



UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICA

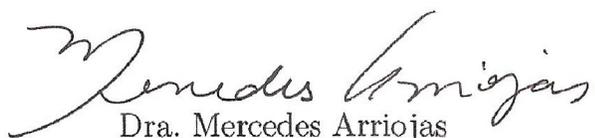
MÉTODO DE VALORES EXTREMOS APLICADO AL RIESGO OPERACIONAL

Trabajo Especial de Grado presentado ante la ilustre Universidad Central de Venezuela por la **Br. Vanessa Parmentier** para optar al título de Licenciada en Matemática.

Tutor: Dra. Mercedes Arriojas.

Caracas, Venezuela
Junio - 2016

Nosotros, los abajo firmantes, designados por la Universidad Central de Venezuela como integrantes del Jurado Examinador del Trabajo Especial de Grado titulado “**Método de Valores Extremos aplicado al Riesgo Operacional**”, presentado por la **Br. Vanessa Parmentier**, titular de la Cédula de Identidad **20.894.486**, certificamos que este trabajo cumple con los requisitos exigidos por nuestra Magna Casa de Estudios para optar al título de **Licenciada en Matemática**.



Dra. Mercedes Arriojas

Tutora



Dr. José Benito Hernández

Jurado



Dra. Elvia Flores

Jurado

Caracas, 06 de Junio del 2016

Agradecimientos

Mi reconocimiento y más sincero agradecimiento va dirigido a aquellas personas que de alguna forma contribuyeron en la realización de este trabajo.

A mi hermana, Susan, por su constante e invaluable apoyo en cada paso de mi vida, por ser mi ejemplo a seguir y por enseñarme que con constancia, organización y mucho esfuerzo se pueden alcanzar las metas que uno se propone.

A mi papá y a mi abuelo, por su incondicional amor y apoyo, por ser mi mayor orgullo e inspiración para lograr todas las metas que me propongo.

A mi mejor amiga, Geika, quien ha sido mi cómplice durante muchos años y me ha alentado a perseguir mis sueños siempre.

A Ricardo por su amor, su constante apoyo y por alegrarme cuando el día no iba bien.

Finalmente, me gustaría agradecerle a todos los profesores de la Escuela de Matemática de la Universidad Central de Venezuela que contribuyeron con mi formación académica, en especial a la Dra. Mercedes Arriojas por su apoyo, comprensión y dedicación y por transmitirme los conocimientos necesarios para la culminación de este trabajo.

Resumen

De diversos estudios de pérdidas por riesgo operacional provenientes de distintos bancos se ha observado que las pérdidas están asociadas a eventos extremos. Estos eventos son estudiados mediante un método de la teoría de valores extremos (EVT) conocido como picos sobre el umbral. Si bien es cierto que la EVT ofrece un conjunto de técnicas estadísticas para el estudio de valores extremos, los resultados tendrán una precisión aceptable sólo cuando los datos satisfagan condiciones específicas. Dos de estas condiciones son que los datos deben ser independientes e idénticamente distribuidos y que se debe contar con suficientes datos para calibrar los modelos. Los datos de pérdidas por riesgo operacional no satisfacen estas condiciones pues generalmente no son independientes y la disponibilidad de los datos es bastante escasa. Esto hace que una aplicación directa de la EVT para el análisis de datos de riesgo operacional sea altamente cuestionable. Debido a esto, en este trabajo se muestra una manera de adaptar la teoría de valores extremos teniendo en consideración la no estacionariedad (dependencia del tiempo) y otras variables (distintos tipos de pérdidas operacionales) para poder hacer un análisis adecuado de los datos.

Palabras claves: Teoría de Valores Extremos; Dominio Máximo de Atracción; Picos sobre el umbral; Distribución generalizada de Pareto; Distribuciones de Valor Extremo; Medidas de Riesgo; Medidas Coherentes de Riesgo; Riesgo Operacional; Valor en Riesgo; Déficit Esperado.

Índice General

Introducción	1
1. Preliminares	4
1.1. Espacio de Probabilidad	4
1.2. Probabilidad Condicional.	6
1.3. Variables Aleatorias	7
1.4. Función de Distribución	9
1.5. Esperanza y Varianza de Variables Aleatorias	13
1.6. Cuantiles o Percentiles	17
2. Medidas de Riesgo	22
2.1. Medidas Coherentes de Riesgo	28
3. Teoría de Valores Extremos. Método POT	33
3.1. Dominio Máximo de Atracción	35
3.2. Distribuciones de Valor Extremo	37
3.3. Método POT y la Distribución Generalizada de Pareto	41
3.4. Estimación por Máxima Verosimilitud	46
3.5. Medidas de Riesgo y el método POT	49
4. Método POT adaptado al modelo no estacionario y covariables	54
4.1. Riesgo Operacional	54
4.2. Método POT avanzado	59
Conclusión	67

Bibliografia

Introducción

En términos de incertidumbre, se puede definir el riesgo como la posibilidad de que ocurran ciertos eventos de interés y las consecuencias derivadas de dichos eventos. En el ámbito financiero, la probabilidad de que ocurran pérdidas financieras originadas por fallas o insuficiencias de procesos, personas, sistemas internos, tecnología y, como consecuencia de eventos externos imprevistos, es lo que se conoce como Riesgo Operacional. Ejemplos de esto son: evasión de impuestos, fraudes, errores de contabilidad, robo de información, desastres naturales, entre otros.

Eventos históricos han demostrado que las pérdidas financieras atribuidas al riesgo operacional pueden ser fatales. Como ejemplo, en 1995 el banco Daiwan sufrió una pérdida de \$1,1 billones debido a especulaciones del corredor Toshihide Iguchi. Por razones como ésta, las instituciones financieras deben gestionar su exposición al riesgo operacional para asegurar su continuidad.

En 1975 se estableció una organización mundial llamada Comité de Supervisión Bancaria de Basilea, con la función de fortalecer la solidez de los sistemas financieros. El comité propuso en 2004 su segundo acuerdo conocido como Basilea II, en el cual se le da importancia al riesgo operacional y se exige un trato explícito para este riesgo, además, se recomienda el empleo de la medida Valor en Riesgo como una herramienta cuantitativa para medir el riesgo operacional, así como también para otras clases de riesgo. En diciembre de 2010 se publica el acuerdo Basilea III, el cual se centra en el riesgo de un fenómeno social conocido como pánico bancario, que ocurre cuando una gran cantidad de clientes de un banco realizan un retiro masivo de sus depósitos bancarios, en un corto período de tiempo, por temor a no poder retirar todo su dinero en el futuro. Este acuerdo complementa los acuerdos Basilea I y Basilea II exigiendo

diferentes niveles de capital requerido para las distintas modalidades de depósitos bancarios y otros préstamos.

De varios estudios de pérdidas por riesgo operacional provenientes de distintos bancos, se obtuvo que en muchos casos las pérdidas alcanzan valores muy altos, muy por encima de los promedios. Estas situaciones se denominan eventos extremos.

Los eventos extremos son, en general, eventos de baja probabilidad que, sin embargo, pueden tener grandes consecuencias. La pérdida que sufrió el banco Daiwan es un ejemplo de evento extremo. La rama de la probabilidad que estudia estos eventos es conocida como Teoría de Valores Extremos (EVT) y uno de los métodos que ofrece esta teoría para cuantificar estos eventos y sus consecuencias es conocido como picos sobre un umbral.

Si bien es cierto que la EVT ofrece un conjunto de técnicas estadísticas para el estudio de valores extremos, los resultados tendrán una precisión aceptable sólo cuando los datos satisfagan condiciones específicas. Dos de estas condiciones son que los datos deben ser independientes e idénticamente distribuidos y que se debe contar con suficientes datos para calibrar los modelos.

Los datos de pérdidas operacionales no satisfacen estas condiciones pues generalmente no son independientes y la disponibilidad de los datos es escasa. Esto hace que una aplicación directa de la EVT para el análisis de datos de riesgo operacional sea altamente cuestionable.

Debido a esto, en este trabajo se muestra una manera de adaptar la teoría de valores extremos teniendo en consideración la no estacionariedad (dependencia del tiempo) y otras variables (distintos tipos de pérdidas operacionales) para poder hacer un análisis adecuado de los datos.

Este trabajo está organizado como sigue:

En el capítulo 1 se introducen aspectos básicos de la teoría de probabilidades que serán necesarios para la comprensión del trabajo.

En el capítulo 2 se definen las medidas valor en riesgo y déficit esperado y se dan algunas de sus propiedades, ejemplos, ventajas y desventajas.

En el capítulo 3 se presenta la teoría de valores extremos y el método picos sobre un umbral, se relacionan las medidas de riesgo con este método y se presenta un ejemplo para comprender la metodología.

En el capítulo 4 se define el riesgo operacional, se presentan algunos ejemplos y se enumeran los tipos de riesgo operacional. Además, se desarrolla el artículo [7], en donde se analiza un grupo de datos formado por tres tipos de pérdidas que parecen depender del tiempo, se adapta el método picos sobre un umbral a estos datos considerando la no estacionariedad y los diferentes tipos de pérdidas y, luego, se presentan y se analizan algunos resultados numéricos referentes a esta adaptación y a las medidas de riesgo de estos datos para cada tipo de pérdida.

Capítulo 1

Preliminares

En este primer capítulo se presentarán algunos conceptos básicos de la Teoría de las Probabilidades que servirán de herramienta para la comprensión de los siguientes capítulos.

1.1. Espacio de Probabilidad

Definición 1.1. Un *experimento aleatorio* es una acción que bajo las mismas condiciones iniciales puede presentar resultados diferentes en cada repetición, es decir, aún cuando se conocen todos los posibles resultados en una realización del experimento, no se puede predecir con exactitud el resultado de la siguiente repetición.

Definición 1.2. El *espacio muestral* asociado a un experimento aleatorio consiste en el conjunto de todos los posibles resultados del experimento. El espacio muestral es denotado por Ω .

Existen dos tipos de espacio muestral: *discreto*, si el espacio es finito o infinito numerable; y *continuo*, si el espacio es infinito no numerable.

Definición 1.3. Un *evento* A asociado a un experimento aleatorio es todo subconjunto del espacio muestral Ω que sea de interés.

Denotaremos a la familia de subconjuntos de Ω formada por los eventos, como \mathcal{F} .

Esta familia de eventos siempre satisface las siguientes propiedades:

- (i) $\Omega \in \mathcal{F}$.
- (ii) Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$.
- (iii) Si $A_n \in \mathcal{F}$, para $n \geq 1$, entonces $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Definición 1.4. Sea Ω un conjunto no vacío. Una σ -álgebra de Ω es una clase o familia \mathcal{F} no vacía de subconjuntos de Ω si verifica las condiciones (i), (ii) y (iii).

Definición 1.5. Sea Ω un espacio muestral y \mathcal{F} una σ -álgebra de Ω . Decimos que una *medida de probabilidad* sobre el espacio muestral Ω es una función $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ tal que satisface los siguientes axiomas:

- (i) $\mathbb{P}(\emptyset) = 0$ y $\mathbb{P}(\Omega) = 1$.
- (ii) \mathbb{P} es no negativa, es decir, $\mathbb{P}(A) \geq 0$, $\forall A \in \mathcal{F}$.
- (iii) \mathbb{P} es σ -aditiva, es decir, si $A_1, A_2, \dots \in \mathcal{F}$ es una colección numerable de eventos disjuntos dos a dos, esto es, $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

La terna $(\Omega, \mathcal{F}, \mathbb{P})$ se denomina *espacio de probabilidad* y el valor $\mathbb{P}(A)$ se denomina *probabilidad de A*.

Proposición 1.1 (Propiedades de la Probabilidad). *Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Si $A, B \in \mathcal{F}$, entonces*

- (i) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
- (ii) Si $A \subseteq B$, entonces $\mathbb{P}(A) \leq \mathbb{P}(B)$.
- (iii) Dados A y B , no necesariamente disjuntos, se cumple que

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

- (iv) Si $\{A_n\}_{n \geq 1} \subseteq \mathcal{F}$ es una sucesión creciente de eventos, es decir, $A_1 \subseteq A_2 \subseteq \dots$, entonces

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

(v) Si $\{A_n\}_{n \geq 1} \subseteq \mathcal{F}$ es una sucesión decreciente de eventos, es decir, $A_1 \supseteq A_2 \supseteq \dots$, entonces

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

1.2. Probabilidad Condicional.

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad.

