechet y de Gumbel ($\xi \geq 0$) la restricción del dominio es $y \geq 0$ se destaca claramente que se estima \bar{F} en la cola superior. Un estimador del cuantil x^* es el siguiente:

$$\hat{x}_p = u + \frac{\beta}{\hat{\xi}} \left(\left(\frac{n}{N_u} (1 - p) \right)^{-\hat{\xi}} - 1 \right)$$

Además para $\hat{\xi} < 0$ un estimador del punto superior de la derecha x^* de F está dado por $\hat{x}^* = u - \frac{\beta}{\hat{\xi}}$. Lo anterior es obtenido poniendo $\hat{x}^* = \hat{x}_1$ es decir ($p = 1$) en la ecuación anterior.

Teorema 2.14 (Distribución Generalizada de Pareto³). Sea X_1, X_2, \dots una sucesión de variables aleatorias i.i.d. con función de distribución común y sea $M_n = \{X_1, \dots, X_n\}$ sea X un término X_i cualquiera de la sucesión y supongamos que F satisface el teorema 3.4.

Si $P(M_n \leq x) \approx G(x; \xi, \mu, \sigma)$, $n \rightarrow \infty$ donde

$$G(x; \xi, \mu, \sigma) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}$$

para algún $\mu, \sigma > 0$ y ξ , entonces, para algún u suficientemente grande la función de distribución de $X - u$ condicionado a que $X > u$ definida sobre

$$\left\{ y : y > 0 \text{ y } \left(1 + \frac{\xi y}{\sigma} \right) > 0 \right\}$$

donde

$$\hat{\sigma} = \sigma + \xi(u - \mu) \tag{2.77}$$

es

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma} \right)^{-\frac{1}{\xi}} \tag{2.78}$$

Demostración. Se considera la v. a. X lo cual tiene una función de distribución F que cumple la suposición del teorema 3.4, entonces

$$F^n \approx \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}$$

Para algunos parámetros $\mu, \sigma > 0$ y ξ Por lo tanto,

$$n \log(F) \approx - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \tag{2.79}$$

Luego para valores grandes de x , una expansión de Taylor indica que:

$$\log F(x) \approx -\{1 - F(x)\}$$

³Para un desarrollo de la estimación de parámetros y cuantiles ver el artículo *Parameter and Quantile Estimation for the Generalized Pareto Distribution* escrito por los autores J. R. M Hosking and J. R. Wallis.

Y sustituyendo está última expresión en (2.79), obtenemos después de re arreglar

$$1 - F(u) \approx \frac{1}{n} \left[1 + \xi \left(\frac{u - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \quad (2.80)$$

Para u grande. Similarmente, para $y > 0$

$$1 - F(u + y) \approx \frac{1}{n} \left[1 + \xi \left(\frac{u + y - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}}$$

por lo tanto,

$$\begin{aligned} P \left(X > u + \frac{y}{X} > u \right) &\approx \frac{n^{-1} \left[1 + \frac{\xi(u + y - \mu)}{\sigma} \right]^{-\frac{1}{\xi}}}{n^{-1} \left[1 + \frac{\xi(u - \mu)}{\sigma} \right]^{-\frac{1}{\xi}}} \\ &= \left[\frac{\xi(u + y - \mu)}{\sigma} \right]^{-\frac{1}{\xi}} \\ &= \left[1 + \frac{\frac{\xi(u + y - \mu)}{\sigma}}{1 + \frac{\xi(u - \mu)}{\sigma}} \right]^{-\frac{1}{\xi}} \\ &= \left[1 + \frac{\xi y}{\sigma} \right]^{-\frac{1}{\xi}} \end{aligned} \quad (2.81)$$

Donde $\hat{\sigma} = \sigma + \xi(u - \mu)$

La familia de distribuciones definida por la ecuación (2.77) se conoce distribución generalizada de Pareto (GPD). Por lo tanto para un umbral u lo suficientemente grande y existe algún $\hat{\sigma}$ que depende de u y algún ξ para los cuales la distribución generalizada de Pareto es una muy buena aproximación de los excesos sobre un umbral u .

Nota 2.1. La ecuación (2.10) tiene una interesante interpretación probabilística. En realidad, Sea X una rv con $df F \in MDA(H_{\xi})$ entonces (2.10) reformulada como

$$\lim_{x \uparrow x_F} P \left(\frac{X - u}{a(u)} > x \mid X > u \right) = \begin{cases} (1 + \xi x)^{-1/\xi} & \text{si } \xi \neq 0, \\ e^{-x} & \text{si } \xi = 0. \end{cases} \quad (2.82)$$

Por lo tanto (2.82) da una aproximación distribucional para los excesos escalados sobre el umbral (alto) u . El factor de escala apropiado es $a(u)$. Esta interpretación resultará crucial en muchas aplicaciones.

A continuación trataremos el resultado de Pickands el cual asegura que las funciones de distribución cuyas colas se pueden aproximar por las distribuciones de Pareto generalizadas son las funciones que se encuentran en los dominios de atracción de las distribuciones de valores extremos. Antes de enunciar el teorema recuerde que cada EVD posee un parámetro llamado *alpha* ver el teorema 2.2.

Definición 23 (Convergencia en Distribución). *Decimos que (A_n) converge en distribución o converge débilmente a la rv $A(A_n \xrightarrow{d} A)$ si para todas las funciones continuas acotadas f , la relación $Ef(A_n) \rightarrow Ef(A)$, $n \rightarrow \infty$ se cumple.*

La convergencia débil puede ser descrita por los dfs F_{A_n} y F_A de A_n y A , respectivamente: $A_n \xrightarrow{d} A$ se mantiene si y solamente si todos los puntos de continuidad y de las df F_A la relación

$$F_{A_n}(y) \rightarrow F_A(y), \quad n \rightarrow \infty \tag{2.83}$$

se satisface. Mas aún si F_A es continua entonces (2.83) puede fortalecer la convergencia uniforme: $\sup_x |F_{A_n}(x) - F_A(x)| \rightarrow 0$, $n \rightarrow \infty$.

La convergencia débil es metrizable, es decir, $\rho(G, H)$ la cual es definida en el espacio de todas las dfs o distribuciones en $[\Omega, \mathcal{F}]$ tal que $A_n \xrightarrow{d} A$ si y solo si $\rho(F_{A_n}, F_A) \rightarrow 0$. Una métrica bien conocida para este propósito es la métrica de Lévy:

$$\rho_L(F, G) = \inf \{ \varepsilon > 0 : \forall x \in \mathbb{R}, \quad F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon \}.$$

Definición 24 (Distribución generalizada de Pareto). *La distribución definida en la ecuación (2.78) es llamada la distribución generalizada de Pareto (GPD).*

Teorema 2.15 (Pickands). *Suponga que una función de distribución $F(\cdot)$ está en el dominio de atracción de una EVD con parámetro α , donde α es como en las ecuaciones (1.4), (1.5) y (1.6), sea*

$$\xi = \begin{cases} \alpha^{-1} & \text{Si } F(\cdot) \text{ está en el dominio de atracción Fréchet} \\ -\alpha^{-1} & \text{Si } F(\cdot) \text{ está en el dominio de atracción Weibull} \\ 0 & \text{Si } F(\cdot) \text{ está en el dominio de atracción Gumbel} \end{cases}$$

Entonces existe una función positiva $\beta(\cdot)$ tal que

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\xi, \beta(u), 0}(x)| = 0$$

Si $F(\cdot)$ tiene cola infinita:

$$\lim_{u \uparrow +\infty} \sup_{0 < x < \infty} |F_u(x) - G_{\xi, \beta(u), 0}(x)| = 0$$

Demostración. El enunciado del teorema lo podemos reescribir de la siguiente manera:

Para todo $\xi \in \mathbb{R}$, $F \in MDA(H_\xi)$ si y solo si

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0 \quad (2.84)$$

para alguna función β positiva.

Del teorema 2.2 y de la nota 2.1 tenemos que $F \in MDA(H_\xi)$ si y solo si

$$\lim_{u \uparrow x_F} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0$$

donde $\beta(u) = a(u)$. Ahora, como la GPD es continua, se sigue la uniformidad de la convergencia (vea definición 23) \square

USO EN FINANZAS

Aunque no haremos aplicaciones en finanzas, presentamos una relación importante entre distribuciones con cola gruesa y distribuciones con cola fina, por ejemplo, el retorno diario es una variable aleatoria ilimitada, o sea, la función de distribución del retorno diario posee cola derecha infinita.

Si p es próximo de 1 se desea $G_{\xi, \beta, S}^{-1}(p)$

Si $\xi_1 \leq \xi_2$ entonces $G_{\xi_1, \beta, S}^{-1}(p) \leq G_{\xi_2, \beta, S}^{-1}(p) \quad \forall p \in [0, 1]; \forall \beta; \forall S.$

Si $\beta_1 \leq \beta_2$ entonces $G_{\xi, \beta_1, S}^{-1}(p) \leq G_{\xi, \beta_2, S}^{-1}(p) \quad \forall p \in [0, 1]; \forall \xi; \forall S.$

Si $S_1 \leq S_2$ entonces $G_{\xi, \beta, S_1}^{-1}(p) \leq G_{\xi, \beta, S_2}^{-1}(p) \quad \forall p \in [0, 1]; \forall \xi; \forall \beta.$

Terminología: Entre dos GPD's la que tiene el parámetro ξ mayor se dice que tiene una **cola más pesada o gruesa**. El que tiene el parámetro ξ menor tiene **cola más leve o fina**. Esta caracterización se puede llevar a una forma más profunda dependiendo del signo de ξ , como se muestra a continuación

$\xi > 0 \rightarrow$ Colas polinomiales.

$\xi = 0 \rightarrow$ Cola exponencial.

$\xi < 0 \rightarrow$ Cola finita.

Interpretación: Decir que la cola de las pérdidas de la inversión A es más leve que la de la inversión B . Significa que la inversión A tiene riesgo menor que B como se observa en la siguiente figura.

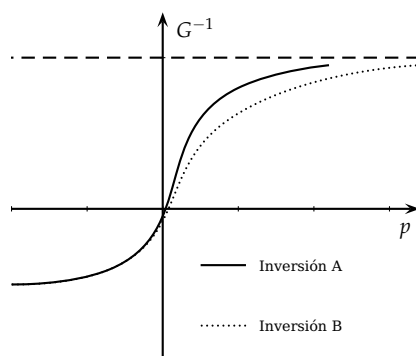


Figura 2.5. Relación entre una cola gruesa y una cola fina

CAPÍTULO 3

ESTIMACIÓN DEL ÍNDICE DE VALOR EXTREMO Y PRUEBAS

INTRODUCCIÓN

Cuando aplicamos la teoría de valor extremo la primera decisión que se debe tomar es si se aplica el método de bloques o picos sobre un umbral de acuerdo a la disponibilidad de datos que se tenga. Si se elige el primer método se trabaja con la distribución de valores extremos y si se elige el segundo se trabaja con la distribución generalizada de Pareto. Identificada la distribución con la cual se va a trabajar se procede a estimar los parámetros de la distribución: Existen tres parámetros uno de forma, uno de escala y otro de localización. El parámetro más importante es el de forma o índice de valor extremo porque después solo hay que trasladar y contraer como en (2.4); ahora, éste define el tipo de distribución, esto es, si $\zeta > 0$ es la distribución tipo 2 Frechet, si $\zeta < 0$ la distribución tipo 3 Weibull, y si $\zeta \rightarrow \pm\infty$, es la distribución tipo 1 Gumbel.

Hay varios métodos para estimar el índice de valor extremo, entre los cuales se tiene: Hill ($\hat{\zeta}_H$) Bajo condiciones apropiadas, el estimador de Hill es consistente solamente para valores positivos de ζ , Pickands ($\hat{\zeta}_P$), Máximo Verosimilitud ($\hat{\zeta}_{MV}$), Momentos ponderados por probabilidad¹ ($\hat{\zeta}_{MPP}$), y Hill Negativo ($\hat{\zeta}_F$), el de Máxima Verosimilitud (MV) está definido para $\zeta > -\frac{1}{2}$, el de Pickands y el de momentos están definidos y son consistentes para todos los valores reales de ζ , el Momentos Ponderados por probabilidad (MPP) es consistente para $\zeta < 1$, y el Hill Negativo para $\zeta < -\frac{1}{2}$.

Para el cálculo del parámetro de forma existen diferentes métodos para estimarlo el estimador de Hill del cual hay dos posibilidades según el valor que tenga mayor que cero o menor de $-\frac{1}{2}$ y el estimador de Pickands que sirve para el caso

¹Para un desarrollo amplio del Método de estimación de parámetros de Momentos ponderados por probabilidad en el artículo Estimation of Generalized Extreme-Value Distribution by the Method of probability-Weighted Moments. Revista American Society for Quality.

general su valor es posiblemente cualquier valor real, Un estimador de momentos para un valor dentro de los reales, un estimador de momentos ponderados según una probabilidad cuando el valor es menor de 1 en nuestro caso se elegirá el estimador de Hill.

También deben estimarse las constantes de normalización, los cuantiles y la probabilidad de la cola.

Pero cuando se usa un método de estimación: Método de Hill, Método de Máxima verosimilitud², Método de Momentos Ponderados por Probabilidad o cualquier otro, la diferencia entre uno y otro es insesgadez, la eficiencia, la consistencia y la suficiencia.

También se debe saber cuál es la distribución del parámetro porque a partir de esta información se construyen las pruebas estadísticas y los intervalos de confianza pertinentes.

Las dos propiedades en las que se concentrará este trabajo son la consistencia y la distribución asintótica de los estimadores.

Definición 25 (Estimador consistente). Sean X_1, X_2, \dots , una sucesión de variables aleatorias iid de una distribución con parámetro θ y $\hat{\theta}$ un estimador para θ . Decimos que $\hat{\theta}$ es un estimador consistente para θ si:

$$\hat{\theta} \xrightarrow{P} \theta \quad \text{o} \quad \lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| \leq \varepsilon) = 1$$

La condición alternativa al teorema 2.3 sirve como una base para las aplicaciones de la teoría del valor extremo.

Considere la relación 2.28; existe una función no decreciente positiva f tal que

$$\lim_{t \rightarrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 - \zeta x)^{-\frac{1}{\zeta}} \quad (3.1)$$

Para todo x para el cual $1 + \zeta x > 0$ donde $x^* = \sup\{x : F(x) < 1\}$

Sea X una v. a. con función de distribución F y sea $F \in D(G_\zeta)$ para algún $\zeta \in R$.

Entonces (3.1) nos dice que $x > 0$ y $x < (0 \vee (-\zeta))^{-1}$

$$\lim_{t \rightarrow x^*} P \left[\frac{x - t}{f(t)} > x/X > t \right] = (1 + \zeta x)^{\frac{1}{\zeta}} \quad (3.2)$$

²Para un desarrollo amplio ver el artículo de *Maximum Likelihood Estimation in a Class of Nonregular Cases* de Richard L. Smith. Revista Bimetrika Trust.

Esto es, la distribución condicional de $\frac{x-t}{f(t)}$ dado $X > t$ tiene una distribución límite cuando $t \rightarrow x^*$.

$$H_{\zeta}(x) := 1 - (1 + \zeta x)^{-\frac{1}{\zeta}} \quad , \quad 0 < x < (0 \vee (-\zeta))^{-1} \quad (3.3)$$

Donde para $\zeta = 0$ el lado derecho es interpretando $1 - e^{-x}$ Esta clase de funciones de distribuciones es llamada la función generalizada de Pareto.