Definición 1.6. Sea $B \in \mathcal{F}$ tal que $\mathbb{P}(B) > 0$. Se define la *probabilidad condicional dado B* como la función de conjuntos $\mathbb{P}(\cdot|B) : \mathcal{F} \rightarrow [0, 1]$, tal que

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \forall A \in \mathcal{F}.$$

Proposición 1.2. Sea $B \in \mathcal{F}$ tal que $\mathbb{P}(B) > 0$. Entonces la función $\mathbb{P}(\cdot|B)$ es una medida de Probabilidad.

Demostración.

- (i) $\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$
- (ii) $\mathbb{P}(A|B) \geq 0$, para cada $A \in \mathcal{F}$.
- (iii) Sea $\{A_n\}_{n \geq 1} \subseteq \mathcal{F}$ una sucesión de eventos disjuntos dos a dos. Entonces

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \middle| B\right) &= \frac{\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \cap B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup_{n=1}^{\infty} (A_n \cap B)\right)}{\mathbb{P}(B)} \\ &= \sum_{n=1}^{\infty} \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n=1}^{\infty} \mathbb{P}(A_n|B). \end{aligned}$$

□

Proposición 1.3 (Propiedades de la Probabilidad Condicional). Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y sean $A, B \in \mathcal{F}$ tal que $\mathbb{P}(B) > 0$.

- (i) Si A y B son disjuntos, entonces $\mathbb{P}(A|B) = 0$.
- (ii) Si $B \subseteq A$, entonces $\mathbb{P}(A|B) = 1$.

Puede darse la situación en que la ocurrencia de un evento B no afecte de manera alguna la ocurrencia de otro evento A . En este caso, se tiene que la probabilidad condicional de A dado B es igual a la probabilidad de A .

Definición 1.7. Dos eventos A y B son *independientes* si $\mathbb{P}(A|B) = \mathbb{P}(A)$. Esto es, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Proposición 1.4. Si A y B son independientes, entonces A y B^c son independientes y también lo son A^c y B^c .

Definición 1.8. Una familia arbitraria de eventos $\mathcal{C} = \{A_i : i \in I\}$ se dice que es *independiente* si para cualquier colección finita de eventos $A_{i_1}, A_{i_2}, \dots, A_{i_n}$ de \mathcal{C} se cumple que

$$\mathbb{P}\left(\bigcap_{j=1}^n A_{i_j}\right) = \prod_{j=1}^n \mathbb{P}(A_{i_j}).$$

1.3. Variables Aleatorias

Definición 1.9. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Una *variable aleatoria* (real) es una función $X : \Omega \rightarrow \mathbb{R}$ tal que, para todo intervalo $I \subseteq \mathbb{R}$, se tiene que

$$X^{-1}(I) = \{\omega \in \Omega : X(\omega) \in I\} \in \mathcal{F}.$$

Es decir, la preimagen por X de cualquier intervalo es un evento.

Ejemplo 1.1. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y $A \in \mathcal{F}$. Entonces, X definida por

$$X(\omega) = \begin{cases} 1, & \text{si } \omega \in A \\ 0, & \text{si } \omega \in A^c \end{cases}$$

es una variable aleatoria.

Esta función es conocida como *función indicatriz de A* y, usualmente, es denotada por I_A .

Proposición 1.5. Si X e Y son variables aleatorias, entonces $X \pm Y$, kX ($k \in \mathbb{R}$), X/Y ($Y \neq 0$) y XY son también variables aleatorias.

Definición 1.10. Dos variables aleatorias X e Y son *independientes* si y sólo si para conjuntos de Borel cualesquiera $A \subseteq \mathbb{R}$ y $B \subseteq \mathbb{R}$ se cumple que

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Definición 1.11. Decimos que el *rango* R_X de una variable aleatoria X es el conjunto

$$R_X = \{x \in \mathbb{R} \mid \exists \omega \in \Omega : X(\omega) = x\}.$$

Definición 1.12. Decimos que el *cardinal* de un conjunto A , $\text{card}\{A\}$, es la cantidad de elementos del conjunto A .

Definición 1.13. Una variable aleatoria X se dice que es *discreta* cuando su rango R_X es un conjunto discreto. En este caso, R_X es finito ($\text{card}\{R_X\} < \infty$) o a lo sumo numerable. En este caso, existe un conjunto $\{x_n\}_{n \geq 1}$ de valores de X tal que

$$\sum_{n \geq 1} \mathbb{P}(X = x_n) = 1. \quad (1.1)$$

Ejemplo 1.2. Consideremos la variable aleatoria definida en el ejemplo 1.1, entonces $R_X = \{0, 1\}$, por lo que $X(\omega)$ es una variable discreta.

Definición 1.14. Sea X una variable aleatoria discreta tal que el conjunto de valores que toma es el conjunto $\{x_n\}_{n \geq 1}$. Se define la *función de probabilidad de X* ó *función de masa de probabilidad de X* como la función $p : \{x_n\}_{n \geq 1} \rightarrow [0, 1]$ tal que

$$p(x_n) = p_n = \mathbb{P}(X = x_n).$$

Ejemplo 1.3. Consideremos el experimento aleatorio que consiste en el lanzamiento de una moneda n veces donde se desea saber la probabilidad de obtener r caras, $r = 0, 1, \dots, n$. En este caso, el espacio muestral está dado por:

$$\Omega = \{(a_1, a_2, \dots, a_n) : a_i = c \text{ o } a_i = s, \quad i = 1, \dots, n.\},$$

donde c y s representan *cara* y *sello*, respectivamente. Así, el evento *se obtiene una cara en tres lanzamientos* está representado por el conjunto $B = \{(c, s, s), (s, c, s), (s, s, c)\}$.

Tomando $\mathcal{F} = \mathcal{P}(\Omega)$ y un evento $A \subseteq \Omega$, la función $\mathbb{P}(A) = \frac{\text{card}\{A\}}{\text{card}\{\Omega\}}$ define una medida de probabilidad en el espacio $(\Omega, \mathcal{F}, \mathbb{P})$.

Si asociamos a los resultados los valores $c = 1$ y $s = 0$, podemos definir la variable aleatoria $X : \Omega \rightarrow \{0, \dots, n\}$ como

$$X = \# \text{ de caras en } n \text{ lanzamientos.}$$

Así, el evento $B = r \text{ caras en } n \text{ lanzamientos}$, puede expresarse como

$$B = \{\omega \in \Omega : X(\omega) = r\}, \quad r \in \{0, 1, \dots, n\}.$$

En particular, cuando $n=3$, se tiene que el conjunto de posibles resultados es

$$\Omega = \{(ccc), (ccs), (csc), (scc), (css), (scs), (ssc), (sss)\}.$$

Y algunos valores de X son

$$X((ccc)) = 3, \quad X((csc)) = 2, \quad X((sss)) = 0.$$

Luego, si asumimos que la moneda es perfectamente simétrica, tenemos que

$$p(0) = \frac{1}{8}, \quad p(1) = \frac{3}{8}, \quad p(2) = \frac{3}{8}, \quad p(3) = \frac{1}{8}.$$

1.4. Función de Distribución

La función de distribución de una variable aleatoria permite calcular las probabilidades de los eventos de interés determinados por la variable aleatoria.

Definición 1.15. Sea X una variable aleatoria. Se define la *función de distribución* de X como la función $F_X : \mathbb{R} \rightarrow [0, 1]$ tal que para cada $x \in \mathbb{R}$

$$F_X(x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\}) = \mathbb{P}(X \leq x).$$

Si X es una variable aleatoria *discreta* tal que el conjunto de valores que toma es el conjunto $\{x_n\}_{n \geq 1}$. Entonces se define

$$F_X(x) = \sum_{x_n \leq x} p_n,$$

donde $p_n = \mathbb{P}(X = x_n)$.

Ejemplo 1.4. Consideremos la variable aleatoria definida en el ejemplo 1.3, como $X = \#$ de caras en 3 lanzamientos. Entonces, se tiene que:

$$\begin{aligned}
 F(0) = p(0) &= \frac{1}{8}, & F\left(\frac{1}{2}\right) &= p(0) = \frac{1}{8}, \\
 F(1) = p(0) + p(1) &= \frac{1}{2}, & F(3) &= p(0) + p(1) + p(2) + p(3) = 1.
 \end{aligned}$$

Luego, la función de distribución de X está dada de la siguiente forma:

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{8}, & 0 \leq x < 1, \\ \frac{1}{2}, & 1 \leq x < 2, \\ \frac{7}{8}, & 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$$

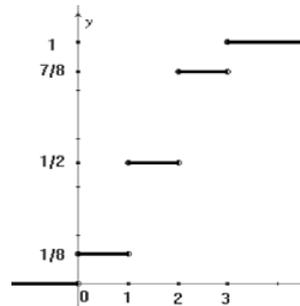


Figura 1.1: Función de Distribución de la v.a $X = \#$ de caras en 3 lanzamientos.

Proposición 1.6 (Propiedades de la Función de Distribución). *Sea F la función de distribución de una variable aleatoria X , entonces:*

- (i) $0 \leq F(t) \leq 1, \quad \forall t \in \mathbb{R}.$
- (ii) $\mathbb{P}(a < x \leq b) = F(b) - F(a), \quad \text{para } a < b.$
- (iii) $F(a) \leq F(b), \text{ si } a < b.$
- (iv) $\lim_{t \rightarrow a^+} F(t) = F(a).$
- (v) $\lim_{t \rightarrow a^-} F(t) = F(a) - \mathbb{P}(X = a).$
- (vi) $\lim_{t \rightarrow \infty} F(t) = 1 \text{ y } \lim_{t \rightarrow -\infty} F(t) = 0.$

Observación: La propiedad (v) indica que para todo $a \in \mathbb{R}$, $\mathbb{P}(X = a)$ equivale al salto de la función de distribución F en el punto a . Por lo tanto, los puntos de continuidad de la distribución F están caracterizados por $\mathbb{P}(X = a) = 0$.

Definición 1.16. Sea X una variable aleatoria con función de distribución F . Se define el *punto final derecho* (o extremo superior) de F , como

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

Definición 1.17. Una función $f : \mathbb{R} \rightarrow \mathbb{R}$ es una *función de densidad* sobre \mathbb{R} si y sólo si f satisface las siguientes condiciones:

- (i) $f(x) \geq 0, \forall x \in \mathbb{R}$.
- (ii) $\int_{-\infty}^{\infty} f(u) du = 1$.

Definición 1.18. Una variable aleatoria X es *continua* si existe una función de densidad f tal que para todo $x \in \mathbb{R}$,

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(u) du.$$

En este caso, f se denomina *función de densidad de probabilidad* de X .

Ejemplo 1.5. Sea X una variable aleatoria. Entonces, los siguientes son ejemplos de distribuciones:

1. Una variable aleatoria X se dice que tiene distribución **normal** o de **Gauss**, con parámetros μ y σ^2 , si su función de densidad de probabilidad está dada por:

$$f(x) = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \forall x \in \mathbb{R}.$$

La función de distribución de una variable aleatoria normal es

$$F(t) = \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^t e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx, \quad \forall t \in \mathbb{R}.$$

Para decir que X está normalmente distribuida con parámetros μ y σ^2 , se emplea la notación $X \sim \mathcal{N}(\mu, \sigma^2)$.

2. Se dice que X tiene distribución **exponencial** con parámetro $\lambda > 0$, si su función de densidad está dada por:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0 \\ 0, & \text{si } x < 0 \end{cases}$$

3. Se dice que X tiene distribución **uniforme** sobre el intervalo $[a, b]$ si su función de densidad está dada por:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{si } a \leq x \leq b \\ 0, & \text{si } x < a \text{ o } x > b \end{cases}$$

4. Sea X una variable aleatoria que toma el conjunto de valores $\{0, 1, 2, \dots\}$. Se dice que X es una variable aleatoria de **Poisson** con parámetro λ , $\lambda > 0$, si la función de probabilidad de X está dada por:

$$p_i = \mathbb{P}(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}, \quad i = 0, 1, 2, \dots$$

Para decir que X tiene distribución de Poisson con parámetro λ , se emplea la notación $X \sim \mathcal{P}(\lambda)$.

5. Sea $\{X_i\}_{i \geq 1}^n$ una sucesión de variables aleatorias iid con función de distribución común F . Sea $M_n = \max(X_1, \dots, X_n)$, $n \geq 1$. Entonces, para $x \in \mathbb{R}$, $n \in \mathbb{N}$, la función de **distribución del Máximo** M_n está dada por:

$$\begin{aligned} \mathbb{P}(M_n \leq x) &= \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \leq x) \\ &= F^n(x). \end{aligned}$$

6. Se dice que X tiene distribución de **Pareto**, con parámetros α y κ , si su función de distribución está dada por:

$$F(x) = \begin{cases} 0, & \text{si } x < \kappa \\ 1 - \left(\frac{\kappa}{x}\right)^\alpha, & \text{si } x \geq \kappa, \end{cases}$$

donde $\kappa > 0$ es el valor mínimo posible de X y $\alpha > 0$.