La relación (3.1) significa en términos generales que de algún umbral alto t en adelante (es decir, $X > t$) la función de distribución puede ser escrita como

$$P(X > x) = 1 - F(x) \cong (1 - F(t)) \left\{ 1 - H_{\zeta} \left(\frac{x-t}{f(t)} \right) \right\} \quad , \quad x > t$$

La cual es una familia paramétrica de distribución de la cola.

Se puede esperar que esta aproximación se cumpla para estadísticos de orden intermedios y extremos. Sea X_1, \dots, X_n v. a. independientes e idénticamente distribuidas con función de distribución F y F_n la correspondiente distribución empírica, esto es, $F_n(x) := n^{-1} \sum_{i=1}^n \mathbf{1}_{(X_i \leq x)}$

Se aplica la última aproximación con $t := X_{n-k,n}$ donde se selecciona

$$K = K(n) \rightarrow \infty \quad , \quad \frac{k}{n} \rightarrow 0 \quad , \quad n \rightarrow \infty$$

entonces

$$1 - F(x) \approx (1 - F(X_{n-k,n})) \left\{ 1 - H_{\zeta} \left(\frac{x - X_{n-k,n}}{f\left(\frac{n}{k}\right)} \right) \right\}$$

Y, ya que $1 - F(X_{n-k,n}) \approx 1 - F_n(X_{n-k,n}) = \frac{k}{n}$

$$1 - F(x) \approx \frac{k}{n} \left\{ 1 - H_{\zeta} \left(\frac{x - X_{n-k,n}}{f\left(\frac{n}{k}\right)} \right) \right\} \quad (3.4)$$

Para poder usar esta aproximación se necesita estimar ζ y la función f en el punto $\frac{n}{k}$.

Esta aproximación es válida para cualquier x más grande que $X_{n-k,n}$ y puede ser usada siempre que $x > X_{n-k,n}$ y puede ser usada siempre que $x > X_{n,n}$, esto es, fuera del rango de observaciones y este es en efecto la base para aplicaciones de valores extremos.

Posteriormente se considerará una aplicación de la relación (2.26) ya que existe una función positiva tal que $\forall x > 0$

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\xi - 1}{\xi} \quad (3.5)$$

La relación (3.3) se convierte en la siguiente aproximación

$$U(x) \approx U(t) + a(t) \frac{\left(\frac{x}{t}\right)^\xi - 1}{\xi}, \quad x > t \quad (3.6)$$

Esta aproximación es usual cuando uno desea estimar un cuantil

$$F^{\leftarrow}(1 - p) = U\left(\frac{1}{p}\right)$$

Con p muy pequeño, ya que este cuantil está luego relacionado con muchos bajos cuantiles.

$U(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right)$ El cual puede ser estimado por un estadístico de orden intermedio. Por lo tanto se selecciona $t := \frac{n}{k}$ con $k = k(n) \rightarrow \infty$ y $\frac{k}{n} \rightarrow 0$, $n \rightarrow \infty$ Entonces para grandes y y por ejemplo $y = \frac{1}{p}$ con p pequeño,

$$U(y) \approx U\left(\frac{n}{k}\right) + a\left(\frac{n}{k}\right) \frac{\left(\frac{Y_k}{n}\right)^\xi - 1}{\xi} \quad (3.7)$$

En orden a hacer esta aproximación aplicable se necesita estimar ξ la función a en el punto $\frac{n}{k}$ y $U\left(\frac{n}{k}\right)$. El cuantil anterior puede ser estimado por un estadístico de orden intermedio, el cual puede ser usado para $y < n$ y también para extrapolar fuera de la muestra.

3.1 ESTIMADOR DE HILL $\xi > 0$

Para introducir el estimador de Hill, se comienza presentando el teorema 1.2.1 (Ver en Haan y Ferreira (2006))

Dado $\xi > 0$, $F \in D(G)$ sí y solo si

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\frac{1}{\xi}}, \quad \xi > 0$$

El parámetro $\alpha = \frac{1}{\xi}$ es llamado el índice de colas de F el teorema 2.3 nos permite escribir esta expresión de la siguiente manera:

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty (1 - F(x)) \frac{dx}{x}}{1 - F(t)} = \xi$$

Ahora la integral parcial da

$$\int_t^\infty (1 - F(s)) \frac{ds}{s} = \int_t^\infty (\log u - \log t) dF(u)$$

Por lo tanto se tiene

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty (\log u - \log t) dF(u)}{1 - F(t)} = \zeta \tag{3.8}$$

En orden a desarrollar un estimador basado en este resultado asintótico reemplazando en (3.8) el parámetro t por el estadístico de orden intermedio $X_{n-k,n}$ y F por la función de distribución empírica F_n . Entonces se logra el estimador de Hill (1975) $\hat{\gamma}_H$ definido por

$$\hat{\gamma}_H = \frac{\int_{X_{n-k,n}}^\infty \log u - \log X_{n-k,n} dF_n(u)}{1 - F_n(X_{n-k,n})}$$

o

$$\hat{\gamma}_H = \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n} \tag{3.9}$$

Teorema 3.1 (Consistencia débil). Sea X_1, X_2, \dots variables aleatorias i. i. d con función de distribución F . Suponga que $F \in D(G_\zeta)$ con $\zeta > 0$ Entonces cuando $n \rightarrow \infty$, $k = k(n) \rightarrow \infty$ y $\frac{k}{n} \rightarrow 0$

$$\hat{\zeta}_H \xrightarrow{P} \zeta$$

Demostración. Por el corolario 1.2.10 (Ver en Haan y Ferreira (2006)) $F \in D(G_\zeta)$ con $\zeta > 0$ implica $\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\zeta$ para $x > 0$, esto es (Proposición B.1.9) (Ver en Haan y Ferreira (2006)) para $x \geq 1$, y $t \geq t_0$

$$(1 - \zeta)x^{\zeta - \zeta'} < \frac{U(tx)}{U(t)} < (1 + \zeta)x^{\zeta + \zeta'}$$

O equivalentemente

$$\log(1 - \zeta) + (\zeta - \zeta') \log x < \log U(tx) - \log U(t) < \log(1 + \zeta) + (\zeta + \zeta') \log x \tag{3.10}$$

Sea Y_1, Y_2, \dots v. a. i. i. d. con distribución común $1 - \frac{1}{y}$, $y \geq 1$

Note que

$$U(Y_i) =^d X_i \quad , \quad i = 1, 2$$

Así que es suficiente probar el resultado para

$$\hat{\zeta}_H = k^{-1} \sum_{i=1}^{k-1} \log U(Y_{n-i,n}) - \log U(Y_{n-k,n})$$

Aplicar (3.10) con $t \cong Y_{n-k,n}x - \frac{Y_{n-i,n}}{Y_{n-k,n}}$ ya que por el lema 1 (ver Haan y Ferreira (2006)) $Y_{n-k,n} \rightarrow \infty$ a. s. $n \rightarrow \infty$ eventualmente

$$\begin{aligned} \log(1 - \varepsilon) + (\zeta - \varepsilon') \log \frac{Y_{n-i,n}}{Y_{n-k,n}} &< \log U(Y_{n-i,n}) - \log U(Y_{n-k,n}) \\ &< \log(1 + \varepsilon) + (\zeta - \varepsilon') \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) \end{aligned}$$

Para $i = 0, 1, 2, 3, \dots, k-1$, por lo tanto

$$\begin{aligned} \log(1 - \varepsilon) + (\zeta - \varepsilon') \frac{1}{k} \sum_{i=1}^n \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) &< \hat{\zeta}_H \\ &< \log(1 + \varepsilon) + (\zeta - \varepsilon') \frac{1}{k} \sum_{i=1}^n \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) \end{aligned}$$

Además es suficiente probar que cuando $n \rightarrow \infty$

$$\frac{1}{k} \sum_{i=1}^{k-1} \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) \xrightarrow{P} 1$$

Teorema 3.2 (Consistencia débil). Sea X_1, X_2, \dots v. a. i. i. d con función de distribución F . Supongamos que para una sucesión de enteros $k = k(n) \rightarrow \infty$, $\frac{k}{n} \rightarrow 0$ y $\frac{k(n+1)}{k(n)} \rightarrow 1$ cuando $n \rightarrow \infty$,

$$\hat{\zeta}_H \xrightarrow{P} \zeta$$

entonces $F \in D(G_\zeta)$

Demostración. Sea F_n la función de distribución empírica de X_1, \dots, X_n y G_n la función de distribución empírica de Y_1, \dots, Y_n las cuales son independientes e igualmente distribuidas $1 - \frac{1}{x}$, $x \geq 1$ luego para cada n

$$1 - F_n(x) =^d 1 - G_n \left(\frac{1}{1 - F(x)} \right)$$

Se escribe

$$\hat{\zeta}_H = \frac{n}{k} \int_{X_{n-k,n}}^{\infty} \left\{ 1 - G_n \left(\frac{1}{1 - F(u)} \right) \right\} \frac{du}{u} = \frac{n}{k} \int_{X_{n-k,n}}^{\infty} (1 - G_n(s)) d \log U(s)$$

con $Y_{1,n} \leq Y_{2,n} \leq \dots \leq Y_{n,n}$ el estadístico de orden de Y_1, \dots, Y_n Se van a usar los siguientes resultados

1. $P(\sup s(1 - G_n(s)) > b) = \frac{1}{b}$, para $b > 1$

$$P(\inf s(1 - G_n(s)) < b) = \frac{e}{a} e^{-\frac{1}{a}} \quad \text{para } 0 < a < 1$$

(Ver Shoroack y Wellner (1986), pp. 345 y 415)

2. $\inf s(1 - G_n(s)) = \inf \frac{ks}{n} Y_{n-[ks],n} = \left(\inf s \frac{Y_{n-[ns],n}}{Y_{n-k,n}} \right) \left(\frac{k}{n} Y_{n-k,n} \right)$ donde los dos factores son independientes.

$$\left(\inf s Y_{k-[ks],n}^* \right) \frac{k}{n} Y_{n-k,n}$$

con los Y^* independientes e iguales en distribución a los Y esto

$$(\inf s(1 - G_n(s))) \frac{k}{n} Y_{n-k,n}$$

3. Del corolario 2.2.2 (ver en Haan y Ferreira (2006))

$$\lim_{n \rightarrow \infty} P \left(1 - \varepsilon \leq \frac{k}{n} Y_{n-k,n} \leq 1 + \varepsilon \right) = 1$$

Para t suficientemente grandes sea $n = n(t)$ el entero que satisface

$$\frac{n(t)}{k(n(t))} \leq t \leq \frac{n(t) + 1}{k(n(t) + 1)}$$

Considere

$$\begin{aligned} P \left(\frac{1 - \varepsilon}{(1 + \varepsilon)^2} t(1 + \varepsilon) \int_{t(1+\varepsilon)}^{\infty} \frac{d \log U(s)}{s} \leq \frac{n}{k} (1 - \varepsilon) \int_{\frac{n}{k}(1+\varepsilon)}^{\infty} \frac{d \log U(s)}{s} \right. \\ \leq Y_{n-k,n} \int_{Y_{n-k,n}}^{\infty} \frac{d \log U(s)}{s} \\ \leq \frac{n}{k} \int_{Y_{n-k,n}}^{\infty} \frac{s(1 - G_n(s))}{1 - \varepsilon} \frac{d \log U(s)}{ds} \\ \left. \leq \frac{\tilde{\zeta} + \varepsilon}{1 - \varepsilon} \right) \end{aligned}$$

La primera desigualdad es verdad por definición: la segunda y la cuarta desigualdad son verdad con probabilidad tendiendo a 1. Para la segunda usamos el resultado (3); la cuarta es verdad por supuesto. Por los resultados (1) y (2) la tercera desigualdad es verdad con probabilidad al menos

$$1 - e(1 - \varepsilon)^{-1} \exp \left(-\frac{1}{1 - \varepsilon} \right) > 0$$

por lo tanto se alcanza la conclusión para cada $\varepsilon > 0$

$$P\left(t \int_t^\infty \frac{d \log U(s)}{s} \leq \varepsilon + \zeta\right) > 0$$

para t suficientemente grande. Por lo tanto para t suficientemente grande,

$$t \int_t^\infty s^{-1} d \log U(s) \leq \zeta + \varepsilon$$

se consigue la otra desigualdad en forma similar, Por tanto

$$\lim_{t \rightarrow \infty} t \int_t^\infty \frac{d \log U(s)}{s} = \zeta$$

Ahora por integración parcial

$$t \int_t^\infty \frac{1}{s} d \log U(s) = t \int_t^\infty \log U(s) \frac{ds}{s} - \log U(t)$$

Por lo tanto por la observación B.2.14(2) (Ver en Haan y Ferreira (2006)) se define para $x > 0$

$$\lim_{t \rightarrow \infty} (\log U(tx) - \log U(t)) = \zeta \log x$$

Es decir, U varía regularmente con índice ζ lo cual implica por la proposición B.1.9 (ver en Haan y Ferreira (2006)) que la función $1 - F$ varía regularmente con índice $-\frac{1}{\zeta}$

Teorema 3.3 (Normalidad asintótica para ζ_h). Suponga que la función de distribución F satisface la condición de segundo orden, esto es, para $x > 0$

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\zeta}{A(t)} = x^\zeta \frac{x^\rho - 1}{\rho} \quad (3.11)$$

o equivalentemente

$$\lim_{t \rightarrow \infty} \frac{1 - \frac{F(tx)}{F(t)} - x^{-\frac{1}{\zeta}}}{A\left(\frac{1}{1-F(t)}\right)} = x^{-\frac{1}{\zeta}} \frac{x^{\frac{\rho}{\zeta}} - 1}{\zeta^\rho} \quad (3.12)$$

Donde $\zeta > 0$, $\rho \leq 0$ y A es una función positiva o negativa con $\lim_{t \rightarrow \infty} A(t) = 0$ y $k = k(n) \rightarrow \infty$, $\frac{k}{n} \rightarrow \infty$, $n \rightarrow \infty$, y

$$\lim_{n \rightarrow \infty} \sqrt{k} A\left(\frac{n}{k}\right) = \lambda \quad (3.13)$$

entonces $\sqrt{k} (\hat{\zeta}_H - \zeta) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \zeta\right)$

Demostración. Se puede escribir la ecuación de segundo orden como

$$\lim_{t \rightarrow \infty} \frac{\frac{x^{-\zeta} U(tx)}{U(t)} - 1}{A(t)} = \frac{x^\rho - 1}{\rho}$$

Por hipótesis $\lim_{t \rightarrow \infty} A(t) = 0$, esto es equivalente a

$$\lim_{t \rightarrow \infty} \frac{\log U(tx) - \log U(t) - \zeta \log x}{A(t)} = \frac{x^\rho - 1}{\rho}$$

Se aplica la desigualdad dada en el teorema B.2.18 (ver en Haan y Ferreira (2006)): para una función diferencial posible A_0 , con $A_0(t) \sim A(t)$, $t \rightarrow \infty$ y para cada $\varepsilon > 0$ existe un t_0 tal que para $t \geq t_0$, $x \geq 1$,

$$\left| \frac{\log U(tx) - \log U(t) - \zeta \log x}{A_0(t)} - \frac{x^\rho - 1}{\rho} \right| \leq \varepsilon x^{\rho+\varepsilon} \quad (3.14)$$

como se prueba en el teorema 3.1 se nota que

$$\hat{\zeta}_H = \frac{1}{k} \sum_{i=0}^{k-1} \log U(Y_{n-1,n}) - \log U(Y_{n-k,n}),$$

Donde los Y_i son independientes y tienen una función de distribución común $1 - \frac{1}{x}$, $x \geq 1$ Por lo tanto se trabaja con esta representación para $\hat{\zeta}$.