1.5. Esperanza y Varianza de Variables Aleatorias

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y sea $X : \Omega \rightarrow \mathbb{R}$ una variable aleatoria.

- **Caso discreto:**

Definición 1.19. Sea X una variable aleatoria *discreta* tal que su conjunto de valores es $\{x_n\}_{n \geq 1}$ y su función de probabilidad es $\{p_n\}_{n \geq 1}$. Se define la *esperanza matemática* o valor esperado de X como el número denotado por $E[X]$ y definido como:

$$E[X] = \sum_{n \geq 1} x_n p_n,$$

siempre y cuando la serie converja, en ese caso, la esperanza existe y se dice que la variable aleatoria X tiene esperanza matemática.

Observe que $E[X]$ es un promedio ponderado de los valores de la variable aleatoria X . La esperanza matemática $E[X]$ también se conoce como *media* o *media ponderada* de X .

- **Caso continuo:**

Definición 1.20. Sea X una variable aleatoria *continua* con función de densidad f . Se define la *esperanza matemática* o valor esperado de X como el número denotado por $E[X]$ y definido como

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx,$$

siempre y cuando la integral anterior sea finita, en ese caso, la esperanza existe y se dice que la variable aleatoria X tiene esperanza matemática.

Ejemplo 1.6. Sea X una variable aleatoria de Poisson con parámetro $\lambda > 0$. Entonces, la esperanza matemática de X es:

$$\begin{aligned} E[X] &= \sum_{n=0}^{\infty} n p_n = \sum_{n=0}^{\infty} n e^{-\lambda} \frac{\lambda^n}{n!} \\ &= \sum_{n=1}^{\infty} e^{-\lambda} \frac{\lambda^n}{(n-1)!} \\ &= \lambda \sum_{n=1}^{\infty} e^{-\lambda} \frac{\lambda^{n-1}}{(n-1)!} \end{aligned}$$

$$\begin{aligned}
&= \lambda \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \\
&= \lambda \sum_{n=0}^{\infty} p_n \\
&= \lambda.
\end{aligned}$$

Es de interés medir la dispersión o distribución de los valores de X alrededor de la media. Esto se conoce como varianza.

Definición 1.21. Sea X una variable aleatoria con media $E[X]$. Se define la *varianza* de X , denotada por $Var[X]$, como el número:

$$Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

Proposición 1.7 (Propiedades de $E[X]$ y $Var[X]$). Sean X e Y variables aleatorias y sean $a, b \in \mathbb{R}$ constantes. Entonces se cumple que:

- (i) $E[X + Y] = E[X] + E[Y]$
- (ii) $E[aX + b] = aE[X] + b$
- (iii) $Var[aX] = a^2 Var[X]$
- (iv) $Var[X + b] = Var[X]$

Proposición 1.8. Sean X e Y variables aleatorias independientes. Entonces:

- (i) $E[XY] = E[X]E[Y]$
- (ii) $Var[X + Y] = Var[X] + Var[Y]$

Ejemplo 1.7. Sea X una variable aleatoria de Poisson con parámetro $\lambda > 0$. Entonces, por propiedades de la esperanza, tenemos que la varianza de X es:

$$\begin{aligned}
Var[X] &= E[X^2] - (E[X])^2 \\
&= E[X^2 - X] + E[X] - (E[X])^2 \\
&= \sum_{n=0}^{\infty} (n^2 - n)p_n + \lambda - \lambda^2
\end{aligned}$$

$$\begin{aligned}
&= \sum_{n=0}^{\infty} n(n-1)e^{-\lambda} \frac{\lambda^n}{n!} + \lambda - \lambda^2 \\
&= \sum_{n=2}^{\infty} e^{-\lambda} \frac{\lambda^n}{(n-2)!} + \lambda - \lambda^2 \\
&= \lambda^2 \sum_{n=2}^{\infty} e^{-\lambda} \frac{\lambda^{n-2}}{(n-2)!} + \lambda - \lambda^2 \\
&= \lambda^2 \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} + \lambda - \lambda^2 \\
&= \lambda^2 \sum_{n=0}^{\infty} p_n + \lambda - \lambda^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda.
\end{aligned}$$

Esperanza Condicional

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Sea $B \in \mathcal{F}$, tal que $\mathbb{P}(B) > 0$ y sea $Q = \mathbb{P}(\cdot|B)$ tal que $Q(A) = \mathbb{P}(A|B)$, con $A \in \mathcal{F}$. Luego, como Q es medida de probabilidad sobre Ω , dada una variable aleatoria X , tiene sentido definir la *esperanza condicional* de X dado B como la esperanza de X con respecto a Q .

Antes de definir la esperanza condicional, definamos la función de probabilidad y función de distribución de una variable aleatoria X dado el evento B .

Definición 1.22. Sea X una variable aleatoria discreta y sea B un evento tal que $\mathbb{P}(B) > 0$.

Se define la *función de probabilidad condicional* de X dado el evento B como:

$$p_{X|B}(x|B) = \mathbb{P}(X = x|B) = \frac{\mathbb{P}(\{X = x\} \cap B)}{\mathbb{P}(B)}.$$

Definición 1.23. Sea X una variable aleatoria y sea B un evento tal que $\mathbb{P}(B) > 0$. Se define la *función de distribución de probabilidad condicional* de X dado el evento B como:

$$F_{X|B}(x|B) = \mathbb{P}(X \leq x|B) = \frac{\mathbb{P}(\{X \leq x\} \cap B)}{\mathbb{P}(B)}.$$

Definición 1.24. Sea X una variable aleatoria con función de distribución F y sea B un evento tal que $\mathbb{P}(B) > 0$.

Se define la *esperanza condicional* de X dado el evento B como:

$$E[X|B] = E_Q[X] = \int x d(Q),$$

donde $Q(\{X \leq x\}) = F_{X|B}(x|B)$.

Sea X una variable aleatoria discreta con función de probabilidad $\{p_i\}_{i \geq 1}$. Sea $A_i = \{X = x_i\}$, entonces $p_i = \mathbb{P}(A_i)$ y $\Omega = \bigcup_{i \geq 1} A_i$, donde $A_i \cap A_j = \emptyset$ para $i \neq j$.

Si $Q(\{X = x_i\}) = p_{X|B}(x_i|B)$, se tiene que

$$E[X|B] = E_Q[X] = \sum_{i \geq 1} x_i Q(A_i) = \sum_{i \geq 1} x_i \frac{P(A_i \cap B)}{P(B)}, \quad (1.2)$$

siempre y cuando las series converjan.

Por otra parte, si se considera $Y = XI_B$, se tiene que Y toma valores $\{x_i\}$ y

$$\{w \mid Y(w) = x_i\} = \{w \mid X(w) = x_i\} \cap B = A_i \cap B.$$

Luego,

$$E[XI_B] = E[Y] = \sum_{i \geq 1} x_i P(A_i \cap B),$$

así

$$E[X|B] = \frac{E[XI_B]}{P(B)}. \quad (1.3)$$

De manera análoga se puede ver que, si X es una variable aleatoria continua, su esperanza condicional dada el evento B también satisface la fórmula (1.3).

Proposición 1.9 (Propiedades de la Esperanza Condicional). *Sean X e Y variables aleatorias, sean $a, b, c \in \mathbb{R}$ constantes y $B \in \mathcal{F}$. Entonces se cumple que:*

- (i) $E[X] = E[E[X|Y]]$.
- (ii) Si $Y = \{\emptyset, \Omega\}$, entonces $E[X|Y] = E[X]$, en casi todas partes.
- (iii) Si $X \geq 0$ casi siempre, entonces $E[X|B] \geq 0$, en casi todas partes.
- (iv) $E[aX + bY + c|B] = aE[X|B] + bE[Y|B] + c$.

1.6. Cuantiles o Percentiles

Definición 1.25. Sea X una variable aleatoria y $\alpha \in [0, 1]$. Decimos que $q \in \mathbb{R}$ es un α -cuantil ó α -percentil de X si se cumple que

$$\mathbb{P}(X < q) \leq \alpha \leq \mathbb{P}(X \leq q),$$

o, de manera equivalente,

$$\lim_{x \rightarrow q^-} F(x) = F(q^-) \leq \alpha \leq F(q), \quad (1.4)$$

donde F es la función de distribución de X .

Observación: De las desigualdades (1.4), tenemos que q es un α -cuantil si y sólo si

$$\alpha \in [F(q^-), F(q)].$$

Algunos casos:

1. Si F es una función *continua y estrictamente creciente*, tenemos que para todo $\alpha \in (0, 1)$, existe $q(\alpha) \in \mathbb{R}$ tal que

$$F(q(\alpha)) = \alpha,$$

además, $q(\alpha)$ es único, pues F es inyectiva. Luego,

$$\mathbb{P}(X < q(\alpha)) = \mathbb{P}(X \leq q(\alpha)) = F(q(\alpha)) = \alpha.$$

Por lo tanto, $q(\alpha)$ es el único α -cuantil. (Ver figura 1.2)

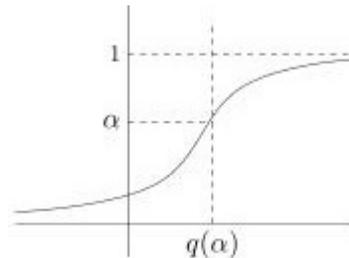


Figura 1.2: α -cuantil de X cuando F_X es continua y estrictamente creciente.

2. Cuando F es una función *continua* pero *no estrictamente creciente*, tenemos que si $\alpha_1 = F(q)$ para todo $q \in [q_1, q_2]$, entonces

$$F(q^-) = F(q) = \alpha_1, \quad \text{para todo } q \in [q_1, q_2].$$

Por lo tanto, α_1 satisface

$$\mathbb{P}(X < q) \leq \alpha_1 \leq \mathbb{P}(X \leq q), \quad \text{para todo } q \in [q_1, q_2]$$

y, en consecuencia, todo $q \in [q_1, q_2]$ es α_1 -cuantil. (Ver figura 1.3)

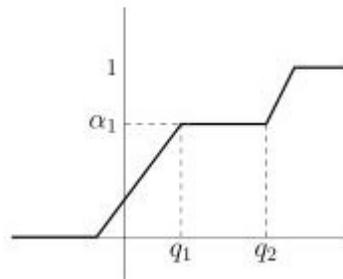


Figura 1.3: α_1 -cuantil de X cuando F_X es continua y no estrictamente creciente.

3. Si F es una función *discontinua* en q_1 , tal que q_1 es α_1 -cuantil, entonces

$$\mathbb{P}(X < q_1) \leq \alpha_1 \leq \mathbb{P}(X \leq q_1),$$

esto es

$$\lim_{x \rightarrow q_1^-} F(x) \leq \alpha_1 \leq F(q_1) = \lim_{x \rightarrow q_1^+} F(x),$$

es decir, el valor α_1 se encuentra en el intervalo del salto de discontinuidad de F en q_1 . (Ver figura 1.4)

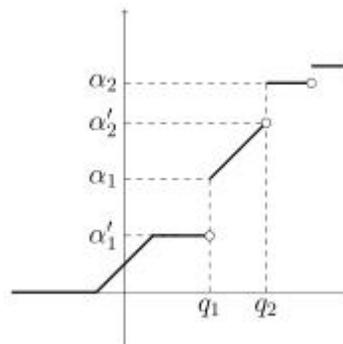


Figura 1.4: α -cuantil de X cuando F_X es discontinua en q_1 .

Definición 1.26. Sea X una variable aleatoria y $\alpha \in [0, 1]$.

Se define el *menor α -cuantil* de X como

$$q_\alpha(x) = q^-(\alpha) = \inf\{x : F(x) \geq \alpha\}.$$

Se define el *mayor α -cuantil* de X como

$$q^\alpha(x) = q^+(\alpha) = \inf\{x : F(x) > \alpha\}.$$

Al intervalo $[q_\alpha(x), q^\alpha(x)]$ se le conoce como *intervalo intercuantílico*.

Proposición 1.10. Dado $s \in \mathbb{R}$

- (i) $q^+(\alpha) \geq s$ si y sólo si $F(s^-) \leq \alpha$
- (ii) $q^-(\alpha) \leq s$ si y sólo si $F(s) \geq \alpha$

Demostración.

(i) (\Rightarrow)

$$\begin{aligned} q^+(\alpha) \geq s &\Leftrightarrow \inf\{x : F(x) > \alpha\} \geq s \\ &\Leftrightarrow F(s) \leq \alpha \\ &\Rightarrow F(s^-) \leq \alpha, \quad \text{pues } F(s^-) \leq F(s). \end{aligned}$$

(\Leftarrow)

$$\begin{aligned} F(s^-) \leq \alpha &\Rightarrow s < \inf\{x : F(x^-) > \alpha\} \\ &\Rightarrow s \leq \inf\{x : F(x) > \alpha\} \\ &\Rightarrow s \leq q^+(\alpha). \end{aligned}$$

- (ii) $q^-(\alpha) \leq s \Leftrightarrow \inf\{x : F(x) \geq \alpha\} \leq s$
 $\Leftrightarrow F(s) \geq \alpha$

□

Por la proposición anterior, se tiene que

$$F(q^-) \leq \alpha \leq F(q) \Leftrightarrow q_\alpha(x) \leq q \leq q^\alpha(x).$$

Es decir, el rango de los posibles valores del α -cuantil q es el intervalo intercuantílico, $[q_\alpha(x), q^\alpha(x)]$.

Definición 1.27. Sea $\alpha \in [0, 1]$ Se define la *función cuantil* asociada a una distribución F como $q : [0, 1] \rightarrow \mathbb{R}$ tal que

$$q(\alpha) = \inf\{x : F(x) \geq \alpha\} = q_\alpha(x).$$

En general, $q(\alpha)$ se denomina *inversa generalizada* de F y se denota también por F^{-1} .

Observación: Si F es biyectiva, entonces F^{-1} es la función inversa de F .

Proposición 1.11. Sea $q(\alpha)$ la función cuantil asociada a F ($q(\alpha) = F^{-1}$). Entonces $q(\alpha)$ es no decreciente y continua a la izquierda.

Demostración. Veamos que $q(\alpha)$ es no decreciente.

Sean $\alpha_1, \alpha_2 \in [0, 1]$ tales que $\alpha_1 < \alpha_2$. Como F es creciente, se tiene que

$$\{x : F(x) \geq \alpha_2\} \subseteq \{x : F(x) \geq \alpha_1\},$$

de donde

$$q(\alpha_1) \leq q(\alpha_2).$$

La demostración de que $q(\alpha)$ es continua por la izquierda se puede ver en [10, pág. 4].

□

Observación: Cuando la función de distribución F es estrictamente creciente y continua, se tiene que $F(q(\alpha)) = \alpha$, como se vió en el caso (1). Esto implica que se pueden calcular los cuantiles a partir de la función de densidad, pues para $\alpha \in (0, 1)$ se tiene que

$$F(q(\alpha)) = \int_{-\infty}^{q(\alpha)} f(u) du = \alpha.$$

Ejemplo 1.8. Sea F una función de distribución exponencial con parámetro $\lambda > 0$. Entonces, como F es estrictamente creciente y continua, se tiene que

$$\alpha = \int_0^{q(\alpha)} \lambda e^{-\lambda u} du$$

$$\begin{aligned} &= \lambda \left(-\frac{e^{-\lambda u}}{\lambda} \right) \Big|_0^{q(\alpha)} \\ &= \lambda \left(-\frac{e^{-\lambda q(\alpha)}}{\lambda} + \frac{1}{\lambda} \right) \\ &= -e^{-\lambda q(\alpha)} + 1, \end{aligned}$$

de donde

$$1 - \alpha = e^{-\lambda q(\alpha)},$$

por lo tanto, la función cuantil asociada a F viene dada por

$$q(\alpha) = -\frac{\ln(1 - \alpha)}{\lambda}.$$

Capítulo 2

Medidas de Riesgo

La noción de riesgo surge ante la necesidad de definir aquellas situaciones que puedan perjudicar a personas, compañías o entidades. Se puede asociar el riesgo a la incertidumbre de obtener una consecuencia negativa de cierta acción.

En 1987, el mercado Dow Jones de Estados Unidos perdió el 22,68 % de su valor equivalente a 500 billones de dólares, se puede decir que se trataba de un riesgo con poca probabilidad de ocurrencia pero de gran impacto negativo para el mercado, además de ser un evento imprevisto.

Las crisis generadas por eventos como este, no afectan únicamente a los grandes grupos económicos; algunos de los efectos inmediatos son el incremento desmedido del desempleo y de los costos de bienes y servicios, entre otros. Así como este evento, hay otros eventos extremos que pueden afectar de manera negativa a individuos o compañías. Todo esto ha sido motivo para que gobiernos e instituciones en todo el mundo reconozcan la importancia de identificar, cuantificar y manejar el riesgo.

En 1975, los presidentes de 11 bancos centrales establecieron una organización mundial llamada Comité de Supervisión Bancaria de Basilea, cuya función es fortalecer la solidez de los sistemas financieros. El comité formula las normas generales de supervisión y directrices con la esperanza de que las autoridades de países miembros y no miembros las implementen.

El acuerdo de Basilea I, publicado en 1988, trata de un conjunto de recomendaciones para establecer un capital mínimo adicional que debía tener una entidad bancaria para

cubrir posibles pérdidas en función de los riesgos que afrontaba, financieros u operativos. Este capital es conocido como capital mínimo requerido. Para el cumplimiento de las normas establecidas en este acuerdo, es necesario poder cuantificar el riesgo. Para esto, se han venido utilizando principalmente dos medidas de riesgo: el Valor en Riesgo y el Déficit Esperado.

El *riesgo financiero* se puede definir como todo evento o acción que pueda representar una pérdida para un individuo o corporación. El riesgo que toma un inversionista al vender un activo, o una compañía de seguros al vender una póliza, son ejemplos de riesgos financieros.

Las 4 clasificaciones de riesgo financiero más usadas son: riesgo de mercado, riesgo de crédito, riesgo de liquidez y riesgo operacional. Éste último es el tipo de riesgo de interés para este trabajo.

El Comité de Basilea define el *riesgo operacional* como el riesgo de un cambio de valor causado por el hecho de que las pérdidas reales difieren de las pérdidas esperadas, debido a inadecuados procesos internos, errores de individuos, fallas en los sistemas, o por eventos externos. [5]

Antes de cuantificar y manejar el riesgo, se definirán algunos conceptos:

Un *portafolio* es una colección de instrumentos financieros: acciones, bonos, cuentas bancarias, inmuebles, entre otros; en los cuales una persona o compañía decide invertir su dinero.

El *valor de un portafolio* en el instante $t \geq 0$, se denota por $V(t)$.

La *posición* es el tipo de compromiso adquirido en el mercado. Se dice que la posición es larga si el inversionista adquiere el compromiso de comprar. Se dice que la posición es corta si el inversionista adquiere el compromiso de vender.

El *retorno* es una medida del cambio de valor de un portafolio durante un período de tiempo $[0, t]$, y está dado por

$$R(t) = \frac{V(t) - V(0)}{V(0)} \quad \text{ó} \quad R(t) = V(t) - V(0).$$

La *pérdida del portafolio* en el período $[s, s + h]$, está dada por

$$L_{[s,s+h]} := -(V(s + h) - V(s)).$$

Se conoce como *Distribución de Pérdida* a la distribución asociada a la pérdida del portafolio $L_{[s,s+h]}$.

Se conoce como *Distribución de Ganancia-Pérdida* a la distribución de

$$-L_{[s,s+h]} := V(s + h) - V(s).$$

Dado que los eventos riesgosos están asociados a la incertidumbre, se estudia el problema de medir y cuantificar el riesgo financiero en el contexto de la teoría de las probabilidades, mediante un espacio de probabilidades, $(\Omega, \mathcal{F}, \mathbb{P})$. Las cantidades asociadas a los precios de los activos serán variables aleatorias sobre este espacio.

En este contexto, se puede definir el riesgo como sigue:

Definición 2.1. Se define como *riesgo* a la variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ que representa el valor neto de un portafolio, su rendimiento o la posible pérdida en un período de tiempo dado.

Se denotará al conjunto de todos los riesgos por

$$\mathcal{R} = \{X : \Omega \rightarrow \mathbb{R}\}.$$

Definición 2.2. Una *medida de riesgo* es cualquier función

$$\rho : \mathcal{R} \rightarrow \mathbb{R}.$$

En adelante, se interpretará X como las pérdidas de un portafolio en un período de tiempo; y, $\rho(X)$ como la cantidad de capital que debe ser agregado a una posición con la pérdida dada por X . Nótese que esta interpretación de X difiere de la interpretación de Artzner (1999,[3]), donde la variable aleatoria X es interpretada como el retorno de una posición o portafolio. Esto lleva a algunos cambios de signo en la discusión de algunos axiomas en comparación con otras presentaciones en la literatura, los cuales se comentarán en la siguiente sección.

Una de las medidas de riesgo más usada es el *Valor en Riesgo* denotada como VaR , por sus siglas en inglés. Se puede describir el VaR como la máxima pérdida en que se puede incurrir durante un período determinado de tiempo, tal que la probabilidad de tener una pérdida mayor sea muy pequeña.

Definición 2.3. Sea X un riesgo y $\alpha \in (0, 1)$, se define el *Valor en Riesgo* de X a un nivel de confianza α , como

$$\begin{aligned} VaR_{\alpha}(X) &= \inf\{x : P(X > x) \leq 1 - \alpha\} \\ &= \inf\{x : P(X \leq x) \geq \alpha\} \\ &= q_{\alpha}(X). \end{aligned}$$

Es decir, $VaR_{\alpha}(X)$ es el α -cuantil de la distribución de X .

Observación. Por lo general, α toma valores cercanos a 1.

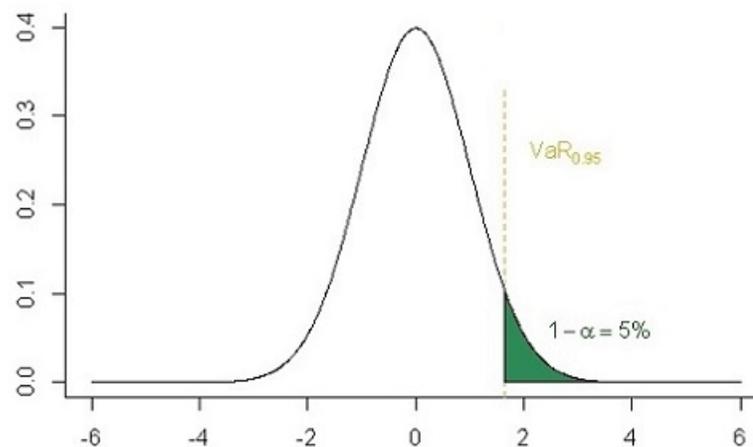


Figura 2.1: Representación gráfica del VaR con un nivel de confianza del 95 %.

Supongamos que la variable que representa las pérdidas de un portafolio sigue una distribución normal estándar, entonces los valores por encima de $VaR_{0,95} = 1,65$ representan las pérdidas posibles en el 5 % de los casos, área sombreada de la figura (2.1). Esto es, existe un 5 % de probabilidad de incurrir en una pérdida mayor a 1.65.

Por ejemplo, si el valor en riesgo diario de un portafolio a un nivel de confianza del 95 % es de 1 millón de dólares ($VaR_{0,95} = 1,000,000$), entonces existe un 5 % de

probabilidad de incurrir en una pérdida mayor a 1 millón, es decir, se estima que en 5 de cada 100 días (o 1 de cada 20 días) la pérdida podría ser mayor a 1 millón de dólares.

Proposición 2.1 (Propiedades del VaR). *Sean X, Y dos riesgos dados que representan pérdidas y sea $\alpha \in (0, 1)$. Entonces se cumple lo siguiente:*

- (i) Si $X \leq Y$, entonces $VaR_\alpha(X) \leq VaR_\alpha(Y)$. (Monotonía)
- (ii) Para $t \geq 0$, entonces $VaR_\alpha(tX) = tVaR_\alpha(X)$. (Homogeneidad positiva)
- (iii) Para $t \in \mathbb{R}$, $VaR_\alpha(X + t) = VaR_\alpha(X) + t$. (Invarianza por traslaciones)
- (iv) $VaR_\alpha(0) = 0$.
- (v) $X \geq 0$, entonces $VaR_\alpha(X) \leq 0$. (Relevancia)

Demostración.