Aplicando (3.14) con $t = Y_{n-k,n} \rightarrow \infty$ a.s. $n \rightarrow \infty$ (Lema 1) y $x = \frac{Y_{n-i,n}}{Y_{n-k,n}}$ y entonces se obtiene, como en el teorema 3.1.

$$\begin{aligned} \hat{\zeta}_H &= \frac{1}{k} \sum_{i=0}^{k-1} \log U(Y_{n-i,n}) - \log U(Y_{n-k,n}) \\ &= \frac{\zeta}{k} \sum_{i=0}^{k-1} \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) + A_0(Y_{n-k,n}) \frac{1}{k} \sum_{i=0}^{k-1} \frac{\left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) - 1}{\rho} \\ &O_\rho(1) |A_0(Y_{n-k,n})| \frac{1}{k} \sum_{i=0}^{k-1} \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^{\rho+\varepsilon} \end{aligned}$$

por lo tanto

$$\begin{aligned} \sqrt{k} (\hat{\zeta}_H - \zeta) &= \zeta \sqrt{k} \left(\frac{1}{k} \sum_{i=0}^{k-1} \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) - 1 \right) \\ &+ \sqrt{k} A_0(Y_{n-k,n}) \frac{1}{k} \sum_{i=0}^{k-1} \frac{\left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^\rho - 1}{\rho} + O_\rho(1) \sqrt{k} |A_0(Y_{n-k,n})| \frac{1}{k} \sum_{i=0}^{k-1} \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^{\rho+\varepsilon} \end{aligned}$$

El primer término, normalizado, es asintóticamente normal por el lema 2. Como en la prueba del lema 2 se tiene

$$\frac{1}{k} \sum_{i=0}^{k-1} \frac{\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^\rho - 1}{\rho} = \frac{1}{k} \sum_{i=0}^{k-1} \frac{Y_i^\rho - 1}{\rho} \quad (3.15)$$

Por lo tanto (3.15) se tiene a

$$\frac{E(Y_i^\rho - 1)}{\rho} = (1 - \rho)^{-1}$$

Por la ley de los grandes números.

Similarmente,

$$\frac{1}{k} \sum \left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{\rho+\varepsilon} \xrightarrow{P} EY_1^{\rho+\varepsilon} = \frac{1}{1 - \rho - \varepsilon}$$

Resta probar que

$$\frac{A_0(Y_{n-k,n})}{A_0\left(\frac{n}{k}\right)} \xrightarrow{P} 1$$

Esto sigue del lema 2.2.3 (ver en Haan y Ferreira (2006)) el hecho que la función $|A_0|$ varía regularmente y las desigualdades de Protter's (Proposición B.1.9 ver en Haan y Ferreira (2006)).

El marco teórico de segundo orden usado en el teorema (3.3) suministra la aproximación más natural a la normalidad asintótica de estimadores como el estimador de Hill. El parámetro ρ controla la velocidad de convergencia para la normalidad asintótica de $\hat{\xi}_H$. Función de distribución F satisface

$$1 - F(x) = c_1 x^{-\frac{1}{\xi}} + c_2 x^{-\frac{1}{\xi} + \frac{\rho}{\xi}} (1 + O(1)) \quad , \quad x \rightarrow \infty$$

y la segunda condición de orden (3.12) con $\xi = 1$ y $\rho = -2$, en efecto es fácil resolver ver si

$$1 - F(x) = c_1 x^{-\frac{1}{\xi}} + c_2 x^{-\frac{1}{\xi} + \frac{\rho}{\xi}} (1 + O(1)) \quad , \quad x \rightarrow \infty$$

para constantes $c_1 > 0$, $c_2 \neq 0$, $\xi > 0$ y $\rho < 0$, entonces, las condiciones de segundo orden (3.12) cumplen con la indicación ξ y ρ con

$$A\left(\frac{1}{1 - F(t)}\right) = \rho \xi^{-1} c_2 c_1^{-1} t^{\frac{\rho}{\xi}}$$

Y por lo tanto (3.11) $A(t) = \rho \zeta^{-1} c_2 c_1^{\rho-1} t^\rho$. Además, si (3.13) se cumple con $\lambda \neq 0$ entonces un simple cálculo muestra que esto es verdad si y solo si

$$k(n) \sim \left(\frac{\lambda \zeta}{\rho c_2} c_1^{1-\rho} \right)^{\frac{2}{1-2\rho}} \eta^{-\frac{2\rho}{1-2\rho}} \quad (3.16)$$

Entonces la tasa de convergencia en el teorema 3.3 es del orden $\eta^{\frac{\rho}{1-2\rho}}$.

Ahora se considerara tres tipos de sucesiones:

1. Suponga $\frac{\sqrt{k}}{|A(\frac{n}{k})|} \rightarrow \infty$ Entonces no es difícil ver usando la desigualdad en la prueba del teorema 3.3 que

$$\frac{\widehat{\zeta}_H - \zeta}{A\left(\frac{n}{k}\right)} \xrightarrow{P} \frac{1}{1-\rho}$$

Ya que para n, k grande debe ser mucho más grande que $n^{-\frac{2\rho}{1-2\rho}}$, se tiene para n grande, $\frac{n}{k}$ mucho más pequeño que

$$n^{\frac{1+2\rho}{1-2\rho}} = n^{\frac{1}{1-2\rho}}$$

por lo tanto la tasa de convergencia $|A\left(\frac{n}{k}\right)|$ es más suave que la tasa $n^{\frac{\rho}{1-2\rho}}$ encontrada después de (3.16).

2. Supongamos que $\sqrt{k}A\left(\frac{n}{k}\right) \rightarrow \infty$. Entonces $k(n) = O\left(n^{-\frac{2\rho}{1-2\rho}}\right)$, y la tasa de convergencia $\frac{1}{\sqrt{k}}$ es otra vez más baja que $\eta^{\frac{\rho}{1-2\rho}}$

3. Supongamos $\sqrt{k}A\left(\frac{n}{k}\right) \rightarrow \lambda \neq 0, \infty$ luego por (3.16) la tasa de convergencia $\frac{1}{\sqrt{k}}$ es del orden $\eta^{\frac{\rho}{1-2\rho}}$. Esta es la situación óptima.

¿Cuál es la mejor selección de λ ? El teorema (3.3) dice que si $\sqrt{k}A\left(\frac{n}{k}\right) \rightarrow \lambda$, entonces

$$\sqrt{k}\left(\widehat{\zeta}_H - \zeta\right) \xrightarrow{d} \zeta N + \frac{\zeta}{1+\rho} \quad (3.17)$$

Con N la normal estándar y por lo tanto

$$\widehat{\zeta}_H - \zeta \approx^d \frac{\zeta N}{\sqrt{k}} + \frac{\zeta}{(1-\rho)\sqrt{k}} \approx^d \frac{\zeta N}{\sqrt{k}} + \frac{A\left(\frac{n}{k}\right)}{1-\rho} \quad (3.18)$$

Para que valor de $k = k(n)$ esta aproximación es la mejor, es decir para que k su error cuadrado medio

$$\frac{\bar{\zeta}^2}{k} + \frac{A^2\left(\frac{n}{k}\right)}{(1-\rho)^2} \quad (3.19)$$

Es mínimo. Para que el tiempo sea continuo se considera el caso especial

$$A(t) = ct^\rho.$$

Escribiendo $r := \frac{n}{k}$. Esto conduce a

$$\arg \min_{\left(\frac{n}{r}\right)=1,2,\dots} \left(\frac{r\bar{\zeta}^2}{n} + \frac{c^2 r^{2\rho}}{(1-\rho)^2} \right)$$

Y para simplificar se considera

$$\arg \min_{t>0} \left(\frac{r\bar{\zeta}^2}{n} + \frac{c^2 r^{2\rho}}{(1-\rho)^2} \right)$$

El ínfimo se alcanza igualando la derivada a cero, es decir

$$\frac{\bar{\zeta}^2}{n} = \frac{-2\rho c^2 t^{2\rho-1}}{(1-\rho)^2}, \quad \text{o} \quad t = \left(\frac{\bar{\zeta}^2 (1-\rho)^2}{-2\rho c^2} \right)^{\frac{1}{2\rho-1}} n^{\frac{1}{1-2\rho}}$$

Equivalentemente, colocando $t = r = \frac{n}{k}$

$$r = \left\lceil \left[\left(\frac{\bar{\zeta}^2 (1-\rho)^2}{-2\rho c^2} \right)^{\frac{1}{2\rho-1}} n^{-\frac{2\rho}{1-2\rho}} \right] \right\rceil$$

donde $\lceil x \rceil$ significa la parte entera de x . Por lo tanto

$$k_0(n) = \left\lceil \left[\left(\frac{\bar{\zeta}^2 (1-\rho)^2}{-2\rho c} \right)^{\frac{1}{2\rho-1}} n^{-\frac{2\rho}{1-2\rho}} \right] \right\rceil \quad (3.20)$$

Y $k_0(n)$ es la selección óptima para la secuencia $k(n)$ bajo las condiciones dadas.

Ahora se considerara el mínimo de la siguiente expresión $E\left(\hat{\zeta}_H - \zeta\right)^2$ o sustituyéndola por la siguiente

$$E\left(\frac{\zeta N}{\sqrt{k}} + \frac{A\left(\frac{n}{k}\right)}{1-\rho}\right)^2 \quad (3.21)$$

Y la sucesión k_0 que optimiza (3.21) servirá como la selección óptima para el estimador $\widehat{\zeta}_H$ también. Se nota que

$$\lim_{n \rightarrow \infty} \sqrt{k_0} A \left(\frac{n}{k_0} \right) = \lim_{n \rightarrow \infty} cn^\rho k^{\frac{1}{2-\rho}} = \frac{\text{sign}(c)\zeta(1-\rho)}{\sqrt{-2\rho}} \quad (3.22)$$

Donde $\text{sign}(c) = 1$ si $c > 0$ y $\text{sign}(c) = -1$ si $c < 0$. Por lo tanto para esta selección de k se tiene

$$\sqrt{k_0} \left(\widehat{\zeta}_H - \zeta \right) \xrightarrow{d} N \left(\frac{\text{sign}(c)\zeta}{\sqrt{-2\rho}}, \zeta^2 \right)$$

y

$$\lim_{n \rightarrow \infty} k_0 \min E \left(\frac{\zeta N}{\sqrt{k}} + \frac{A \left(\frac{n}{k} \right)}{1-\rho} \right)^2 = \lim_{n \rightarrow \infty} k_0 \min \left(\frac{\zeta^2}{k} + \frac{A^2 \left(\frac{n}{k} \right)}{(1-\rho)^2} \right)$$

por (3.20), la cual es igual

$$\lim_{n \rightarrow \infty} k_0 \left(\frac{\zeta^2}{k} + \frac{A^2 \left(\frac{n}{k} \right)}{(1-\rho)^2} \right) = \zeta^2 + \frac{1}{(1-\rho)^2} \left(\frac{\text{sign}(c)\zeta(1-\rho)}{\sqrt{-2\rho}} \right)^2 = \zeta^2 \left(1 - \frac{1}{2\rho} \right)$$

Por (3.19)

Anteriormente solamente se ha considerado el caso especial $A(t) = ct^\rho$ con $\rho < 0$. Esto es lo que a menudo se asume en aplicaciones de la teoría de valor extremo. Sin embargo esta simplificación no es factible en el caso de que $\rho = 0$. Próximamente se considerará el problema de optimizar en el caso más general de condición de segundo orden no solamente el caso especial $A(t) = ct^\rho$.

Igual sucede en el límite con una sucesión de óptimos, es decir, si $k_0(n)$ representa una sucesión donde cada $k_0(n)$ es un óptimo entonces $k(n) = k_0(n)$ cuando $n \rightarrow \infty$ es también óptimo. Esto implica que se puede reemplazar la función A por cualquier función A^* con $A^*(t) \sim A(t)$, cuando $t \rightarrow \infty$ sin pérdida de generalidad.

Similarmente como antes se puede hallar

$$\arg \min_{t>0} \left(\frac{t\zeta^2}{n} + \frac{A^2(t)}{(1-\rho)^2} \right) \quad (3.23)$$

Donde la función $|A|$ esta variando regularmente con índice $\rho \leq 0$. Para $\rho = 0$ es razonable asumir que existe un función A^* con $|A^*(t)| \sim |A(t)|$ cuando $n \rightarrow \infty$ es

monótona decreciente. En este caso se puede asumir sin pérdida de generalidad que la función A^2 satisface

$$\lim_{t \rightarrow \infty} \frac{A^2(tx) - A^2(t)}{q(t)} = -\log x \quad , \quad x > 0$$

Con q una función positiva, Entonces para cada $\rho \leq 0$ existe una función decreciente positiva $s \in RV_{2\rho-1}$ tal que $n \rightarrow \infty$

$$A^2(t) \int_1^\infty s(u) du \tag{3.24}$$

Se tiene para $c > 1$ y t suficientemente grande

$$\frac{t\tilde{\xi}}{n} + \frac{c^{-1}}{(1-\rho)^2} \int_t^\infty s(u) du < \frac{t\tilde{\xi}^2}{n} + \frac{A^2(t)}{1-\rho^2} \frac{t\tilde{\xi}^2}{n} + \frac{c}{(1-\rho)^2} \int_t^\infty s(u) du \tag{3.25}$$

El ínfimo sobre $t > 0$ para los lados derecho e izquierdo puede calcularse la derivada igualando a cero. Para el lado derecho se tiene

$$\frac{\tilde{\xi}^2(1-\rho)^2}{cn} = s(t)$$

lo cual es equivalente a

$$t = s^{\leftarrow} \left(\frac{\tilde{\xi}^2(1-\rho)^2}{cn} \right)$$

y el ínfimo es

$$\begin{aligned} & \frac{\tilde{\xi}^2}{n} s^{\leftarrow} \left(\frac{\tilde{\xi}^2(1-\rho)^2}{cn} \right) + \frac{c}{(1-\rho)^2} \int_\rho^\infty s(u) du \quad p = s^{\leftarrow} \left(\frac{\tilde{\xi}^2(1-\rho)^2}{cn} \right) \\ & = \frac{c}{(1-\rho)^2} \left\{ \frac{\tilde{\xi}^2(1-\rho)^2}{cn} s^{\leftarrow} \left(\frac{\tilde{\xi}^2(1-\rho)^2}{cn} \right) + \int_\rho^\infty s(u) du \right\} \\ & = \frac{c}{(1-\rho)^2} \int_0^r s(u) du \end{aligned}$$

donde en la última etapa se usó

$$v s^{\leftarrow}(v) + \int_{s^{\leftarrow}(v)}^\infty s(u) du = \int_0^v s^{\leftarrow}(u) du \tag{3.26}$$