- (i) Si $X \leq Y$, entonces $\mathbb{P}(X \leq t) \geq \mathbb{P}(Y \leq t)$ para todo $t \in \mathbb{R}$, lo que implica que

$$\{t \in \mathbb{R} : \mathbb{P}(Y \leq t) \geq \alpha\} \subseteq \{t \in \mathbb{R} : \mathbb{P}(X \leq t) \geq \alpha\},$$

de donde

$$\inf\{t \in \mathbb{R} : \mathbb{P}(X \leq t) \geq \alpha\} \leq \inf\{t \in \mathbb{R} : \mathbb{P}(Y \leq t) \geq \alpha\},$$

y así

$$VaR_\alpha(X) \leq VaR_\alpha(Y).$$

- (ii) Sea $t > 0$, entonces

$$\begin{aligned} VaR_\alpha(tX) &= \inf\{x : \mathbb{P}(tX \leq x) \geq \alpha\}, \\ &= \inf\{x : \mathbb{P}\left(X \leq \frac{x}{t}\right) \geq \alpha\}, \quad \text{si } ty = x, \text{ entonces} \\ &= \inf\{ty : \mathbb{P}(X \leq y) \geq \alpha\}, \\ &= t \inf\{y : \mathbb{P}(X \leq y) \geq \alpha\}, \\ &= tVaR_\alpha(X). \end{aligned}$$

- (iii) Sea $t \in \mathbb{R}$, entonces

$$\begin{aligned} VaR_\alpha(X + t) &= \inf\{x : \mathbb{P}(X + t \leq x) \geq \alpha\}, \\ &= \inf\{x : \mathbb{P}(X \leq x - t) \geq \alpha\}, \quad \text{si } y + t = x, \text{ entonces} \end{aligned}$$

$$\begin{aligned}
&= \inf\{y + t : \mathbb{P}(X \leq y) \geq \alpha\}, \\
&= \inf\{y : \mathbb{P}(X \leq y) \geq \alpha\} + t, \\
&= VaR_\alpha(X) + t.
\end{aligned}$$

(iv) Para $t \in (0, 1)$, se tiene que

$$tVaR_\alpha(0) = VaR_\alpha(t0) = VaR_\alpha((1-t)0) = (1-t)VaR_\alpha(0),$$

por lo tanto, $VaR_\alpha(0) = 0$.

(v) Si $X \geq 0$, entonces $VaR_\alpha(X) \geq VaR_\alpha(0) = 0$ □

Ejemplo 2.1. Sea X un riesgo. Si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces la distribución de X es creciente y estrictamente continua, así

$$\alpha = \mathbb{P}(X \leq q_\alpha),$$

de donde

$$\begin{aligned}
\alpha &= \mathbb{P}(X \leq VaR_\alpha(X)) \\
&= \mathbb{P}\left(\frac{X - \mu}{\sigma} \leq \frac{VaR_\alpha(X) - \mu}{\sigma}\right),
\end{aligned}$$

si Φ^{-1} es la función de distribución normal $(0, 1)$, entonces

$$\Phi^{-1}(\alpha) = \frac{VaR_\alpha(X) - \mu}{\sigma}$$

es decir

$$VaR_\alpha(X) = \sigma\Phi^{-1}(\alpha) + \mu.$$

Algunas ventajas que caracterizan al VaR son:

- Es una medida de riesgo universal, ya que puede ser aplicado a cualquier tipo de activo.
- Es simple, es decir, posee una fácil interpretación.
- Es completo, pues resume en un sólo número todas las posibles fuentes de riesgo financiero existentes en un portafolio.

2.1. Medidas Coherentes de Riesgo

Ante la inquietud acerca de qué propiedades debería tener un estadístico o medida que, aplicado al riesgo de un portafolio, pueda ser considerado como una medida de riesgo sensible a la realidad, se definen las *medidas coherentes de riesgo*.

Definición 2.4. Una medida de riesgo ρ se dice *coherente* si, para $X, Y \in \mathcal{R}$, satisface las siguientes propiedades:

- (i) Si $X \leq Y$, entonces $\rho(X) \leq \rho(Y)$. (Monotonía)
- (ii) Para todo $t \geq 0$, $\rho(tX) = t\rho(X)$. (Homogeneidad positiva)
- (iii) Para todo $t \in \mathbb{R}$, $\rho(X + t) = \rho(X) + t$. (Invarianza por traslaciones)
- (iv) $\rho(X + Y) \leq \rho(X) + \rho(Y)$. (Subaditividad)

Los axiomas anteriores se pueden interpretar de una manera más intuitiva como sigue:

- Monotonía: posiciones que conducen a mayores pérdidas, requieren más capital de riesgo.
- Homogeneidad positiva: el riesgo de un portafolio es proporcional a su tamaño.
- Invarianza por traslaciones: añadir (o restar) una cantidad t a una posición que conduce a la pérdida, aumenta (o disminuye) el capital requerido en t .
- Subaditividad: la diversificación del portafolio disminuye el riesgo.

El *VaR* no es una medida coherente, pues no cumple la propiedad de subaditividad. Consideremos el siguiente contraejemplo:

Ejemplo 2.2. Supongamos que se tienen dos bonos idénticos, X_1 y X_2 , que caen en mora con probabilidad 4%, es decir el 4% de las veces el deudor se retrasa en su pago, obteniéndose una pérdida de 100 cuando esto ocurre, y de 0 cuando no. Esto es, para $i = 1, 2$,

$$\mathbb{P}(X_i = 100) = 0,04,$$

$$\mathbb{P}(X_i = 0) = 0,96.$$

Entonces, el 95 % VaR de los bonos es

$$VaR_{0,95}(X_i) = \inf\{x : \mathbb{P}(X_i \leq x) \geq 0,95\} = 0.$$

Por otra parte, suponiendo que los bonos son independientes, se tiene que

$$\begin{aligned} \mathbb{P}(X_1 + X_2 = 200) &= \mathbb{P}(X_1 = 100)\mathbb{P}(X_2 = 100) \\ &= 0,04^2 = 0,016, \\ \mathbb{P}(X_1 + X_2 = 100) &= \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 100) + \mathbb{P}(X_1 = 100)\mathbb{P}(X_2 = 0) \\ &= 2 * 0,96 * 0,04 = 0,0768, \\ \mathbb{P}(X_1 + X_2 = 0) &= \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 0) \\ &= 0,96^2 = 0,9216. \end{aligned}$$

Luego,

$$\begin{aligned} \mathbb{P}(X_1 + X_2 \leq 0) &= 0,9216, \\ \mathbb{P}(X_1 + X_2 \leq 100) &= 0,9216 + 0,0768 = 0,9984, \\ \mathbb{P}(X_1 + X_2 \leq 200) &= 1. \end{aligned}$$

Por lo tanto, se tiene que el 95 % VaR de la suma de los bonos es

$$VaR_{0,95}(X_1 + X_2) = \inf\{x : \mathbb{P}(X_1 + X_2 \leq x) \geq 0,95\} = 100.$$

Así,

$$VaR_{0,95}(X_1 + X_2) = 100 > 0 = VaR_{0,95}(X_1) + VaR_{0,95}(X_2).$$

La propiedad de subaditividad es importante, ésta indica que un portafolio formado por sub-portafolios tendrá un riesgo no mayor a la suma de los riesgos de los sub-portafolios que lo componen. Es decir, agregar riesgos individuales no debe aumentar el riesgo global.

Debido a esto, y a otras desventajas del VaR , como la falta de información sobre lo que podría ocurrir si la pérdida supera el VaR y que no describe la peor pérdida posible, proporcionando información sobre el tamaño y la frecuencia de ésta, se propusieron algunas medidas de riesgo alternativas. Una de estas medidas alternativas es el *Déficit Esperado* (ES, por sus siglas en inglés), el cual mide las posibles pérdidas que excedan el VaR .

Definición 2.5. Sea X un riesgo y $\alpha \in (0, 1)$, se define el *Déficit Esperado* a nivel α de X como

$$ES_\alpha(X) = E[X|X > VaR_\alpha(X)].$$

Ejemplo 2.3. Sea X un riesgo. Si $X \sim \mathcal{N}(0, 1)$, entonces por el ejemplo (2.1), sabemos que:

$$VaR_\alpha(X) = \Phi^{-1}(\alpha).$$

Así, usando la fórmula de la esperanza condicional dada por (1.3), se tiene que

$$\begin{aligned} ES_\alpha(X) &= E[X|X > \Phi^{-1}(\alpha)] \\ &= \int_{-\infty}^{\infty} y \phi_{X|X > \Phi^{-1}(\alpha)}(y) dy \\ &= \int_{-\infty}^{\infty} y \frac{\phi(y) I_{\{X > \Phi^{-1}(\alpha)\}}}{\mathbb{P}(X > \Phi^{-1}(\alpha))} dy \\ &= \int_{\Phi^{-1}(\alpha)}^{\infty} y \frac{\phi(y)}{1 - \alpha} dy = \frac{1}{1 - \alpha} \left[-\phi(y) \right]_{\Phi^{-1}(\alpha)}^{\infty} \\ &= \frac{1}{1 - \alpha} \phi(\Phi^{-1}(\alpha)), \end{aligned}$$

donde ϕ es la densidad normal $(0, 1)$ y Φ la distribución normal $(0, 1)$.

Proposición 2.2. Sea X un riesgo y sea $\alpha \in (0, 1)$. Entonces $ES_\alpha(X)$ es una medida coherente de riesgo.

Demostración. Sean X, Y dos riesgos dados y $\alpha \in (0, 1)$. Por medio de las propiedades de la esperanza y la fórmula de la esperanza condicional (1.3) se puede demostrar lo siguiente:

(i) Monotonía: Si $X \leq Y$,

$$\begin{aligned} ES_\alpha(Y) &= E[Y|Y > VaR_\alpha(Y)] \\ &= VaR_\alpha(Y) + E[Y - VaR_\alpha(Y)|Y > VaR_\alpha(Y)] \\ &= VaR_\alpha(Y) + \frac{E[(Y - VaR_\alpha(Y)) I_{\{Y - VaR_\alpha(Y) > 0\}}]}{\mathbb{P}(Y - VaR_\alpha(Y) > 0)} \end{aligned}$$

$$\begin{aligned}
&= VaR_\alpha(Y) + \frac{E[(Y - VaR_\alpha(Y))I_{\{Y - VaR_\alpha(Y) > 0\}}I_{\{X - VaR_\alpha(X) > 0\}}]}{\mathbb{P}(Y - VaR_\alpha(Y) > 0)} \\
&\quad + \frac{E[(Y - VaR_\alpha(Y))I_{\{Y - VaR_\alpha(Y) > 0\}}I_{\{X - VaR_\alpha(X) \leq 0\}}]}{\mathbb{P}(Y - VaR_\alpha(Y) > 0)} \\
&\geq VaR_\alpha(Y) + \frac{E[(Y - VaR_\alpha(Y))I_{\{Y - VaR_\alpha(Y) > 0\}}I_{\{X - VaR_\alpha(X) > 0\}}]}{\mathbb{P}(Y - VaR_\alpha(Y) > 0)} \\
&\geq VaR_\alpha(Y) + \frac{E[(Y - VaR_\alpha(Y))I_{\{X - VaR_\alpha(X) > 0\}}]}{\mathbb{P}(X - VaR_\alpha(X) > 0)} \\
&= VaR_\alpha(Y) + E[(Y - VaR_\alpha(Y)) | X - VaR_\alpha(X) > 0] \\
&= E[Y | X - VaR_\alpha(X) > 0] \\
&\geq E[X | X - VaR_\alpha(X) > 0] \\
&= ES_\alpha(X).
\end{aligned}$$

(ii) Homogeneidad positiva: Sea $t > 0$, entonces

$$\begin{aligned}
ES_\alpha(tX) &= E[tX | tX > VaR_\alpha(tX)] \\
&= E[tX | tX > tVaR_\alpha(X)] \\
&= E[tX | X > VaR_\alpha(X)] \\
&= tE[X | X > VaR_\alpha(X)] \\
&= tES_\alpha(X).
\end{aligned}$$

(iii) Invarianza por traslación: Sea $t \in \mathbb{R}$, entonces

$$\begin{aligned}
ES_\alpha(X + t) &= E[X + t | X + t > VaR_\alpha(X + t)] \\
&= E[X + t | X + t > VaR_\alpha(X) + t] \\
&= E[X + t | X > VaR_\alpha(X)] \\
&= E[X | X > VaR_\alpha(X)] + t \\
&= ES_\alpha(X) + t.
\end{aligned}$$