Por el lado izquierdo de (3.25) se tiene el mismo resultado pero reemplazando ac por c^{-1} se sigue que el ínfimo (3.23) es

$$\frac{1}{(1-\rho)^2} \int_0^r s^{\leftarrow}(u) du$$

Y se alcanza cuando

$$t \sim s^{\leftarrow} \left(\frac{\xi^2(1-\rho)^2}{n} \right)$$

Es decir, (ya que t se reemplaza por $\frac{n}{k}$)

$$k(n) \sim \frac{n}{s^{\leftarrow} \left(\frac{\xi^2(1-\rho)^2}{n} \right)}$$

Ya que una secuencia óptima $k_0 = k_0(n)$ en el sentido de minimizar

$$\frac{\xi^2}{k} + \frac{A^2 \left(\frac{n}{k} \right)}{(1-\rho)^2}$$

Es dado por

$$k_0 = \left[\frac{n}{s^{\leftarrow} \left(\frac{\xi^2(1-\rho)^2}{n} \right)} \right]$$

¿Qué puede decirse acerca de la distribución asintótica de $\sqrt{k_0}(\widehat{\xi}_H - \xi)$?

$$\sqrt{k_0}(\widehat{\xi}_H - \xi) \xrightarrow{d} \xi N + \frac{\sqrt{k_0} A \left(\frac{n}{k_0} \right)}{1-\rho}$$

Así que se tiene que evaluar $\sqrt{k_0} A \left(\frac{n}{k_0} \right)$ para n grande. Por (3.24) y (3.26)

$$\begin{aligned} k_0 A^2 \left(\frac{n}{k_0} \right) &\sim \frac{n}{s^{\leftarrow} \left(\frac{\xi^2(1-\rho)^2}{n} \right)} \left\{ \int_{\rho}^{\infty} s(u) du \right\} \\ &= \xi^2(1-\rho)^2 \left\{ \int_0^m s^{\leftarrow}(u) du - \frac{\xi^2(1-\rho)}{n} s^{\leftarrow} \left(\frac{\xi^2(1-\rho)^2}{n} \right) \right\} \\ &= \xi^2(1-\rho)^2 \left\{ \frac{\int_0^m s^{\leftarrow}(u) du}{\frac{\xi^2(1-\rho)^2}{n} s^{\leftarrow} \left(\frac{\xi^2(1-\rho)^2}{n} \right)} - 1 \right\} \end{aligned}$$

Ahora por el teorema B.1.5 (ver en Haan y Ferreira (2006)) ya que

$$\left(\frac{1}{s} \right)^{\leftarrow} \in RV_{\frac{1}{(1-2\rho)'}}$$

$$\lim_{t \rightarrow \infty} \frac{\int_0^{\frac{1}{x}} s^{\leftarrow}(u) du}{\frac{1}{x} s^{\leftarrow} \left(\frac{1}{x} \right)} = \lim_{t \rightarrow \infty} \frac{\int_x^{\infty} \left(\frac{1}{s} \right)^{\leftarrow}(u) \frac{du}{u^2}}{\frac{1}{x} \left(\frac{1}{s} \right)^{\leftarrow}(x)} = \frac{1-2\rho}{-2\rho} \quad (3.27)$$

Se debe notar que $s^{\leftarrow} \left(\frac{1}{x} \right) = \left(\frac{1}{s} \right)^{\leftarrow} (x)$. Por lo tanto para $\rho < 0$

$$\lim_{n \rightarrow \infty} k_0 A^2 \left(\frac{n}{k_0} \right) = \xi^2 (1 - \rho)^2 \left(\frac{1 - 2\rho}{-2\rho} - 1 \right) = \frac{\xi^2 (1 - \rho)^2}{-2\rho}$$

Y

$$\sqrt{k_0} (\hat{\xi}_H - \xi) \longrightarrow N \left(\frac{\text{sign}(A)\xi}{\sqrt{-2\rho}}, \xi^2 \right)$$

Para $\rho = 0$ el límite en (3.27) debe ser interpretado como infinito. Esto significa que para $\rho = 0$ minimizando el error cuadrado medio conseguimos una secuencia optima k_0 para el cual

$$\sqrt{k_0} (\hat{\xi}_H - \xi) + b_n \xrightarrow{d} N(0, \xi^2)$$

Donde b_n es una sucesión que varia suavemente tendiendo a más o menos infinito. Esta afirmación no es común para obtener intervalo de confianza para ξ . Todo lo que se puede decir es que

$$\frac{\sqrt{k_0}}{b_n} (\hat{\xi}_H - \xi) \longrightarrow 1$$

Si $\rho = 0$ se toma la sucesión $k(n)$ a bit más pequeño, no se logra una normalidad asintótica. Tomamos $k_\lambda := k_\lambda(n)$ tal que

$$\lim_{n \rightarrow \infty} k_\lambda A^2 \left(\frac{n}{k_\lambda} \right) = \lambda^2 > 0 \tag{3.28}$$

Si escribimos

$$f(t) = \frac{\lambda^2 t}{\int_t^\infty s(u) du}$$

con s como en (3.24). Entonces la expresión anterior es creciente y es RV_1 además se tiene

$$n \sim f \left(\frac{n}{k_\lambda} \right)$$

En contraste con el caso $\rho < 0$ se tiene para la selección óptima

$$n \sim \frac{\xi^2 (1 - \rho)^2}{s \left(\frac{n}{k_0} \right)}$$

Ahora, las funciones

$$f(t) = \frac{\lambda^2 t}{\int_t^\infty s(u) du} \quad \text{y} \quad \frac{1}{s(t)}$$

son ambos RV_1 pero por el teorema B.1.5. (ver en Haan y Ferreira (2006))

$$\lim_{t \rightarrow \infty} s(t)f(t) = 0$$

Propiedades Asintóticas

El estimador de Hill es asintóticamente normal. Esto es, para una muestra independiente e idénticamente distribuida de tamaño n y con un $k = K(n)$ una sucesión intermedia

$$\sqrt{k}(\hat{\xi} - \xi) \approx \sqrt{\text{var}_{\xi}}N + \lambda b_{\xi, \rho}$$

Con N normal estándar, donde las constantes λ , var_{ξ} y $b_{\xi, \rho}$ son conocidos (ver teorema 3.3) para el estimador de Hill.

El estimador de Hill tiene sistemáticamente la varianza asintótica más pequeña en su rango de posibles valores. Los estimadores de Hill y de momentos tienen la varianza asintótica más pequeña para valores positivos de ξ .

Los estimadores de Máxima Verosimilitud y de Hill Negativo tienen la varianza asintótica más pequeña para valores negativos de ξ .

Es más complicado comparar el sesgo de los estimadores, ya que en general el sesgo depende de ambos parámetros ξ y ρ entre otras características de la distribución subyacente.

Para determinar una secuencia óptima $k_0(n)$ se considera la representación

$$\hat{\xi} - \xi \approx \sqrt{\text{var}_{\xi}} \frac{N}{\sqrt{k}} + \frac{\lambda b_{\xi, \rho}}{\sqrt{k}} \approx \sqrt{\text{var}_{\xi}} \frac{N}{\sqrt{k}} + b_{\xi, \rho} A\left(\frac{n}{k}\right)$$

Con $k = k(n)$ una sucesión intermedia tal que $\sqrt{k}A\left(\frac{n}{k}\right) \rightarrow \lambda \in R$ normal estándar. Por simplicidad se supone $A(t) = ct^{\rho}$ para alguna constante c real y $\rho < 0$ buscamos una secuencia k para la cual

$$\frac{\text{var}_{\xi}}{k} + b_{\xi, \rho}^2 c^2 \left(\frac{n}{k}\right)^2 \text{ es mínimo}$$

Se obtiene

$$k_0(n) = \left[\left(\frac{\text{var}_{\xi}}{-2\rho c^2 b_{\xi, \rho}^2} \right)^{\frac{1}{1-2\rho}} n^{-\frac{2\rho}{1-2\rho}} \right]$$

Donde $\llbracket x \rrbracket$ denota la parte entera de x .

Luego para esta selección de k se tiene

$$\sqrt{k_0}(\hat{\xi} - \xi) \xrightarrow{d} N\left(\frac{\text{sign}(c)\sqrt{\text{var}_\xi}}{\sqrt{-2\rho}}, \text{var}_\xi\right)$$

Y

$$\lim_{n \rightarrow \infty} k_0 \text{mín} E\left(\frac{\text{var}_\xi N}{\sqrt{k}} + \frac{A\left(\frac{n}{k}\right)}{1-\rho}\right)^2 = \text{var}_\xi \left(1 - \frac{1}{2\rho}\right)$$

Se observa que con los supuestos datos una comparación de la calidad de los diferentes estimadores se reduce a una comparación de la varianza asintótica.

De la teoría asintótica uno obtiene intervalos de confianza correspondientes. La aproximación más común es asumir

$$\hat{\xi} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\text{var}_\xi}{k}} < \xi < \hat{\xi} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\text{var}_\xi}{k}}$$

Donde var_ξ es la respectiva varianza asintótica con ξ reemplazada por su estimador y $Z_{\frac{\alpha}{2}}$ es el cuantil $1 - \frac{\alpha}{2}$ para la distribución normal estándar.

Si formulamos la hipótesis Nula y alterna se tiene:

$$H_0 : \xi \geq 0 \quad \text{versus} \quad H_a : \xi < 0$$

La hipótesis nula es rechazada si para un nivel de significancia del 5%,

$$\hat{\xi}_P < \frac{-2.96}{\sqrt{k}}$$

para el estimador de Pickands.

3.2 FUNCIÓN DE DISTRIBUCIÓN DE EXCESOS DE LA MUESTRA

A continuación se presenta el procedimiento que se usa para calcular altos cuantiles de una distribución de colas pesadas por encima de una abscisa x . De acuerdo con la definición 21 si X es una v.a. con función de distribución F y punto final x_F , para un número fijo $u \in \mathbb{R}$, con $u < x_F$,

$$F_u(x) := P[X - u \leq x \mid X > u], \quad x \geq 0 \quad (3.29)$$

la cual llamábamos función de distribución de **excesos** por encima del umbral u de la variable aleatoria X con función de distribución $F(\cdot)$; y en la definición

22 teníamos que $e(u) = E(X - u | X > u)$, la cual llamamos **función media de excesos** de X .

En estas, la desigualdad $u < x_f$ garantiza que el evento $\{X > u\}$ ocurra con probabilidad no nula. La función de excesos sobre un umbral, $F_u(\cdot)$ es conocida como la función de distribución de tiempo de *vida residual* en estadística médica. En un contexto de seguros, $F_u(\cdot)$ es conocida como la función de distribución de *exceso de pérdidas*.

La función de distribución de excesos sobre el umbral $F_u(\cdot)$ está relacionada con la propia función de distribución $F(\cdot)$

$$\begin{aligned}
 F_u(x) &= P[X - u \leq x | X > u] = \frac{P[X - u \leq x, X > u]}{P[X > u]} \\
 &= \begin{cases} 0 & \text{si } x < 0 \\ \frac{P[u \leq X \leq x + u]}{P[X > u]} & \text{si } x \geq 0 \end{cases} \\
 &= \begin{cases} 0 & \text{si } x < 0 \\ \frac{F(x + u) - F(u)}{1 - F(u)} & \text{si } x \geq 0 \end{cases}
 \end{aligned}$$

Si se rescata la primera y la última igualdad se puede escribir

$$F_u(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{F(x + u) - F(u)}{1 - F(u)} & \text{si } x \geq 0 \end{cases} \quad (3.30)$$

Y esta relación tiene una gran importancia debido a que se puede interpretar de la siguiente manera: La cola de $F(\cdot)$ a la derecha de u se transforma en $F_u(\cdot)$ siguiendo tres pasos:

- (a) Trasladar hacia la izquierda la cola de $F(\cdot)$ a la derecha de u de tal manera que el punto $(u, F(u))$ coincida con el punto $(0, 0)$.
- (b) Estirar la cola trasladada multiplicando por el factor $\frac{1}{1 - F(u)}$ mayor que uno.
- (c) Para que la función quede definida para todos los valores de x se considera el semieje $(-\infty, 0]$ y así la función resultante después de la transformación es una función de distribución.

Por lo tanto se ha construido la función de distribución $F_u(\cdot)$ a partir de la propia $F(\cdot)$. La idea de este trabajo es estimar la cola de la distribución $F(\cdot)$ usando un procedimiento descrito en los literales **[(1),(2),(3),(4)]** que es en cierto sentido el inverso de el procedimiento **[(a),(b),(c)]** y este procedimiento es el siguiente

La estimación de la cola de una función de distribución $F(\cdot)$ a la derecha de u : Comenzar construyendo $F_u(\cdot)$ la función de distribución de excesos de $F(\cdot)$ por encima del umbral u . Una vez construida aplique el teorema 2.15 para encontrar la distribución de Pareto generalizada (denotada en el teorema por $G_{\xi,\beta(u),0}$) que corresponde a $F_u(\cdot)$ y se transforma esta función siguiendo el procedimiento que se muestra a seguir

- (1) Eliminar la parte de $G_{\xi,\beta(u),0}$ a la izquierda de 0.
- (2) Encoger la gráfica de $G_{\xi,\beta(u),0}$ a la derecha del 0 multiplicando por el valor $1 - F(u)$ el cual es menor que uno.
- (3) Trasladar la parte encogida de tal manera que el punto $(0,0)$ coincida con el punto $(u, F(u))$.
- (4) Por último para obtener una función de distribución por medio de este procedimiento se considera el semieje $(-\infty, u)$.

Con este procedimiento descrito en los pasos **[(1),(2),(3),(4)]** se obtiene que $F_u(\cdot)$ se transforma en la cola de $F(\cdot)$ a la derecha de u , donde $F_u(\cdot)$ es como en (3.30) y $G_{\xi,\beta(u),0}$ se transforma en una distribución muy cerca de la cola de $F(\cdot)$; pues $G_{\xi,\beta(u),0}$ estaba muy próxima de $F_u(\cdot)$. La condición suficiente para que este procedimiento funcione es que la $F(\cdot)$ esté en el dominio de atracción de una de las D.V.E.

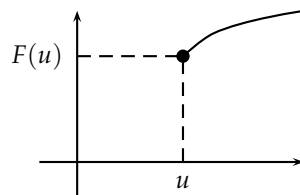


Figura 3.1. Gráfica de la función $G_{\xi,\beta(u),0}^T(\cdot)$ resultante de la transformación **[(1), (2), (3), (4)]** aplicada a $G_{\xi,\beta(u),0}(\cdot)$, que por su parte, es la función proporcionada por el teorema 2.15 y que aproxima a $F_u(\cdot)$. La condición suficiente para que este procedimiento funcione es que la $F(\cdot)$ esté en el dominio de atracción de una de las D.V.E.

Si se denota por $G_{\xi,\beta(u),0}(\cdot)$ a $G_{\xi,\beta(u),0}^T(\cdot)$ después de la transformación [(1), (2), (3),(4)] entonces la expresión de $G_{\xi,\beta(u),0}^T(\cdot)$, este procedimiento es

$$G_{\xi,\beta(u),0}^T(x) = \begin{cases} 0 & x < u \\ F(u) + [1 - F(u)]G_{\xi,\beta(u),0}(x - u) & x \geq u \end{cases}$$

Finalizamos esta sección proporcionando un resumen de los conceptos dados hasta el momento y los cuales son usados en la solución del problema de estimación de altos cuantiles de funciones de distribución $F(\cdot)$ de colas pesadas.

3.3 RESUMEN

Si la cola $F(\cdot)$ es infinita, (recuerde que si existe x tal que $F(x) = 1$ entonces F es de cola finita), entonces:

$$\left. \begin{array}{l} F(\cdot) \in \boxed{\text{Dominio de Atracción de Fréchet}} \quad \Phi_\alpha(\cdot) \text{ con } \alpha > 0 \implies \xi = \alpha^{-1} \\ \text{ó} \\ F(\cdot) \in \boxed{\text{Dominio de Atracción de Gumbel}} \quad \Lambda(\cdot) \implies \xi = 0 \end{array} \right\} \begin{array}{l} \text{En ambos} \\ \text{casos la} \\ G_{\xi,\beta(u),0} \\ \text{tiene cola} \\ \text{infinita} \end{array}$$

Cola $F(\cdot)$	Dominio de Atracción	$G_{\xi,\beta(u),0}$ dado por el T de Pickards	Cola
finita	Weibull $\Psi_\alpha(\cdot)$ $\alpha > 0$	$\xi = -\alpha^{-1}$ $\alpha > 0$	Finita
	Gumbel $\Lambda(\cdot)$	$\xi = 0$	Infinita y de la distribución exponencial

3.4 PROCEDIMIENTO DE INFERENCIA ESTADÍSTICA

En la sección anterior se obtuvo el siguiente resultado: dado un valor u , la cola a la derecha de u , $F_u(\cdot)$ de una distribución $F(\cdot)$, desconocida, se puede aproximar por la función $G_{\xi,\beta(u),0}^T(\cdot)$, determinada por los parámetros u , ξ , $\beta(u)$, $F(u)$ mediante

$$G_{\xi,\beta(u),0}^T(x) = \begin{cases} 0 & x < 0 \\ F(u) + [1 - F(u)]G_{\xi,\beta(u),0}(x - u) & x \geq u \end{cases} \quad (3.31)$$

Sin embargo, surgen preguntas del tipo ¿cómo obtener u ? ó ¿cómo estimar ξ , $\beta(u)$, $F(u)$?. El objetivo de esta sección es proporcionar un método que de respuesta a estas preguntas. Una condición necesaria para que se pueda aplicar el método es que la función de distribución $F(\cdot)$ de la cual conocemos solo una muestra, esté en el dominio de atracción de las EVD's. El procedimiento que se desea realizar se puede sistematizar en dos pasos, a saber:

(a) Encontrar el menor umbral u posible donde los excesos de la muestra encima de u pueden ser aproximados por una GPD con $s = 0$ ($s =$ parámetro de localización); llamaremos u^{opt} el valor óptimo del umbral u .

(b) Entre todas las GPD's con $s = 0$ se escoge la que mejor se adhiera a los excesos de la muestra encima del umbral u^{opt}

3.5 PROCESO GENERAL DE LA APLICACIÓN

Suponga que X, X_1, \dots, X_n son variables aleatoria independientes e idénticamente distribuidas con función de distribución $F(\cdot)$, y $F(\cdot)$ pertenece al máximo dominio de atracción de una GED \mathbf{H} para algún $\zeta \in R$ ver el teorema 2.2.

Primero se fija un umbral u^{opt} y se denota por

$$N_{u^{opt}} = \text{card}\{i : i = 1, \dots, n, X_i > u^{opt}\}$$

al número de excedentes de x_1, \dots, x_n por encima de u^{opt} . Seguidamente se fija $y_1, y_2, \dots, y_{N_{u^{opt}}}$ como el conjunto ordenado de los correspondientes excesos, generado por estos excedentes; ver figura 3.2.

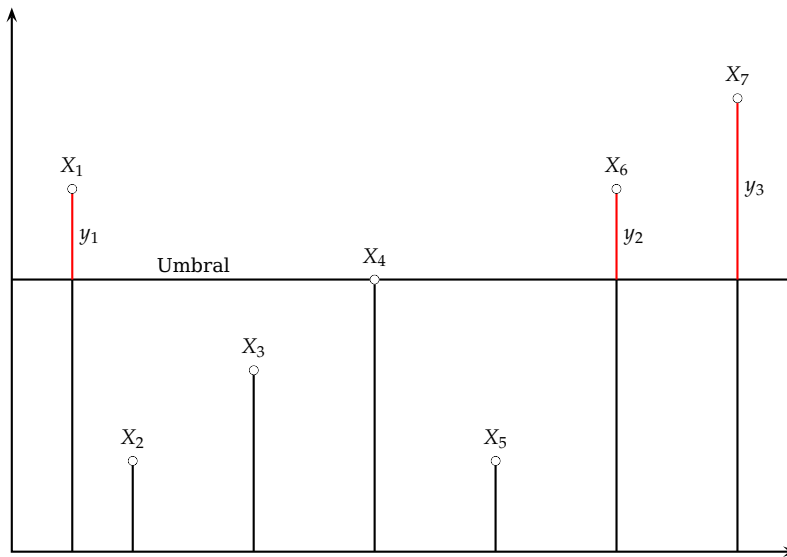


Figura 3.2. Datos $X_1, X_2, X_3, X_4, X_5, X_6, X_7$ y los correspondientes excesos $y_1, \dots, y_{N_{u^{opt}}}$ sobre el umbral u . En este caso $N_{u^{opt}} = 3$.

La función de distribución de excesos de X es dada por

$$F_u(y) = P(X - u \leq y | X > u) = P(Y \leq y | X > u), \quad y \geq 0;$$

ver la definición 21. La última relación se puede escribir como

$$\bar{F}(u + y) = \bar{F}(u)\bar{F}_u(y). \quad (3.32)$$

De acuerdo con la definición 7 de distribución generalizada de Pareto GPD: una GPD $G_{\xi,\beta,0}$ con parámetros ξ , β y S tiene una función de sobrevivencia

$$\bar{G}_{\xi,\beta,0}(x) = \begin{cases} \left(1 + \xi \frac{x}{\beta}\right) & \text{si } \xi \neq 0, \\ e^{-x/\beta} & \text{si } \xi = 0, \end{cases}$$

donde

$$\begin{aligned} x &\in [0, \infty) & \text{si } \xi \geq 0, \\ x &\in [0, -\beta/\xi] & \text{si } \xi \leq 0. \end{aligned}$$

El teorema 2.15 proporciona un resultado para $1 - F_u(y)$ que citamos a continuación

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} \left| \bar{F}_u(x) - \bar{G}_{\xi,\beta(u),0}(x) \right| = 0$$

para una apropiada función positiva $\beta(u)$. Con base en este resultado, para valores grandes de u , la siguiente aproximación cobra sentido

$$1 - F(y) = \bar{F}_u(y) \approx \bar{G}_{\xi,\beta(u),0}(y). \quad (3.33)$$

Note que β está en función del umbral u . En la práctica, u tiene que tomarse suficientemente grande en magnitud. Dado que u , ξ y $\beta = \beta(u)$ son estimados de los datos excedentes (los y_i), entonces los datos estimados dependen de u . Escoger un umbral óptimo u es similar al problema de escoger un número k de orden estadístico para el estimador de Hill. Un valor de u muy alto resulta de pocos excedentes y en consecuencia en estimadores con mucha varianza. Para u muy pequeño los estimadores son sesgados. Un método muy usado en la práctica es basado en la linealidad de la función de la media de excesos para la GPD. Más adelante se discutirá con más profundidad este tema.

La relación (3.32) sugiere un método para estimar la cola de $F(\cdot)$ usando la estimación de $\bar{F}_u(y)$ y $\bar{F}(u)$ por separado. La función de distribución empírica proporciona un estimador natural para $\bar{F}(u)$ este es

$$1 - \hat{F}(u) = \frac{1}{n} \sum_{i=1}^n I_{X_i > u} = \frac{N_{u^{opt}}}{n}$$

donde $I_{X_i > u} = 1$ si $X_i > u$ y cero en caso contrario. Por otro lado, la aproximación por la distribución de Pareto generalizada (3.33) proporciona un estimador de la forma

$$1 - \hat{F}_u(y) = \bar{G}_{\xi,\hat{\beta},0}(y)$$

Para apropiados $\hat{\xi}$ y $\hat{\beta}$ que dependen de $N_{u^{opt}}$. Por lo tanto el estimador que resulta para la cola $\bar{F}(u + y)$ para $y > 0$ es de la forma

$$1 - \hat{F}(u + y) = \frac{N_{u^{opt}}}{n} \left(1 + \hat{\xi} \frac{y}{\hat{\beta}} \right)^{-1/\hat{\xi}} \quad (3.34)$$

En el caso de Fréchet y Gumbel $\xi \geq 0$, el dominio de restricción en (3.34) es $y \geq 0$, obviamente pensando en estimar la cola superior de $F(\cdot)$. Surge inmediatamente un estimador de el cuantil x_p :

$$\hat{x}_p = u^{opt} + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_{u^{opt}}} (1 - p) \right)^{-\hat{\xi}} - 1 \right). \quad (3.35)$$

Para el caso en que $\xi < 0$ una estimación para el punto final del lado derecho x_F de $F(\cdot)$ es dado por

$$\hat{x}_F = u^{opt} - \frac{\hat{\beta}}{\hat{\xi}}.$$

Esta última ecuación se obtiene sustituyendo $\hat{x}_F = \hat{x}_1$ es decir $p = 1$ en la ecuación (3.35). Recordamos que las propiedades estadísticas de los estimadores resultantes dependen de forma crucial de las propiedades en distribución del proceso puntual de excedentes ($N_{u^{opt}}$). La forma de proceder que se mostró en esta sección proviene de la hidrología. Durante los últimos 30 años los profesionales de esta área han desarrollado este procedimiento de estimación bajo el nombre de picos sobre el umbral POT (del inglés *Peak Over Threshold*).

3.6 CÁLCULO DE u^{opt}

Se retoma ahora la discusión sobre como hallar el umbral óptimo u^{opt} a partir del cual será aproximada la cola de la distribución $F(\cdot)$. El método está basado en la linealidad de la función de la media de excesos $e(u)$ para la GPD.

Si el parámetro de forma ξ de la GPD $G_{\xi, \beta, 0}(\cdot)$ satisface la condición

$$\xi < 1$$

entonces la función de la media de excesos de esa GPD es dada por

$$e(u) = \frac{\beta + \xi u}{1 - \xi} \quad \beta + \xi u > 0.$$

De este modo la parte correspondiente a las abcisas positivas de la función de la media de excesos, $e(u)$, $u > 0$, de una GPD $G_{\xi, \beta, 0}(\cdot)$ con la condición $\xi < 1$ pero no negativo, es una función lineal en u con tangente igual a

$$\frac{\xi}{1 - \xi}.$$

Este hecho sobre la tangente de la función de la media de los excesos sugiere una aproximación gráfica para escoger $u > 0$ tal que $\hat{e}(x)$ sea aproximadamente lineal para $x \geq u$.

Ahora se debe pensar en la forma de aproximar la función $e(\cdot)$, un primer acercamiento sugiere que la aproximación sea dada usando la función $\hat{e}(\cdot)$ construida usando la muestra X_1, \dots, X_n mediante la siguiente regla:

$$\hat{e}(u) = \frac{1}{N_{u^{opt}}} \sum_{i=1}^{N_{u^{opt}}} (x_i - u) \quad \text{con} \quad \min\{x_1, \dots, x_n\} < u < \max\{x_1, \dots, x_n\}.$$

Esta función será llamada función de la media muestral de excesos de la muestra x_1, \dots, x_n .

3.7 ESTIMANDO ξ Y β

Una vez que el valor del umbral óptimo se ha determinado, los excesos de la muestra x_1, \dots, x_n por encima de este umbral óptimo se pueden ver como si fuesen generados por una GPD con $S = 0$. Por tanto, el próximo paso es calcular estimaciones de los parámetros ξ y β de la GPD $G_{\xi, \beta, 0}(\cdot)$, que aproxima mejor los excesos de la muestra por encima del umbral óptimo. Recuerde que u^{opt} denota el valor del umbral óptimo y que $\hat{\xi}$ y $\hat{\beta}$, denotan las estimaciones de los parámetros ξ y β respectivamente, $N_{u^{opt}}$ denota el número de excesos y

$$y_1, y_2, \dots, y_{N_{u^{opt}}} \tag{3.36}$$

denotan los excesos ordenados. Se necesita determinar una familia de funciones paramétricas, a saber, las GPD's con $S = 0$ que mejor se ajusten a la muestra (3.36) y para calcular los parámetros de la función que representa a esta muestra, utilizaremos el principio de máxima verosimilitud.

Con base en la muestra (3.36) se construye la función $l(\xi, \beta)$ y se encuentra el punto $(\hat{\xi}, \hat{\beta})$ donde la función $l(\xi, \beta)$ asume su máximo.

$$l(\xi, \beta) = -N(u^{opt}) \log \beta - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^{N(u^{opt})} \log \left(1 + \frac{\xi}{\beta} y_i \right) \tag{3.37}$$

en el dominio

$\beta > 0$ debido a su interpretación

$\xi > 0$ pues esta expresión ya fue construida.

Desafortunadamente no hay una solución analítica para las coordenadas $(\hat{\xi}, \hat{\beta})$, el punto donde la función (3.37) asume el valor máximo. Comúnmente, esto lleva al cálculo aproximado de $(\hat{\xi}, \hat{\beta})$, utilizando métodos de optimización numérica. En el presente caso emplearemos un método implementado en el apéndice A.4 para optimizar una función de dos variables no lineal. Este método tiene un mejor desempeño cuando hace uso de las derivadas parciales de la expresión que será maximizada. En nuestro caso estas derivadas son:

$$\begin{aligned}\partial l(\xi, \beta) / \partial \xi &= \frac{1}{\xi^2} \sum_{i=1}^{N(u^{opt})} \log \left(1 + \frac{\xi}{\beta} y_i \right) - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^{N(u^{opt})} \frac{y_i}{\beta + \xi y_i} \\ \partial l(\xi, \beta) / \partial \beta &= -\frac{N(u^{opt})}{\beta} - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^{N(u^{opt})} \frac{-\beta y_i}{\xi(\beta + \xi y_i)}\end{aligned}$$

En las condiciones en que el punto $(\hat{\xi}, \hat{\beta})$ maximice la función $l(\xi, \beta)$ se dirá que $G_{\hat{\xi}, \hat{\beta}, 0}(\cdot)$ es la función que mejor se adhiere a la muestra (3.36) de acuerdo con el principio de máxima verosimilitud.

3.8 CÁLCULO DE $F(u^{opt})$

Después de recibir de los pasos anteriores el valor de u^{opt} y la función $G_{\hat{\xi}, \hat{\beta}, 0}(\cdot)$, en este paso se construye una aproximación para la cola de $F(\cdot)$ a la derecha de u^{opt} . La función de aproximación $G_{\hat{\xi}, \hat{\beta}, 0}^T(\cdot)$ es dada por la fórmula (3.31). Se observa que en la fórmula (3.31) el único valor que faltaba por determinar era el de $F(u^{opt})$, este valor es estimado mediante la cantidad $(n - N_{u^{opt}})/n$. El cual corresponde a la estimación de $F(u^{opt})$ usando la distribución muestral. Con esto se concluye la ejecución de la parte práctica.

CAPÍTULO 4

CASO DE APLICACIÓN

En esta fase del trabajo se plantean los elementos más relevantes del método propuesto para la estimar altos cuantiles de distribuciones de probabilidad de colas pesadas, estos elementos son:

1. Dada una distribución F , considerar la función de F_u que está dada por $P(X < x \mid u < x)$, la cual tiene la información sobre los valores extremos de la distribución F .
2. Aproximar a F_u con una distribución generalizada de Pareto, y usar esta GPD para hallar los cuantiles y las probabilidades de interés.