(iv) Subaditividad:

$$\begin{aligned}
ES_\alpha(X) &= E[X | X > VaR_\alpha(X)] \\
&= VaR_\alpha(X) + E[X - VaR_\alpha(X) | X > VaR_\alpha(X)] \\
&= VaR_\alpha(X) + \frac{E[(X - VaR_\alpha(X))I_{\{X - VaR_\alpha(X) > 0\}}]}{\mathbb{P}(X - VaR_\alpha(X) > 0)}
\end{aligned}$$

$$\begin{aligned}
&= VaR_\alpha(X) + \frac{E[(X - VaR_\alpha(X))I_{\{X - VaR_\alpha(X) > 0\}}I_{\{X+Y - VaR_\alpha(X+Y) > 0\}}]}{\mathbb{P}(X - VaR_\alpha(X) > 0)} \\
&\quad + \frac{E[(X - VaR_\alpha(X))I_{\{X - VaR_\alpha(X) > 0\}}I_{\{X+Y - VaR_\alpha(X+Y) \leq 0\}}]}{\mathbb{P}(X - VaR_\alpha(X) > 0)}, \\
&\geq VaR_\alpha(X) + \frac{E[(X - VaR_\alpha(X))I_{\{X+Y - VaR_\alpha(X+Y) > 0\}}]}{\mathbb{P}(X + Y - VaR_\alpha(X + Y) > 0)} \\
&= VaR_\alpha(X) + E[(X - VaR_\alpha(X)) | X + Y - VaR_\alpha(X + Y) > 0] \\
&= E[X | X + Y > VaR_\alpha(X + Y)].
\end{aligned}$$

De igual manera, se tiene que

$$ES_\alpha(Y) \geq E[Y | X + Y > VaR_\alpha(X + Y)].$$

Luego,

$$\begin{aligned}
ES_\alpha(X + Y) &= E[X + Y | X + Y > VaR_\alpha(X + Y)] \\
&= E[X | X + Y > VaR_\alpha(X + Y)] + E[Y | X + Y > VaR_\alpha(X + Y)] \\
&\leq ES_\alpha(X) + ES_\alpha(Y).
\end{aligned}$$

□

Capítulo 3

Teoría de Valores Extremos y el Método POT

La Teoría de Valores Extremos (EVT, Extreme Value Theory) es una rama de la teoría de la probabilidad que desarrolla técnicas y modelos para describir y estudiar los eventos inesperados, anormales o extremos. Los eventos extremos son aquellos de grandes consecuencias, pero de probabilidad bastante baja como, por ejemplo, una catástrofe natural.

Esta teoría tiene muchas aplicaciones en la práctica, algunas son: el estudio de inundaciones extremas, sequías, terremotos, desastres naturales en general y la gestión de tráfico en telecomunicaciones, entre otros. La teoría de valores extremos se ha convertido recientemente en una herramienta fundamental en aplicaciones para finanzas y seguros, particularmente, en la gestión o manejo de riesgos. Como ejemplo de evento extremo, se puede mencionar la caída del mercado de valores de octubre de 1987, el lunes 19 de octubre el mercado Dow Jones de Estados Unidos perdió el 22,68 % de su valor equivalente a \$500 billones, cuando la peor caída desde 1962 fue del 5,25 %, en esa misma semana los mercados de valores de Hong Kong cayeron un 45,8 %, los de Australia un 41,8 %, los de España un 31 %, entre otros, ocasionando una desestabilización económica global.

La EVT ofrece métodos para cuantificar eventos extremos y sus consecuencias. Para Gumbel (1958,[17]), el objetivo de la EVT es analizar valores extremos observados y, con ello, poder predecir valores extremos en el futuro. Los métodos más conocidos para describir los valores extremos son: el método de máximos por bloque (block

maxima, en inglés) y el método de picos sobre un umbral (POT, por sus siglas en inglés).

El método POT es la técnica más usada para analizar el extremo derecho (o izquierdo) de una función de distribución, área conocida como la *cola* de la función. Se puede definir la cola derecha de una función de distribución F como:

$$\bar{F}(x) = \mathbb{P}_F(X > x) = 1 - F(x),$$

para un valor de x tal que $\bar{F}(x)$ sea “pequeño”.

Una interpretación gráfica de la cola de una distribución se puede ver en la figura 3.1, en donde la cola está representada por el área anaranjada, esto es, con α cercano a 0, $\bar{F}(q_\alpha) = \mathbb{P}(X > q_\alpha)$.

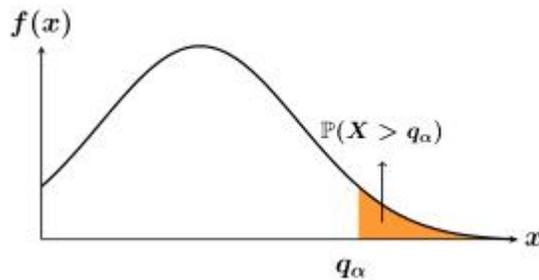


Figura 3.1: La cola de la función $f(x)$ representada por el área anaranjada.

Se pueden clasificar las distribuciones por la forma de su cola, si es pesada o ligera: Las distribuciones de cola pesada (o gruesa) son aquellas cuyas colas son “más grandes” que la cola de una distribución exponencial. Esto es, $F(x)$ es una distribución de cola pesada si

$$\bar{F}(x) > ae^{-bx}, \quad \forall a > 0, b > 0. \quad (3.1)$$

Las distribuciones de cola ligera son aquellas cuyas colas son “más pequeñas” que la cola de una distribución exponencial. Esto es, $F(x)$ es una distribución de cola ligera si

$$\bar{F}(x) < ae^{-bx}, \quad \forall a > 0, b > 0. \quad (3.2)$$

Un ejemplo de esto son las distribuciones Pareto y uniforme. La distribución Pareto es de cola pesada y la distribución uniforme es de cola ligera, como se puede observar en la figura 3.2.

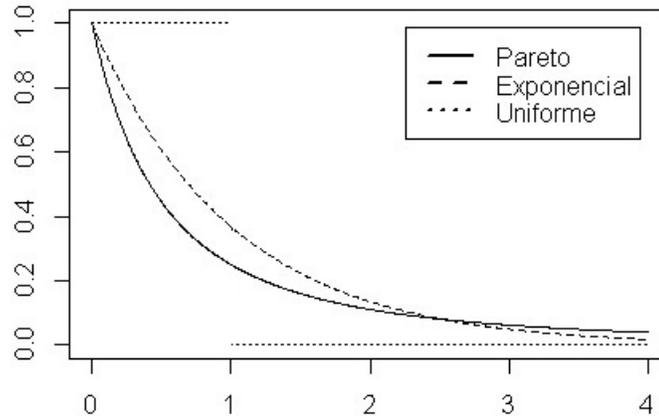


Figura 3.2: Funciones de densidad de las distribuciones: Pareto, Exponencial y Uniforme.

3.1. Dominio Máximo de Atracción

Definición 3.1. Sea X una variable aleatoria. Se dice que X es *degenerada* o que tiene *distribución degenerada* en x_0 si la función de probabilidad de X viene dada por

$$\mathbb{P}(X = x_0) = 1, \quad \mathbb{P}(X \neq x_0) = 0.$$

Es decir, la variable aleatoria es constante con probabilidad 1. En consecuencia, la función de distribución de X viene dada por

$$F_X(x) = \begin{cases} 0, & \text{si } x < x_0 \\ 1, & \text{si } x \geq x_0. \end{cases}$$

Supongamos que X_1, X_2, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas con función de distribución F . Por el ejemplo 1.5.5, sabemos que la función de distribución del máximo $M_n = \max\{X_1, X_2, \dots, X_n\}$ es

$$F^n(x) = \prod_{i=1}^n \mathbb{P}(X_i \leq x).$$

Sea $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$ el punto final derecho de F .

Si $x < x_F$, entonces $F(x) < 1$ y en consecuencia

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

Si $x > x_F$, entonces $F(x) = 1$ y en consecuencia

$$\lim_{n \rightarrow \infty} F^n(x) = 1.$$

Es decir, M_n tiende en distribución a una variable aleatoria degenerada, cuando $n \rightarrow \infty$.

Sin embargo, en algunos casos se puede considerar una normalización de la sucesión de variables M_n , cuyo límite en distribución sea no degenerado. Esto es, existen sucesiones $a_n > 0$ y b_n ($n \geq 1$) tales que

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

tiene una distribución no degenerada cuando $n \rightarrow \infty$, es decir,

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

para cada punto de continuidad de G , y G una función de distribución no degenerada.

Definición 3.2. Sea X una variable aleatoria no degenerada. Se dice que X (o su correspondiente función de distribución) es *máximo-estable* si para X, X_1, \dots, X_n variables aleatorias iid, constantes $a_n > 0$ y $b_n \in \mathbb{R}$ y para $n \geq 2$ se satisface que:

$$\text{máx}(X_1, \dots, X_n) \stackrel{d}{=} a_n X + b_n, \tag{3.3}$$

donde $\stackrel{d}{=}$ significa igualdad en distribución.

Suponiendo que (X_n) es una sucesión de variables aleatorias máximo-estables, entonces (3.3) puede reescribirse como sigue

$$\frac{M_n - b_n}{a_n} \stackrel{d}{=} X.$$

Por lo que cada distribución máximo-estable es una distribución límite para el máximo de variables aleatorias independientes e idénticamente distribuidas. Más aún, las distribuciones máximo-estables son las únicas distribuciones límites para el máximo normalizado. Propiedad que destacan Embrechts ([12], pág. 121) con el siguiente resultado:

Teorema 3.1 (Propiedad límite de distribuciones máximo-estables). *La clase de distribuciones máximo-estables coincide con la clase de todas las posibles distribuciones límite (no degeneradas) del máximo (normalizado adecuadamente) de variables aleatorias independientes e idénticamente distribuidas.*

Definición 3.3. Se dice que F pertenece al *dominio máximo de atracción* de una distribución máximo-estable H , se denota $F \in \mathcal{MDA}(H)$, si existen constantes $a_n > 0$ y $b_n \in \mathbb{R}$ tales que

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

Es decir,

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

3.2. Distribuciones de Valor Extremo

El siguiente teorema es un resultado clásico de la Teoría de Valores Extremos, el cual establece que sólo hay tres posibles distribuciones límites para el máximo normalizado.

Teorema 3.2 (Fisher-Tippett). *Sean X_1, \dots, X_2 variables aleatorias independientes e idénticamente distribuidas con función de distribución F . Si existe una función de distribución no degenerada H tal que $F \in \mathcal{MDA}(H)$, entonces H pertenece a alguna de las tres siguientes familias de distribuciones:*

$$\text{Fréchet:} \quad \Phi_\alpha(x) = \begin{cases} 0, & \text{si } x \leq 0 \\ e^{-x^{-\alpha}}, & \text{si } x > 0. \end{cases} \quad (3.4)$$

$$\text{Weibull:} \quad \Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha}, & \text{si } x < 0 \\ 1, & \text{si } x \geq 0. \end{cases} \quad (3.5)$$

$$\text{Gumbel:} \quad \Lambda(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}. \quad (3.6)$$

Observación. El teorema 3.2 no garantiza la existencia de un límite no degenerado para M_n , ni indica cuál es el límite cuando existe. Lo que sí indica es que, cuando el límite existe, tiene que ser una de las tres distribuciones: Fréchet, Weibull o Gumbel, cualquiera sea la distribución inicial F .

Las tres familias de distribuciones del teorema 3.2 son conocidas como Distribuciones de Valor Extremo (EVD, por sus siglas en inglés). Y el parámetro α es conocido como *índice de colas*.

En la figura 3.3 se presentan las funciones de densidad y distribución para cada una de las familias de distribución de valor extremo. La distribución Weibull posee un extremo superior (o punto final derecho) finito, mientras que la distribución Fréchet presenta un extremo inferior (o punto final izquierdo) finito. La distribución que presenta la tasa de crecimiento más rápida es la distribución Weibull, siendo la distribución Fréchet la que presenta la tasa de crecimiento más lenta.