Para realizar la tarea descrita en los puntos anteriores se debe enfrentar varias etapas las cuales conllevan un reto estadístico: se desconoce totalmente la función de distribución subyacente F , y este hecho implica que de 1. no se pueda construir la función F_u y de 2. dado que no se puede construir la función F_u se tiene también desconocimiento de la función de distribución generalizada de Pareto.

Recapitulando: para hallar el cuantil de interés se debe tener a la mano la función de distribución generalizada de Pareto que es determinada por los parámetros ξ , β y s . Ahora, en la fórmula para el cálculo de estos parámetros aparecen cantidades como el umbral u^{opt} , es decir, que el primer paso para lograr el acometido de hallar los cuantiles es encontrar el umbral $\bar{F}(u + y) = \bar{F}(u)\bar{F}_u(y)$ al óptimo u^{opt} .

Con el objeto de evaluar el desempeño del método usado en este trabajo, se realizará una prueba siguiendo la siguiente especificación: construir una función de distribución que se llamará $F^{prueba}(x)$ y a partir de ella generar una muestra. Después se produce una aproximación de la cola a la derecha de $F^{prueba}(x)$.

De esta manera se prueba el método, pues en este caso se conoce el valor exacto de las probabilidades que son dadas por la función de distribución de prue-

ba $F^{prueba}(x)$ y estas se comparan con las calculadas usando el método. Estas comparaciones se muestran en las tablas 4.1 y 4.2, más adelante en esta sección volvemos a centrar la atención en los resultados resumidos en estas tablas.

4.1 PRUEBA 1

En esta sección probamos el método suponiendo¹ que la función de distribución de la muestra se encuentra en el dominio de atracción de Fréchet, ya que esta se aplica a las funciones con colas infinitas. La función de distribución que se usará para generar las muestras es

$$F^{prueba}(x) = \begin{cases} \Phi(x) & \text{si } x \leq 1 \\ (1 - \log(x + 1.71)) \times (1 - \Phi(x)) & \text{si } 1 < x < s \\ \left\{ 1 - (1 - d) \left(1 - \left[1 + \xi \frac{x-s}{\beta} \right]^{-1/\xi} \right) - d \right\} & \text{si } x \geq s \end{cases} \quad (4.1)$$

donde a ξ le asignamos el valor de 0.7

donde a β le asignamos el valor de 1.0

donde a s le asignamos el valor de 1.281

donde a d le asignamos el valor de 0.9

y

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy \quad y \in R$$

es la función de distribución estándar.

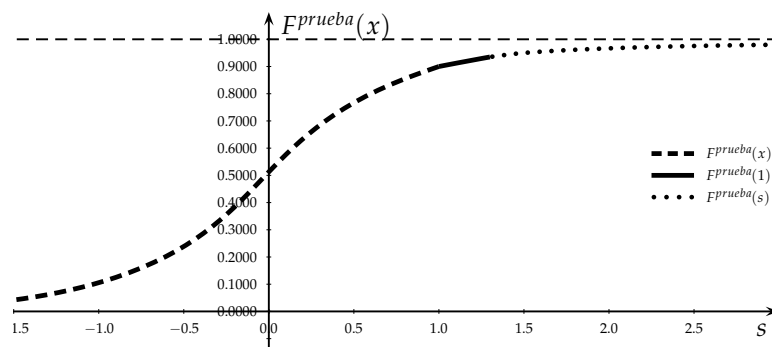


Figura 4.1. Gráfica de (4.1)

¹Recuerde que una condición necesaria para que el método funcione es que la función de distribución de la muestra pertenezca a una de estas tres familias: el dominio de atracción de las distribuciones de Fréchet, el dominio de atracción de las distribuciones de Weibull o el dominio de atracción de la distribución de Gumbel.

Ahora sigue usar estas función de distribución para generar muestras; aplicar a las muestras el método POT con el objetivo de estimar las colas de las respectivas funciones de distribución y, analizar la eficiencia del método, comparando los resultados desde el punto de vista de susceptibilidad de esta a la presencia del ruido.

4.2 GENERANDO LA MUESTRA

Hay una variedad de métodos para generar variables aleatorias. Cada método se aplica solo a un subconjunto de distribuciones y para una distribución en particular un método puede ser más eficiente que otro. A continuación exponemos el método de la transformación inversa.

Lema 4.1. Si la variable aleatoria X tiene una función de distribución acumulada $F(x)$, entonces la variable $r = F(x)$ esta distribuida uniformemente $U(0,1)$. Por lo tanto, X se puede obtener generando números uniformes y calculando $x = F^{-1}(r)$.

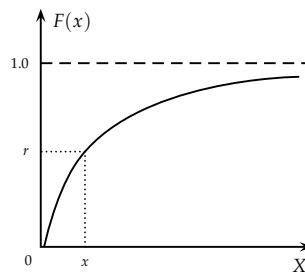


Figura 4.2. Transformación inversa

Demostración. Sea $r = g(x)$ tal que $x = g^{-1}(r)$:

$$F_U(r) = P(U \leq r) = P(X \leq g^{-1}(r)) = F_X(g^{-1}(r))$$

Seleccionemos $g(x) = F_X(x)$, o $r = F_X(x)$, y que r sea una variable aleatoria entre 0 y 1 con distribución dada por

$$F_U(r) = F_X(g^{-1}(r)) = F_X(F_X^{-1}(r)) = r \quad \text{y} \quad f(r) = \frac{dF}{dr} = 1$$

o sea que r está distribuida uniformemente ente 0 y 1. Este método nos permite generar variables aleatorias siempre que se pueda determinar $F^{-1}(x)$ analíticamente o empíricamente. □

Según el lema anterior, podemos obtener el conjunto de valores $F^{-1}(r_1)$, $F^{-1}(r_2)$, \dots , $F^{-1}(r_n)$, que corresponde a una muestra aleatoria generada por la distribución $F(x)$. En el proceso de generación de esta muestra, se debe tener presente que, debemos encontrar $F^{-1}(r)$ dados r y $F(x)$. El problema radica en la ausencia de una expresión analítica simple para la inversa de la función $F^{prueba}(\cdot)$, dada en (4.1). Este hecho nos obligó a calcular $F^{-1}(r_i)$ como solución de la ecuación $F(x) - r_i = 0$, por intermedio de un método numérico apropiado. El código del programa que realiza este método es presentado en la Sub-sección A.2, que el lector puede encontrar en el Apéndice A2.

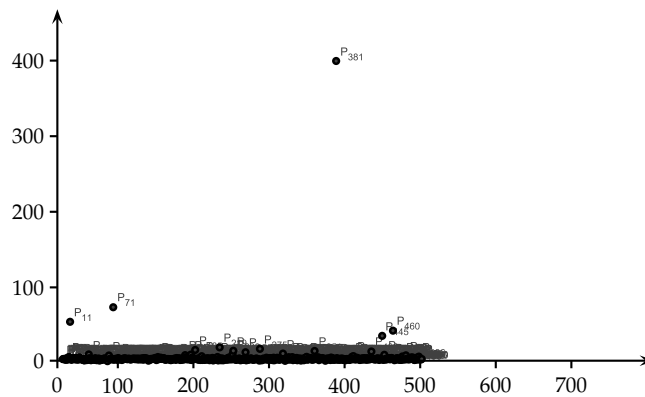


Figura 4.3. Datos obtenidos de la muestra, en el eje de las abscisas se representa el tiempo y el eje de las ordenadas la magnitud de cada dato. Inicialmente se pueden detectar puntos extremos como el punto 11, punto 71 y el punto 381.

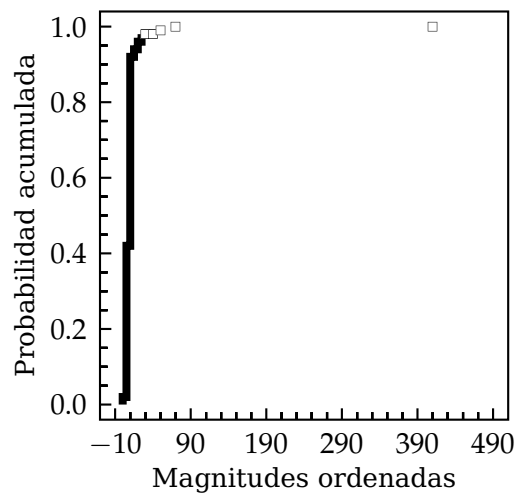


Figura 4.4. Función de distribución muestral o empírica construida a partir de los datos suministrados X_1, \dots, X_{500} .

El método descrito arriba fue utilizado para generar una muestra de $n = 500$ puntos de la función de distribución $F^{prueba}(\cdot)$. Esta muestra es presentada en la Figura 4.3 y su función de distribución muestral en la figura 4.4.

Es interesante notar que la muestra de la función $F^{prueba}(\cdot)$ está estirada a la derecha, una observación natural, que se debe a la presencia de la función de variación lenta, $L(\cdot)^2$, en la construcción de $F^{prueba}(\cdot)$ y al hecho de que $L(\cdot)$ sea creciente y mayor que 1.

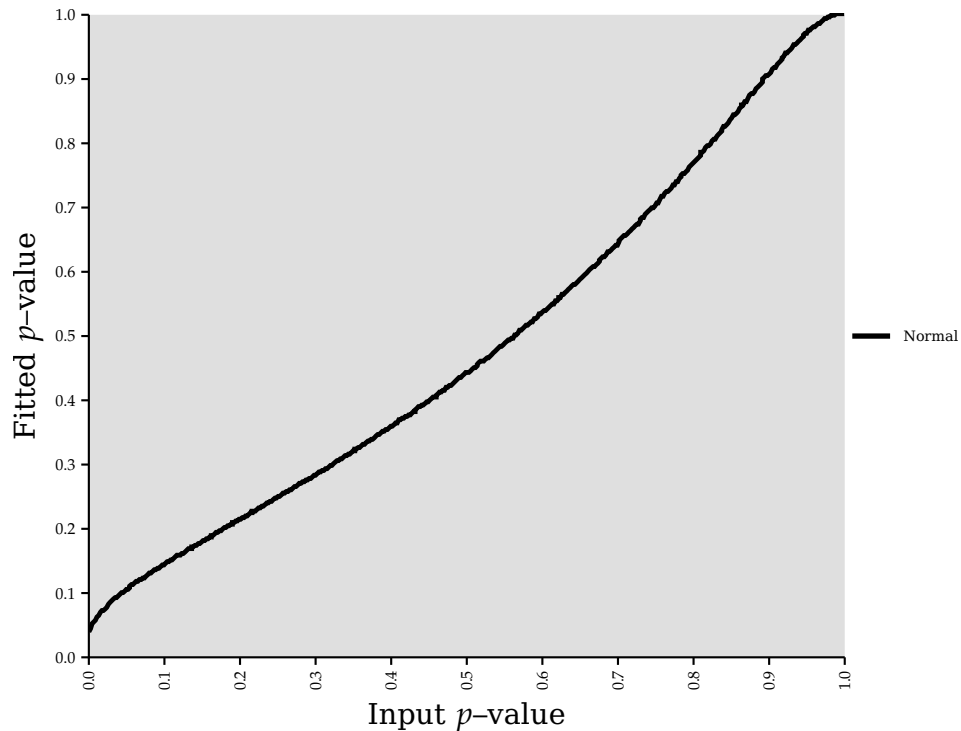


Figura 4.5. Gráfico de probabilidad-probabilidad de los datos de $F^{prueba}(x)$

Este es el gráfico de probabilidad-probabilidad de los datos de $F^{prueba}(x)$ comparados con la distribución normal. No se nota que tenga un ajuste de normalidad. por el contrario se ven datos muy alejados de la línea recta que se debería tener en teoría para que los datos fueran normales.

²Se está haciendo referencia a la función que figura en la definición de $F^{prueba}(\cdot)$ y que está dada por

$$L(x) = \begin{cases} 1 & \text{si } x < 1 \\ \log(x + 1.71) & \text{si } x \geq 1. \end{cases}$$

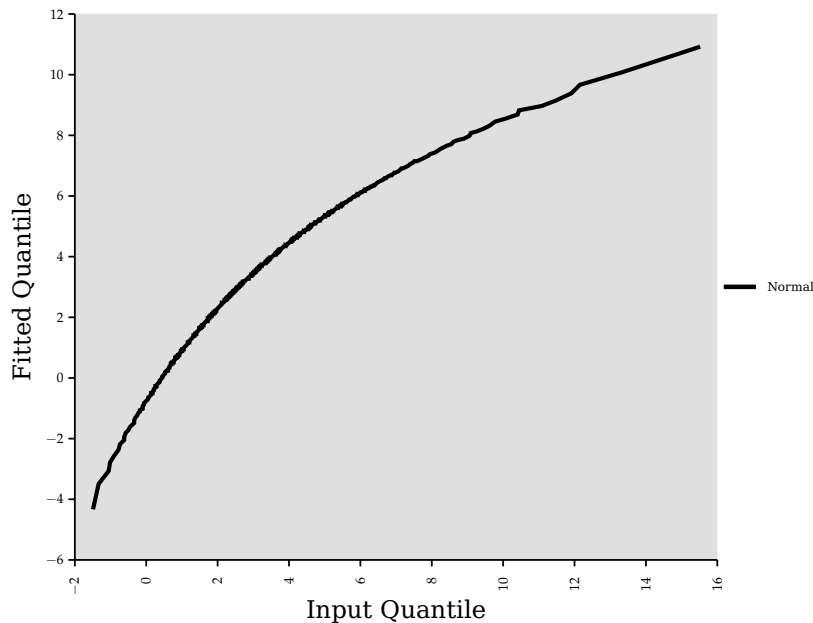


Figura 4.6. Gráfico cuantil-cuantil de los datos de $F_{prueba}(x)$

El gráfico 4.6 corresponde al cuantil-cuantil de los datos de las $F_{prueba}(x)$ comparados con la distribución normal. Se observa que los datos están muy alejados de la línea recta que se debería tener en teoría para que los datos fueran normales. Es decir, los datos no se ajustan a un proceso normal y que posiblemente correspondan a una distribución con colas pesadas.

4.3 APLICACIÓN DEL MÉTODO POT

La aplicación del método POT comienza, conforme la construcción del mismo presentada en la sección 3.4, con el establecimiento de la cantidad denotada por u^{opt} llamada umbral óptimo.

Para esta tarea es necesario construir la función de la media muestral de excesos y analizar su gráfico, buscando identificar el menor dato, para el cual el gráfico sea similar al de una recta con tangente no negativa.

En la Subsección A.3 del Apéndice se presenta el código que produce la función de la media muestral de excesos. El gráfico de esta función, relativo a la muestra de $n = 500$ puntos originados de la función $F_{prueba}(\cdot)$ de (4.1), está en la figura 4.7.

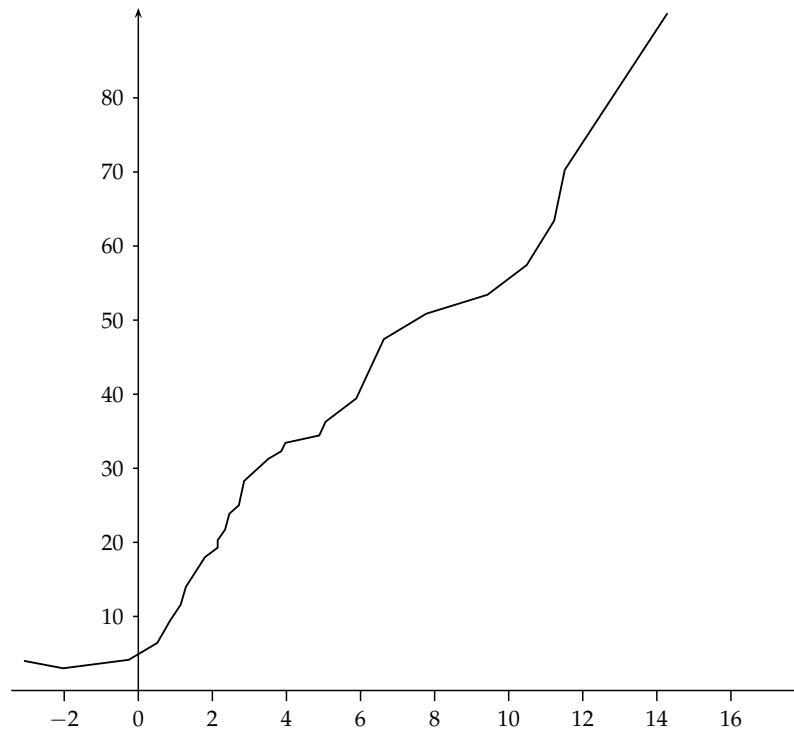


Figura 4.7. Gráfica de la función de distribución muestral o empírica

En esta gráfica se representa la función de distribución muestral o empírica. Construida a partir de los datos suministrados X_1, \dots, X_{500} . Los valores muy próximos de los mayores puntos de la muestra se descartan como candidatos al umbral óptimo. Estos valores no entran en el análisis del comportamiento de la función de la media muestral de excesos.

Se sabe de la sección 3.6 que la función de la media de excesos de una función de distribución debe ser lineal para el caso en que la cola de esta distribución sea Pareto, esto es un hecho rigurosamente probado en Embrecht P. et al. (1997), este aspecto hace que la región de "dientes grandes" sea excluida del análisis de la linealidad, una vez que el tamaño de dientes corresponde a la distancia entre los puntos de la muestra, y donde los puntos de la muestra están expandidos, la aproximación de la función de media de excesos por la función de la media muestral de excesos es pobre.

Con esto en mente en el análisis del gráfico de la figura 4.7 se decide que en el caso de la muestra de que se está tratando con $n = 500$ puntos, retirada de la función $F^{prueba}(\cdot)$, la linealidad comienza en la abscisa 1. Consecuentemente, asumimos que el umbral óptimo, u^{opt} , es 1 para esta muestra de tamaño $n = 500$.

Una forma de determinar el umbral óptimo en casos reales es tomando el punto que separa entre el 10 % y el 15 % de los mayores valores de los datos de los otros valores de la muestra.

Una vez determinado el valor del umbral óptimo, el próximo paso del método POT es calcular estimaciones de los parámetros ξ y β de la GPD $G_{\xi,\beta,0}(\cdot)$, que mejor aproxima los excesos de la muestra por encima del umbral óptimo. Recuerde, que $\hat{\xi}$ y $\hat{\beta}$ denotan, respectivamente, las estimaciones de los parámetros ξ y β . Recordamos también que en el ámbito del presente trabajo estas estimaciones son las proporcionadas por el método de máxima verosimilitud. Como se explica en la sección 3.7, este método atribuye $\hat{\xi}$ y $\hat{\beta}$ los valores que maximizan la expresión (3.37) que fue construida con base en la expresión de la GPD con colas infinitas. La solución de este problema proporciona los resultados $\hat{\xi} = 1.177$ y $\hat{\beta} = 0.942$. Estas estimaciones corresponden a los valores aproximados de los parámetros ξ y $\beta(u)$ de la GPD garantizada por el teorema de 2.15 para aproximar la función de distribución de los excesos de $F(\cdot)$ a partir de la abscisa u^{opt} . La aproximación de la cola producida por el método es dada por $G_{\hat{\xi},\hat{\beta},0}^T(\cdot)$ que se obtuvo de la GPD $G_{\hat{\xi},\hat{\beta},0}(\cdot)$ por medio del procedimiento 2.1 en la sección 3.4.

Por último el objetivo final del problema de estimación de la cola, en la mayoría de los casos, es aproximar los cuantiles $F^{-1}(p)$ de una distribución desconocida $F(\cdot)$ con los cuantiles de $\hat{F}^{-1}(p)$ correspondientes a la aproximación $\hat{F}(\cdot)$ de la cola de $F(\cdot)$. El resultado de la comparación de los valores de $F^{-1}(p)$ con los valores de $\hat{F}_u^{-1}(p)$, donde $\hat{F}_u(\cdot)$ denota la estimación de la cola de $F(\cdot)$ calculada por el método para el valor del umbral óptimo se muestran en la tabla 4.1 donde además se muestra el procedimiento para varios valores tomados para el umbral óptimo.

p	$F^{-1}(p)$	$\hat{F}_{0.8}^{-1}(p)$	$\hat{F}_1^{-1}(p)$	$\hat{F}_{1.2}^{-1}(p)$
0.9	1.42	1.523	1.556	1.537
0.95	3.01	3.019	3.085	3.013
0.995	27.01	41.271	38.172	42.247

Tabla 4.1. Comparación entre $\hat{F}^{-1}(p)$ y $F^{-1}(p)$

En la tabla 4.2 se observa que las estimaciones $\hat{\xi}$ están por encima del verdadero ξ en todos los casos, sin embargo los valores de ξ tienden a ser estables mientras que el umbral crece, lo que no sucede con el parámetro β , para este último no se observa que se torne estable a medida que el valor de umbral óptimo crece, sin

embargo esto no es nada que contradiga la teoría, pues el teorema 2.15 dice que el parámetro β es función de u , $\beta(u)$, dice que β depende de u pero no dice que se tenga que estabilizar e medida que u crece.

Umbral	0.8	1	1.2
$\hat{\xi}$	1.127	1.177	1.219
$\hat{\beta}$	0.538	0.942	1.090
excedentes	104	75	66

Tabla 4.2. Comparación entre $\hat{\xi}$ y ξ

La comparación entre los valores $F^{-1}(p)$ con los de $\hat{F}_u^{-1}(p)$ proporciona razones para concluir que el método funciona adecuadamente en el caso considerado en este trabajo.

4.4 UNA APLICACIÓN A LA DESERCIÓN

Las altas tasas de deserción y bajas tasas de graduación se han convertido en un asunto de creciente interés para las instituciones de educación superior y las autoridades educativas Castaño E. et al. (2004, 2006). Según diversos autores e investigadores siempre se apunta a que los principales causas de deserción responden a factores, institucionales, socioeconómicos, académicos e individuales, entre otros Castaño E. et al. (2004, 2006), *Australian vocational education training statistics, Completion and attrition rates for apprentices and trainees* (2010), Vélez White C. M. et al. (2009), Montero Rojas E. et al. (1996).

En torno a este problema se han desarrollado diferentes modelos y programas para medir la deserción académica. Ahora, bajo la premisa de que los factores que intervienen en la deserción antes mencionados pueden variar, de país a país, de ciudad a ciudad, de una institución educativa a otra, de una facultad a otra, e incluso, de un programa académico a otro, presentamos en esta sección, un modelo versátil, en el cual se consideran explícitamente factores relevantes que ayudan a explicar el fenómeno de deserción.

La versatilidad del modelo consiste en que se utiliza una variable que en este documento llamaremos Z^{Sc} , ya que la asignación de su valor está dado por el peso de los diversos factores que pueden o se quieran analizar sobre el problema de la deserción, en esta caso se utilizó el método de Scoring Gutierrez Girault M. A. (2007), Schreiner M. (2001), para determinar el valor de Z^{Sc} , es decir, el scoring

es un método que permite ponderar el valor de Z^{Sc} según la relación existente entre las variables. A partir de la información suministrada por el estudiante podemos caracterizar la variable Z^{Sc} , en nuestro caso tomamos las variables x_1 : corte examen de admisión, x_2 : estrato y x_3 : escolaridad del segundo hermano, $Z^{Sc} = (x_1, x_2, x_3)$. Un análisis exploratorio de esta variable realizado a 147 alumnos que desertaron en el periodo 2007-I al 2010-II muestra que el supuesto de normalidad no se cumple (ver definición 10), ya que la curtosis es de 16.25, como lo muestran la tabla 4.3 y la figura 4.8.

Estocásticos descriptivos	
Media	0.20842191
Error típico	0.02457157
Mediana	0.12118415
Moda	# N/A
Desviación estándar	0.29791445
Varianza de la muestra	0.08875302
Curtosis	16.2513039
Coefficiente de asimetría	3.81779089
Rango	1.83542323
Mínimo	0.02191058
Máximo	1.85733382
Suma	30.6380209
Cuenta	147

Tabla 4.3. Medidas descriptivas de la variable Z^{Sc}

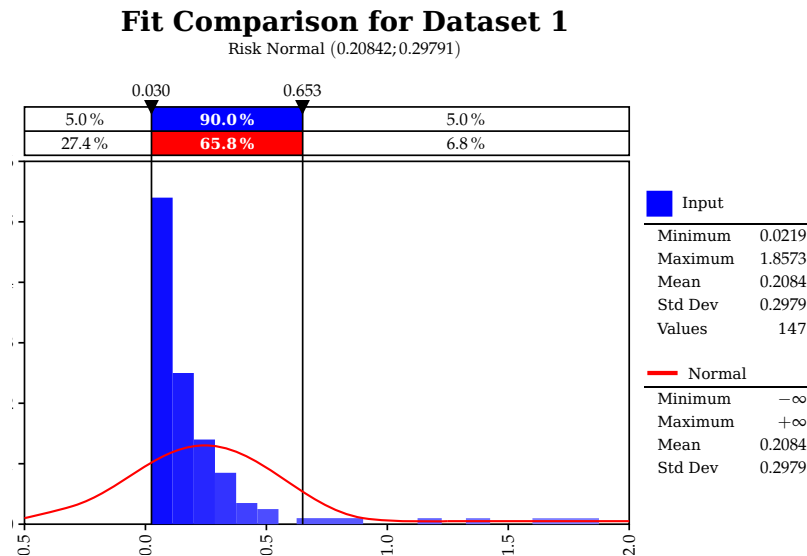


Figura 4.8. Ajuste de normalidad para Z^{Sc}

Se analizaron solo estas tres variables debido a que la finalidad en esta sección es de mostrar cómo funciona el modelo expuesto en este trabajo para predecir la deserción, adviértase al lector que el Z^{Sc} debe ser calculado de acuerdo a las

necesidades requeridas y a las variables que se involucran, mas aun los modelos para predecir las variables involucradas, lo debe exponer el investigador, una buena ayuda la pueden encontrar en (Castaño E. et al. 2004, 2006, *Australian vocational education training statistics, Completion and attrition rates for apprentices and trainees* 2010, Vélez White C. M. et al. 2009, Montero Rojas E. et al. 1996), es decir, para calcular Z^{Sc} podemos a groso modo definir dos etapas, la primera, consiste en determinar el tipo de variables que describen a Z^{Sc} , esto se puede lograr utilizando los diversos métodos estadísticos, regresión multivariada, análisis factorial, análisis estructural de varianza y covarianza, etc. Esta etapa le corresponde al investigador ya que describir las variables que determinan a Z^{Sc} depende de los diferentes entornos en los que cohabita el individuo de estudio, en Castaño E. et al. (2004, 2006), se hace un análisis sobre esta situación.

La segunda etapa consiste en lo que llaman los analistas de riesgo “el Scoring” Gutierrez Girault M. A. (2007), Schreiner M. (2001), procedimiento utilizado en el modelo que estamos exhibiendo, debido a que proporciona los pesos de cada variable que están describiendo a Z^{Sc} .

El método

El Z^{Sc} fue tomado de tal manera que entre mayor sea su valor, mayor es la probabilidad de que haya una deserción, esto es, al estudiante se le pondera el Z^{Sc} de acuerdo a lo descrito anteriormente, ahora, según el valor que toma esta variable, el modelo le puede ayudar a decidir si el estudiante esta en alto riesgo de deserción o no, para articular un proceso preventivo con psicólogos, profesores, padres de familia, etc. Esto de acuerdo a las variables preestablecidas para Z^{Sc} .

Para ajustar la variable Z^{Sc} se tomaron los datos sobre, corte que obtuvo el estudiante en el examen de admisión a la universidad, estrato socio económico, y por medio del análisis estructural de varianza y covarianza se obtuvo la influencia en la deserción de acuerdo a si el segundo hermano tenia estudios universitarios o no, correspondientes a 147 alumnos que desertaron en una ventana de tiempo desde 2007-1 hasta 2010-2.

A continuación presentamos el desarrollo del método expuesto en las secciones 4.2 y 4.3. En la figura 4.9 se muestra el diagrama de dispersión de los datos iniciales para el análisis de los altos cuantiles.

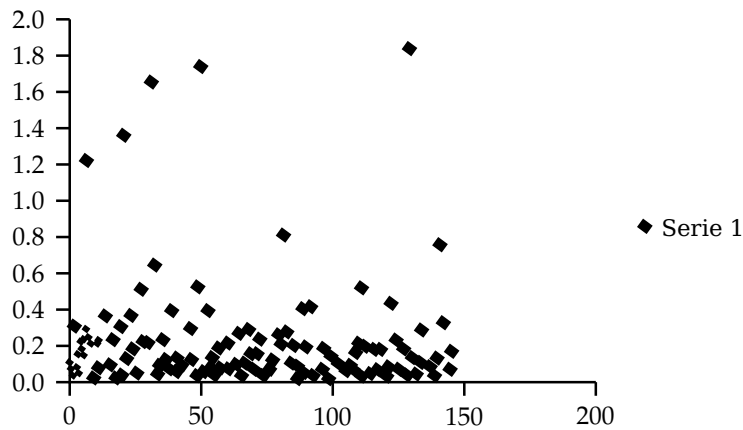


Figura 4.9. Gráfico de dispersión de Z^{Sc}

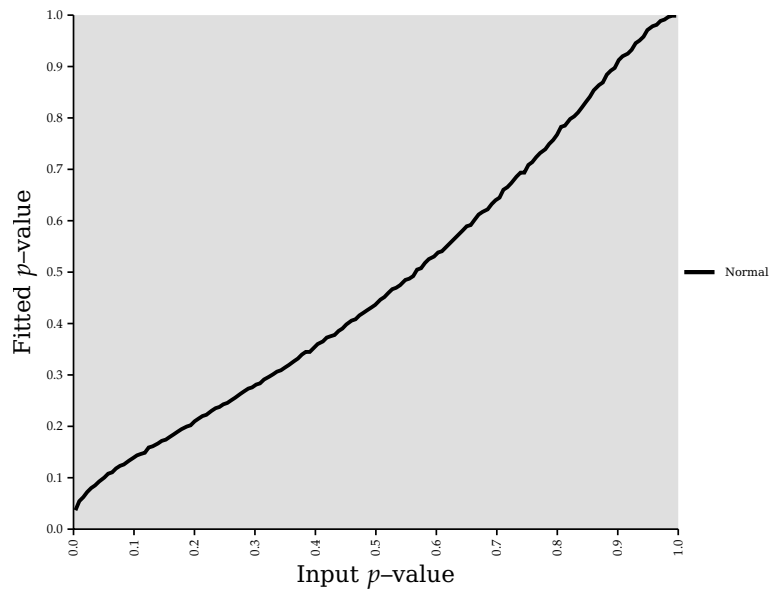


Figura 4.10. Gráfico de probabilidad–probabilidad de los datos de la Z^{Sc}

Los gráficos 4.10 y 4.11 corresponden a probabilidad–probabilidad y quantil–quantil correspondiente a los datos de la Z^{Sc} de la deserción comparados con la distribución normal. Como se mencionó en la sección 4.2 podemos observar que los datos no se ajustan a un proceso normal y que posiblemente corresponda a una distribución con colas pesadas.