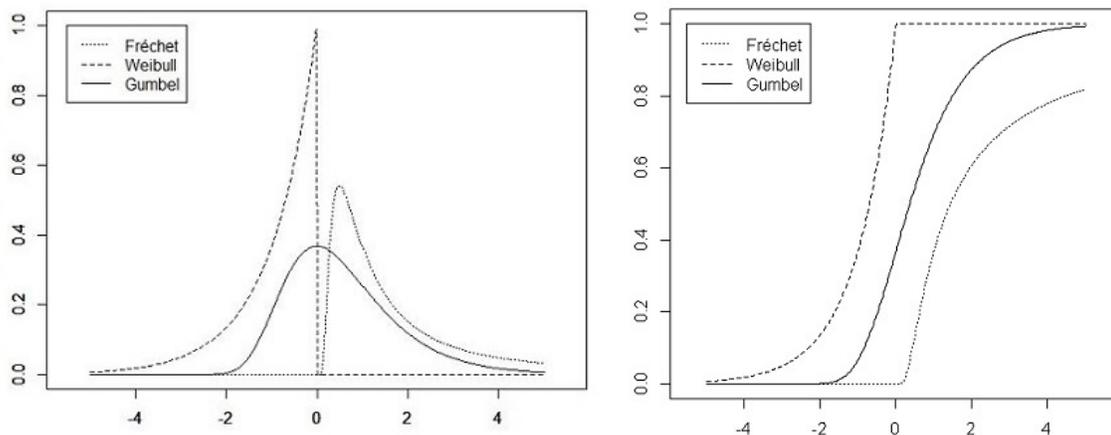


Figura 3.3: Funciones de densidad (izquierda) y distribución (derecha) de las distribuciones de valor extremo. Para Fréchet y Weibull ($\alpha = 1$).

Las distribuciones Fréchet y Weibull son ejemplos de distribuciones de *colas pesadas* o *gruesas*, pues satisfacen la ecuación 3.1.

Ejemplo 3.1. Sea F una distribución exponencial de parámetro $\lambda = 1$, es decir, $F(x) = 1 - e^{-x}$ para $x > 0$, entonces $F^n(x) = (1 - e^{-x})^n$ y

$$\begin{aligned} F^n(x + \ln(n)) &= (1 - e^{-x - \ln(n)})^n \\ &= \left(1 - \frac{1}{n}e^{-x}\right)^n \\ &\rightarrow e^{-e^{-x}} = \Lambda(x). \end{aligned}$$

Por lo tanto, con una normalización $a_n = 1$ y $b_n = \ln(n)$, M_n tiene como límite una distribución Gumbel. Es decir, F pertenece al dominio máximo de atracción de la distribución Gumbel.

De igual manera, si $a_n = n^{1/\alpha}$ y $b_n = 0$ entonces M_n tiene como límite la distribución Fréchet, si el límite existe. Y, si $a_n = n^{-1/\alpha}$ y $b_n = 0$ entonces M_n tiene como límite la distribución Weibull.

Existe una representación alternativa de las distribuciones de valor extremo conocida como la representación Jenkinson-Von Mises. En esta reformulación se representan las tres distribuciones de valor extremo mediante una distribución de parámetro ξ , conocida como la *distribución generalizada de valor extremo* (GEVD, por sus siglas en inglés), donde ξ es el parámetro de forma, que indica el espesor de la cola de la distribución: mientras más grande sea ξ , más pesada (o gruesa) será la cola de la distribución, como se observa en la figura 3.4, donde la cola de la distribución con $\xi = 0,5$ es más pesada que la cola de la distribución con $\xi = 0$ que, a su vez, es más pesada que la cola de la distribución con $\xi = -0,5$.

Definición 3.4 (Representación Jenkinson-Von Mises de las EVD). Se define la función de *distribución generalizada de valor extremo* (GEVD) H_ξ como

$$H_\xi(x) = \begin{cases} e^{-(1 + \xi x)^{-1/\xi}}, & \text{si } \xi \neq 0, \\ e^{-e^{-x}}, & \text{si } \xi = 0, \end{cases}$$

donde $1 + \xi x > 0$ y $\xi \in \mathbb{R}$.

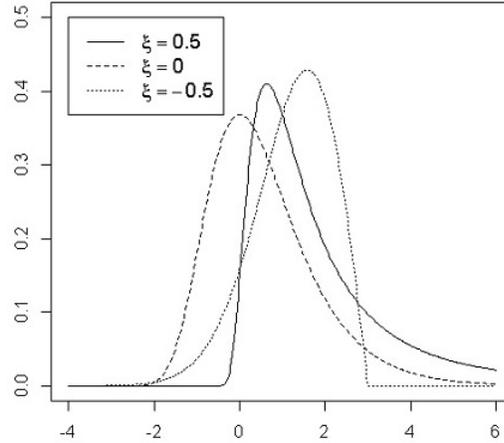


Figura 3.4: Función de densidad de la distribución generalizada de valor extremo para $\xi = -0,5$, $\xi = 0$ y $\xi = 0,5$ (distribuciones Weibull, Gumbel y Fréchet, respectivamente).

Si $\xi > 0$ se tiene la distribución Fréchet con parámetro de forma $\alpha = 1/\xi$, si $\xi < 0$ se tiene la distribución Weibull con parámetro de forma $\alpha = -1/\xi$ y si $\xi = 0$ se obtiene la distribución Gumbel.

Esta expresión conjunta de las tres distribuciones límite permite hacer inferencia estadística sin seleccionar previamente una de las tres distribuciones posibles: Fréchet, Weibull y Gumbel.

En general, el dominio máximo de atracción de las distribuciones de valor extremo, $\mathcal{MDA}(H_\xi)$, depende de las características de la distribución que se esté analizando, como lo indican Embrechts ([12], pág. 158) mediante el siguiente teorema:

Teorema 3.3 (Caracterización del $\mathcal{MDA}(H_\xi)$). Para $\xi \in \mathbb{R}$, las siguientes afirmaciones son equivalentes:

- (i) $F \in \mathcal{MDA}(H_\xi)$.
- (ii) Existe una función medible positiva $a(\cdot)$ tal que para $1 + \xi x > 0$,

$$\lim_{u \uparrow x_F} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \begin{cases} (1 + \xi x)^{-1/\xi}, & \text{si } \xi \neq 0, \\ e^{-x}, & \text{si } \xi = 0. \end{cases}$$

(iii) Para $x, y > 0$, $y \neq 1$,

$$\lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1}, & \text{si } \xi \neq 0, \\ \frac{\ln(x)}{\ln(y)}, & \text{si } \xi = 0. \end{cases}$$

Donde $U(\alpha) = F^{-1}(1 - \alpha^{-1})$, con $\alpha > 0$, y F^{-1} es la función cuantil de F .

Observación. La notación $u \uparrow x_F$ significa que u se aproxima a x_F de manera creciente.

Particularmente, las distribuciones que presentan colas pesadas, usuales en aplicaciones financieras, se encuentran en el dominio máximo de atracción de las distribuciones Gumbel y Fréchet, aunque principalmente se concentran en el de la distribución Fréchet.

3.3. Método POT y la Distribución Generalizada de Pareto

Un tipo de metodología desarrollada en la teoría del valor extremo es la asociada a la distribución de los excesos sobre un umbral, conocido como Picos sobre el Umbral (POT, por sus siglas en inglés).

Supongamos que X, X_1, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas con función de distribución F tal que $F \in \mathcal{MDA}(H_\xi)$, para algún $\xi \in \mathbb{R}$.

Dado un umbral alto u , se tiene que

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

es el número de *excedentes* sobre u por X_1, \dots, X_n . Y los *excesos* correspondientes son $Y_k = X_{i_k} - u$ para $k = 1, \dots, N_u$, $i_k = 1, \dots, n$ y $X_{i_k} > u$.

Ejemplo 3.2. Supongamos que se tienen los siguientes datos: 6, 2, 10, 2, 8, 1, 4, 7. Si el umbral es $u = 6$, entonces se tiene que los excedentes son 10, 8, 7, es decir, 3 de los valores son mayores a 6, ($N_u = 3$). Luego, los excesos correspondientes son $10 - 6, 8 - 6, 7 - 6$, es decir, 4, 2, 1. (Ver figura 3.5)

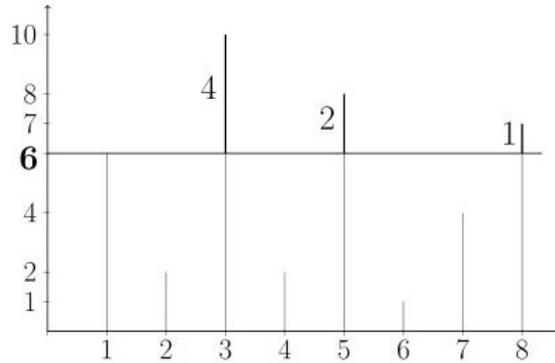


Figura 3.5: Gráfica de los datos del ejemplo 3.2 y sus correspondientes excedentes.

Luego, la *función de distribución condicional de los excesos* de X viene dada por

$$\begin{aligned}
 F_u(y) &= \mathbb{P}(X - u \leq y | X > u) \\
 &= \frac{\mathbb{P}(u < X \leq y + u)}{\mathbb{P}(X > u)} \\
 &= \frac{F(u + y) - F(u)}{1 - F(u)}
 \end{aligned} \tag{3.7}$$

para $0 \leq y < x_F - u$.

Es decir, $F_u(y)$ es la probabilidad de que un valor de la muestra exceda el umbral u por un valor no mayor a “ y ”, dado que el umbral es excedido.

Recordando que $\bar{F}(x) = 1 - F(x)$, se puede escribir la relación (3.7) como

$$\bar{F}(u + y) = \bar{F}(u)\bar{F}_u(y), \tag{3.8}$$

pues

$$\begin{aligned}
 F_u(y) &= \frac{F(u + y) - F(u)}{1 - F(u)} \\
 &= \frac{F(u + y) - 1 + 1 - F(u)}{1 - F(u)} \\
 &= \frac{\bar{F}(u) - \bar{F}(u + y)}{\bar{F}(u)}
 \end{aligned}$$

de donde

$$\begin{aligned}
 \bar{F}(u + y) &= \bar{F}(u) - \bar{F}(u)F_u(y) \\
 \bar{F}(u + y) &= \bar{F}(u)(1 - F_u(y)).
 \end{aligned}$$

La *Distribución Generalizada de Pareto* (GPD, por sus siglas en inglés) tiene una gran importancia en los modelos de umbrales en particular y en la teoría de valores extremos en general. Embrechts (1997,[12]) destaca un resultado de Balkema - de Haan (1974,[4]) y Pickands (1975,[28]) donde se establece que la distribución apropiada para modelar los excesos sobre el umbral u es la distribución generalizada de Pareto, la cual es usualmente expresada como una función de distribución de dos parámetros $G_{\xi,\beta}$.

Teorema 3.4. Para cada $\xi \in \mathbb{R}$, $F \in \mathcal{MDA}(H_\xi)$ si y sólo si

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |\bar{F}_u(x) - \bar{G}_{\xi,\beta(u)}(x)| = 0,$$

para alguna función $\beta(u)$ positiva, estimada a partir de los datos.

En otras palabras, para un umbral u suficientemente grande, el teorema anterior sugiere que

$$\bar{F}_u(x) = \mathbb{P}(X - u > x | X > u) \approx \bar{G}_{\xi,\beta(u)}(x), \quad x > 0, \quad (3.9)$$

es decir, la función de distribución condicional de los excesos de X es aproximadamente igual a la distribución generalizada de Pareto.

Definición 3.5. Se define la función de *distribución generalizada de Pareto* con parámetros $\xi \in \mathbb{R}$ y $\beta > 0$, como

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}}, & \text{si } \xi \neq 0, \\ 1 - e^{-\frac{x}{\beta}}, & \text{si } \xi = 0, \end{cases}$$

donde

$$\begin{aligned} x &\geq 0, & \text{si } \xi &\geq 0, \\ 0 \leq x &\leq -\frac{\beta}{\xi}, & \text{si } \xi &< 0. \end{aligned}$$

La Distribución Generalizada de Pareto $G_{\xi,\beta}$ presenta los siguientes casos particulares:

- cuando $\xi \rightarrow 0$, $G_{\xi,\beta}$ es la distribución exponencial de parámetro $\frac{1}{\beta}$;

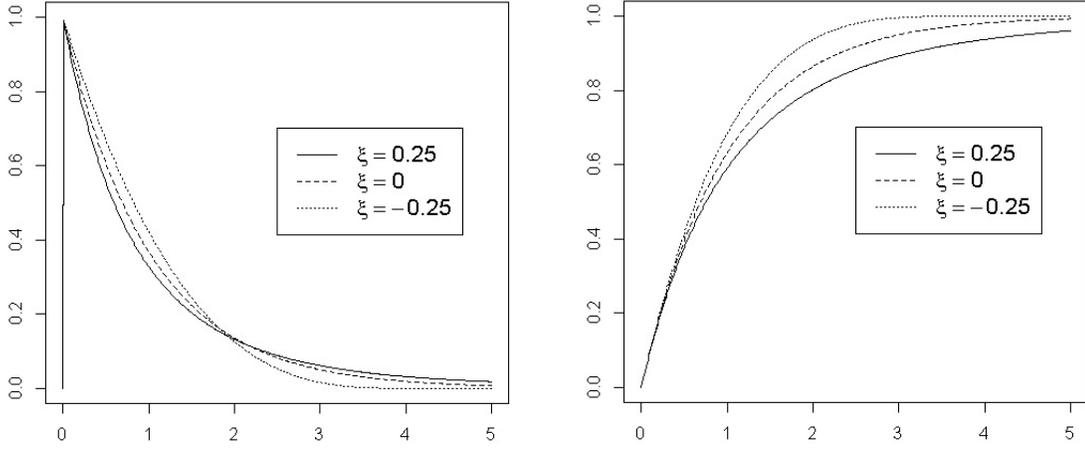


Figura 3.6: Funciones de densidad (izquierda) y de distribución (derecha) de la GPD con $\beta = 1$ y distintos valores de ξ .