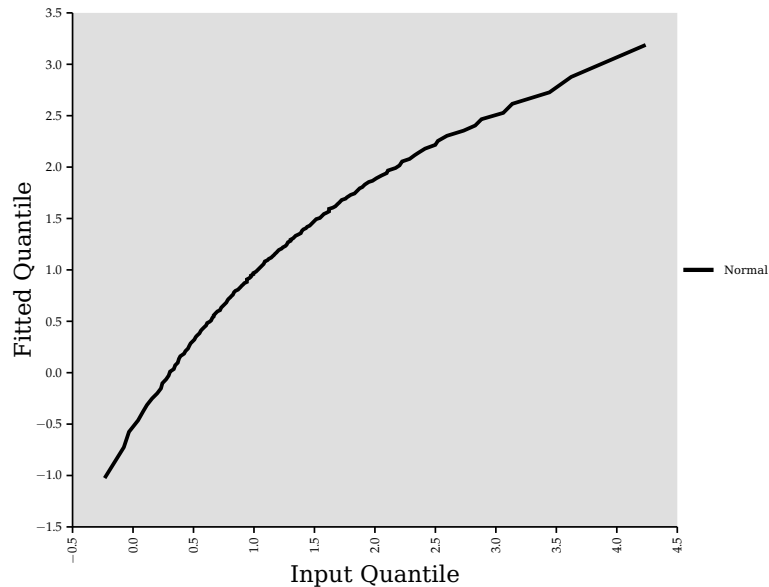


Figura 4.11. Gráfico cuantil-cuantil de los datos de la Z

Para aplicar el método de estimación de altos cuantiles, se comienza con la determinación del umbral óptimo denominado u^{opt} . En la figura 4.5 se muestra la función de excesos de la muestra, exhibida en la sección 4.3.

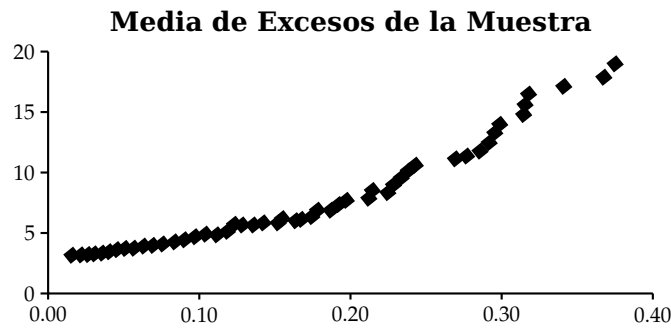


Figura 4.12. Gráfica de la media de excesos de la muestra para 147 datos provenientes de los alumnos que desertaron el primer semestre del programa de Matemáticas de 2007-1 hasta 2010-2.

En este caso el umbral óptimo es $u^{opt} = 0.2380$. Observe que se nota una propiedad de concavidad hasta el punto 1.3 aproximadamente. Este umbral es excedido por 32 datos y representa el percentil 78.

Igualmente a lo que se hizo para el caso de la ilustración del método en la sección 4.2 y 4.3 El paso siguiente consiste en realizar un programa de optimización en dos variables para obtener los parámetros de la distribución generalizada de

Pareto, que servirá como estimación para la cola de la distribución desconocida de los estudiantes que desertaron. Para este caso las estimaciones son $\zeta = 0.64$ y $\beta = 0.81$

Así, que la estimación para el percentil 99,99% a partir de los datos es $X_p = 168$ (ver figura 4.13). Esto significa que la próxima vez que un estudiante obtenga un Z^{Sc} igual o superior a 168, este se encuentra en alto riesgo de que deserte en el primer semestre. Esto nos permite tomar como política que a todos los estudiantes que vayan a ingresar a la Universidad en el programa se les mida el puntaje Z^{Sc} y para aquellos a los que este puntaje quede por encima o mayor se les proporcione un tratamiento especial en cuanto a asesorías y seguimiento continuo para prevenir que abandonen el programa y lograr mejorar los indicadores de la retención estudiantil con calidad.

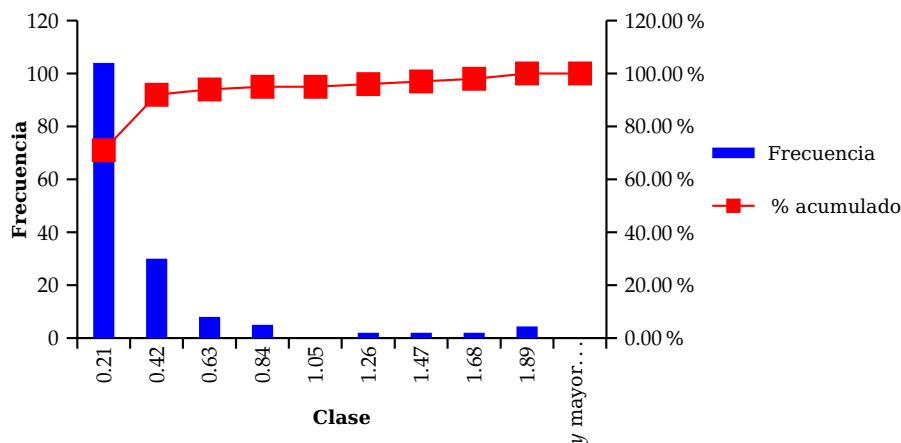


Figura 4.13. Polígono de frecuencia acumulada para la variable Z^{Sc}

CONCLUSIONES

En este trabajo se estudio el problema de estimar altos cuantiles de distribuciones de probabilidad que pertenecen al dominio de atracción de alguna de las distribuciones de valores extremos. El método, denominado picos sobre el umbral POT fue usado en un ejemplo que pretendía mostrar la estrategia de ejecución que brinda esta metodología y luego se aplicó a un conjunto de datos reales.

En ambos casos la determinación del umbral óptimo se baso en la función de la media de excesos de la muestra la cual presentó una concavidad al comienzo y esto permitió decidir que el umbral óptimo no se encontraba en esta región.

En el caso de la función de prueba (4.1) se observa que las estimaciones del parámetro de forma $\hat{\zeta}$ se encuentran por encima del verdadero ζ , pero sin embargo, los valores de ζ tienden a ser estables mientras que el umbral crece, por un cambio relativo de 50 % en u^{opt} se tuvo solo un cambio relativo de 8 % en el parámetro ζ .

Si el umbral óptimo u^{opt} es igual a 0.8 entonces el parámetro de forma ζ es igual a 1.127 y Si el umbral óptimo u^{opt} es igual a 1.2 entonces el parámetro de forma ζ vale 1.219.

Cuando se usó el percentil 0.995, los valores de los cuantiles que proporcionaron las mejores aproximaciones fueron los casos para $\zeta = 1,127$ y para $\zeta = 1,177$ aunque ambos estuvieron por encima del valor verdadero que era de 27,01.

El método también fue usado para ajustar valores extremos a partir de una base de datos provenientes de la deserción estudiantil. En este caso, como en la mayoría de veces que se tienen casos prácticos, no se conoce la función de distribución subyacente, así que el método de picos sobre el umbral se usa como modelo para ajustar distribuciones generalizadas de Pareto con parámetros independientes del tiempo. Para este caso real, la estimación estadística de los parámetros de las distribuciones generalizadas de Pareto condujo a los resultados $\zeta = 0,64$ y

$\beta = 0,81$. Estas estimaciones permitieron sugerir a manera de pronóstico cuando un estudiante abandonará la carrera en el primer semestre de la universidad.

La continuación de este trabajo está enfocada en la vía de las cópulas de valores extremos, investigando la relación existente entre estas cópulas y el método de picos sobre el umbral para dimensiones mayores que uno.

Se estimula a continuar con este estudio porque el modo de estimar altos cuantiles es tanto de importancia en la teoría como en la aplicación a las diferentes empresas del sector real y privado, en la cual se unen conceptos al rededor de lo que se conoce como la toma de decisiones y los riesgos inherentes en cada una de ellas.

Este tipo de análisis proporcionará bases para afirmar si es valioso la generalización a mayores dimensiones o no. El tema de estimar, además de un cuantil particular, un intervalo de confianza para este cuantil también es otra dirección que se avista para las futuras investigaciones.

APÉNDICE

Todos los códigos presentes fueron hechos y rodados por el autor en *Visual Basic* para Aplicaciones contenido en *Excel 2007*.

A.1

```
Public Function funcion_prueba(x, xi, beta, S, d)
%'
' Inputs are x    = abcisa
'              xi = parametro de forma o indice de cola
'              beta = parametro de escala
'              s   = parametro de localizacion
'              d   = sirve en este ejercicio para contraer la funcion
%'
%'
Dim d2
If x <= 1 Then
    funcion_prueba = Application.NormSDist(x)
Else
    d2 = Log(x + 1.71)
    If x < S Then
        funcion_prueba = 1 - d2*(1 - Application.NormSDist(x))
    Else
        funcion_prueba = 1 - d2*(1 - (1 - d)*(1 - (1 + xi*(x - S)/beta)^(-1/xi)) - d)
    End If
End If
End Function
```

A.2

```
Sub Create_Sample_random_Table()
'A Aplica la funcion solver 500 veces
Dim i
    For i = 1 To 500
        Cells(15 + i, 1).Select
```

```
Selection.Copy
Range("A13").Select
ActiveSheet.Paste
Application.CutCopyMode = False
    SolverOk SetCell:=Range("E13"), ByChange:=Range("B13"),_
        MaxMinVal:=3, ValueOf:=0
    SolverSolve UserFinish:=True
    Cells(15 + i, 7) = i
    Cells(15 + i, 8) = Range("B13")
    SolverFinish KeepFinal:=2
Range("A13").Select
Selection.ClearContents
Next
' Clear the range C1:C2
Range("A13:B13").Clear
End Sub
```

A.3

```
Sub Find_Suma_Excesos()
M = 500
For i = 1 To (M - 1)
    Cells(14 + i, 3).Select
    Selection.Copy
    Range("A13").Select
    ActiveSheet.Paste
    Application.CutCopyMode = False
    valor = Range("A13")
    N = M - i
    suma = 0
    For j = (i + 1) To M
        valor_aux = Cells(14 + j, 3)
        'Range("C1").Value = valor_aux
        resta = valor_aux - valor
        'Range("d1").Value = resta
        suma = resta + suma
        valor_aux = 0
    Next j
    Cells(14 + i, 5) = suma
Next i
End Sub
```

Bibliografía

- Arias Pineda G. L. (n.d.), *Modelos de pérdidas agregadas (LDA) y de la teoría del valor extremo para cuantificar el riesgo operativo. Teoría y Aplicaciones*, maestría en matemáticas aplicadas, medellín: eafit edn.
- Australian vocational education training statistics, Completion and attrition rates for apprentices and trainees* (2010), australiano: ncvr edn.
- Beirlant J., Goegebeur Y. & Otros (2004), *Statistics of Extremes: Theory and Applications.*, Wiley.
- Castaño E., Gallón S. & Otros (2004), *Deserción estudiantil universitaria: una aplicación de modelos de duración*, Lecturas de Economía, No. 60, Medellín: Enero–Junio, pp 39–465.
- Castaño E., Gallón S. & Otros (2006), *Análisis de los factores asociados a la deserción y graduación estudiantil universitaria*, Lecturas de Economía, No. 65, julio–diciembre, pp 9–35.
- Coles S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer.
- Davison A. & Smith R. (1990), *Models for Exceedance Over High Thresholds*, J. R. Statistic. Soc. B 52, pp 393-442.
- De Haan, L. & Ferreira, A. (2006), *Extreme value theory: an introduction*, Springer Verlag.
- Embrecht P., Kluppelberg C. & Mikosh T. (1997), *Modelling extremal event, Volume 33 of Applications of Mathematics (New York)*, Berlin: Springer-Verlag.
- Gnedenko B. (1943), *Sur la distribution limite du terme maximum d'une série aléatoire.*
- Goldie C. & Kluppelberg C. (1997), *Subexponential distributions.*

- Gutierrez Girault M. A. (2007), *Modelos de credit scoring —qué, cómo, cuándo y para qué—*.
- Jorion P. (1997), *Value at Risk: The New Benchmark for Controlling Market Risk*, Homewood, IL:Irwin.
- Kotz S. & Nadarajah S. (2000), *Extreme value distributions. Theory and applications*, Imperial College Press.
- Matthys, G., Delafosse, E. & Otros (2004), *Estimating catastrophic quantile levels for heavy-tailed distributions*, *Insurande: Mathematics and Economics* 34(3), 517-537.
- Montero Rojas E., Villalobos Palma J. & Valverde Bermudez A. (1996), *Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la universidad de Costa Rica: un análisis multinivel*, revista electrónica de investigación y evaluación educativa, vol. 13, no. 2, pp. 215-234 edn.
- Morgan Guaranty Trust Company Reuters Ltda. (1996), *Risk Metrics Technical Document*, New York, 4th edition.
- Peng, L. & Qi, Y. (2006), *Confidence regions for high quantiles of a heavy tailed distribution*, *Ann. Statist.* 34(4), 1964-1986.
- Pickands, J. (1971), *The two-dimensional Poisson process and extremal processes*, *J. Appl. Probab.* 8, 745-756.
- Pickands, J. (1975), *Statistical inference using extreme order statistics*, *Ann. Statist.* 3, 119-131.
- Rust, H. W., Kallache, M. & Otros (2006), *Confidence intervals for flood return level estimatives using a bootstrap approach*.
- Salvadori G., De Michele C. & Otros (n.d.), 'Extremes in nature an approach using copulas'.
- Saunders, A. (1996), *Financial Institutions Management: A Modern Perspective*, Homewood, IL:Irwin, 2th edition.
- Schreiner M. (2001), *Ventajas y desventajas del scoring estadístico para las microfinanzas*, *Microfinance risk management y center for social development*, Washington University in St. Louis.

- Sigman, K. (1999), *A Primer on Heavy-Tailed Distributions*, Queueing Syst. Theory Appl., **33**(1-3), 261-275.
- Smith, R. (1989), *Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone*, Statist. Sc. 4, 367-393.
- Smith, R. (2004), *Statistics of extremes with applications in environment, insurance, and finance*, Extreme Values in Finance, Telecommunications, and the Environment, Finkenstadt, B. and Rootzen, H. Eds. CRC/Chapman & Hall,. Boca Raton, Ch. 1.
- Tsay, R. S. (2002), *Analysis of Financial Time Series*, Wiley, 2th edition.
- Van Noortwijk, J. M. (1999), *Quantiles of Generalised gamma distributions from a bayesian point of view*.
- Vélez White C. M., Burgos Mantilla G. & Otros (2009), *Deserción estudiantil en la educación superior colombiana, Metodología de seguimiento, diagnóstico y elementos para su prevención.*, Ministerio de Educación Nacional.