- cuando $\xi = -1$, $G_{\xi,\beta}$ es la distribución uniforme en $[0, \beta]$;
- cuando $\xi > 0$, $G_{\xi,\beta}$ es una versión parametrizada de la distribución Pareto.

De otra manera, también se puede decir que $G_{\xi,\beta}$ es una distribución de cola pesada si $\xi > 0$ o de cola ligera si $\xi < 0$.

Ahora bien, la ecuación (3.8) sugiere un método para estimar la cola de F , estimando $\overline{F}_u(y)$ y $\overline{F}(u)$ por separado. Un estimador natural para $\overline{F}(u)$ viene dado por la función de distribución empírica

$$\widehat{\overline{F}(u)} = \overline{F}_n(u) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i > u\}} = \frac{N_u}{n}. \quad (3.10)$$

Por otra parte, la aproximación (3.9) motiva un estimador para $\overline{F}_u(y)$, con u grande, de la forma:

$$\widehat{\overline{F}_u(y)} = \overline{G}_{\widehat{\xi}, \widehat{\beta}}(y), \quad \widehat{\xi} = \widehat{\xi}_{N_u}, \quad \widehat{\beta} = \widehat{\beta}_{N_u}. \quad (3.11)$$

En consecuencia, por (3.8), (3.10) y (3.11), un estimador para la cola $\overline{F}(u+y)$ ($y > 0$) es

$$\widehat{\overline{F}(u+y)} = \frac{N_u}{n} \left(1 + \widehat{\xi} \frac{y}{\widehat{\beta}} \right)^{-1/\widehat{\xi}}. \quad (3.12)$$

Además, recordando la definición de la inversa generalizada de F , (1.27), se puede calcular un estimador para el α -cuantil q_α invirtiendo la ecuación (3.12), esto es

$$1 - \alpha = \widehat{\overline{F}}(q_\alpha) = \frac{N_u}{n} \left(1 + \widehat{\xi} \frac{\widehat{q}_\alpha - u}{\widehat{\beta}} \right)^{-1/\widehat{\xi}}$$

de donde

$$\widehat{q}_\alpha = u + \frac{\widehat{\beta}}{\widehat{\xi}} \left(\left(\frac{n}{N_u} (1 - \alpha) \right)^{-\widehat{\xi}} - 1 \right). \quad (3.13)$$

Definición 3.6. Se define la *función media de excesos* de X como

$$e(u) = E[X - u | X > u].$$

El siguiente teorema, tomado de Embrechts (1997,[12]), provee una técnica gráfica para escoger un umbral u alto tal que la aproximación 3.9 se cumpla.

Teorema 3.5. *Supongamos que X tiene una GPD con parámetros $\xi < 1$ y $\beta > 0$. Entonces para $u < x_F$,*

$$e(u) = E[X - u | X > u] = \frac{\beta + \xi u}{1 - \xi}, \quad \beta + \xi u > 0. \quad (3.14)$$

Y, si $v > u$, se tiene que

$$e(v) = \frac{\beta + \xi(v - u)}{1 - \xi}. \quad (3.15)$$

Se sigue de Embrechts (1997,[12]) que un método de selección de u , tal que éste sea óptimo, está basado en la linealidad de la función media de excesos $e(u)$, mediante la igualdad 3.14.

Dada una muestra X_1, \dots, X_n independiente e idénticamente distribuida se define la *función media de excesos empírica* como una versión muestral de la función media de excesos $e(u)$, como sigue:

$$e_n(u) = \frac{1}{N_u} \sum_{i \in \Delta_n(u)} (X_i - u), \quad u > 0, \quad (3.16)$$

donde $\Delta_n(u) = \{i : X_i > u, i = 1, \dots, n\}$ y $N_u = \text{card} \Delta_n(u)$.

Luego, por (3.14), se puede escoger $u > 0$ tal que el gráfico de $(u, e_n(u))$ siga aproximadamente una línea recta para $x \geq u$.

Cuando la distribución es de cola pesada ($\xi > 0$), la función media de excesos empírica es una función creciente de u . Cuando la distribución es de tipo exponencial, el gráfico de la media de excesos es aproximadamente constante, pues este es el caso cuando $\xi = 0$. Y, cuando la distribución es de cola ligera ($\xi \leq 0$), el gráfico de la media de excesos decrece a 0.

Al final de este capítulo, se aplicará este método de selección de u en un ejemplo de datos reales (ejemplo 3.3). Además, se verá que la función media de excesos empírica es una función creciente, dado que la distribución de los datos tiene cola pesada.

3.4. Estimación por Máxima Verosimilitud

Uno de los métodos más usados para hallar estimadores de los parámetros desconocidos de una distribución teórica se conoce como estimación por máxima verosimilitud.

Definición 3.7. Se conoce como *función de verosimilitud*, \mathcal{L} , de una muestra observada a la densidad conjunta (o función de probabilidad conjunta) de la muestra aleatoria X_1, \dots, X_n considerada como función del parámetro o parámetros desconocidos. Esto es,

$$\mathcal{L}(\beta_1, \dots, \beta_k; X_1, \dots, X_n) = f_{X_1, \dots, X_n}(x_1, \dots, x_n; \beta_1, \dots, \beta_k),$$

donde β_1, \dots, β_k son los parámetros desconocidos de la distribución.

Para una muestra independiente e idénticamente distribuida X_1, \dots, X_n con función de densidad f , se tiene que

$$\mathcal{L}(\beta_1, \dots, \beta_k; X_1, \dots, X_n) = f_{X_1, \dots, X_n}(x_1, \dots, x_n; \beta_1, \dots, \beta_k) = \prod_{i=1}^n f(x_i; \beta_1, \dots, \beta_k).$$

El método de máxima verosimilitud consiste en elegir los parámetros β_1, \dots, β_k tales que maximicen la función de verosimilitud, \mathcal{L} . Es decir, si \mathcal{L} es una función

derivable dos veces respecto a los parámetros, entonces los valores β_1, \dots, β_k vienen dados a partir de la siguiente ecuación

$$\frac{\partial \mathcal{L}}{\partial \beta_i}(\beta_1, \dots, \beta_k; X_1, \dots, X_n) = 0.$$

Además, los parámetros β_1, \dots, β_k deben verificar el criterio de Sylvester para máximos:

Si definimos las submatrices de la matriz M_k de las segundas derivadas de \mathcal{L} como sigue:

$$M_1 = \left(\frac{\partial^2 \mathcal{L}}{\partial \beta_1^2} \right), M_2 = \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial \beta_1^2} & \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \partial \beta_2} \\ \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \partial \beta_2} & \frac{\partial^2 \mathcal{L}}{\partial \beta_2^2} \end{pmatrix}, \dots, M_k = \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial \beta_1^2} & \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \partial \beta_2} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \partial \beta_k} \\ \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \partial \beta_2} & \frac{\partial^2 \mathcal{L}}{\partial \beta_2^2} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial \beta_2 \partial \beta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \partial \beta_k} & \frac{\partial^2 \mathcal{L}}{\partial \beta_2 \partial \beta_k} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial \beta_k^2} \end{pmatrix}$$

entonces, los parámetros β_1, \dots, β_k deben verificar que los determinantes de las submatrices de índice impar sean negativos y los de índice par sean positivos, esto es

$$\det(M_i) < 0, \quad \det(M_j) > 0,$$

para $i, j = 1, \dots, k$ tales que i es impar y j es par.

Ahora bien, observe que

$$\frac{\partial \mathcal{L}}{\partial \beta_i} = 0 \text{ si y sólo si } \frac{\partial(\ln \mathcal{L})}{\partial \beta_i} = 0.$$

Así, se calculan los valores de β_1, \dots, β_k tales que maximicen la función $\ln \mathcal{L}$; pues, es más conveniente trabajar con el logaritmo de la función de máxima verosimilitud cuando ésta depende de funciones exponenciales, como las distribuciones generalizadas de valor extremo y la distribución Pareto.

Usualmente, $\ln \mathcal{L}$ es llamada *log-verosimilitud* y denotada por l :

$$l(\beta_1, \dots, \beta_k) = \ln \mathcal{L}(\beta_1, \dots, \beta_k; X_1, \dots, X_n) = \sum_{i=1}^n \ln f(x_i; \beta_1, \dots, \beta_k). \quad (3.17)$$

Estimación de los parámetros de la GPD

Sean Y_1, \dots, Y_n variables aleatorias independientes e idénticamente distribuidas con función de distribución $G_{\xi, \beta}$, distribución generalizada de Pareto con parámetros ξ y β , entonces la función de densidad $g_{\xi, \beta}$ correspondiente es

$$g_{\xi, \beta}(y) = \begin{cases} \frac{1}{\beta} \left(1 + \xi \frac{y}{\beta}\right)^{-1/\xi-1}, & \text{si } \xi \neq 0, \\ \frac{1}{\beta} e^{-y/\beta}, & \text{si } \xi = 0, \end{cases} \quad (3.18)$$

donde

$$\begin{aligned} y &\geq 0, & \text{si } \xi &\geq 0, \\ 0 \leq y &\leq -\frac{\beta}{\xi}, & \text{si } \xi &< 0. \end{aligned}$$

Habiendo determinado el umbral u , los parámetros de la distribución Generalizada de Pareto se pueden estimar por máxima verosimilitud.

Supongamos que los valores y_1, \dots, y_{N_u} son los N_u excesos del umbral u . Entonces, para $\xi \neq 0$, la función de log-verosimilitud, $l(\xi, \beta)$, viene dada por (3.17) y (3.18), como sigue:

$$l(\xi, \beta) = \sum_{i=1}^{N_u} \ln \left(\frac{1}{\beta} \left(1 + \xi \frac{y_i}{\beta}\right)^{-1/\xi-1} \right) \quad (3.19)$$

$$= \sum_{i=1}^{N_u} -\ln \beta + \left(-\frac{1}{\xi} - 1\right) \ln \left(1 + \xi \frac{y_i}{\beta}\right) \quad (3.20)$$

$$= -N_u \ln \beta - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{N_u} \ln \left(1 + \frac{\xi}{\beta} y_i\right) \quad (3.21)$$

siempre que $1 + \frac{\xi}{\beta} y_i > 0$ para $i = 1, \dots, N_u$.

Si $\xi = 0$, la función de log-verosimilitud viene dada por

$$l(\beta) = \sum_{i=1}^{N_u} \ln \left(\frac{1}{\beta} e^{-\frac{y_i}{\beta}} \right)$$

$$\begin{aligned}
 &= \sum_{i=1}^{N_u} -\ln \beta - \frac{y_i}{\beta} \\
 &= -N_u \ln \beta - \frac{1}{\beta} \sum_{i=1}^{N_u} y_i.
 \end{aligned}$$

Analíticamente no es posible maximizar la función log-verosimilitud, es decir, no es posible dar una fórmula explícita para $\hat{\xi}$ y $\hat{\beta}$. Sin embargo, se puede derivar y resolver numéricamente, obteniendo los estimadores de máxima verosimilitud $\hat{\beta}_{N_u}, \hat{\xi}_{N_u}$, para $\xi > -1/2$. [12]

Para muestras relativamente grandes, se pueden usar como apoyo herramientas de software matemático como R-Project, Matlab, entre otros.

3.5. Medidas de Riesgo y el método POT

El Valor en Riesgo se puede describir como la máxima pérdida en circunstancias normales o como la mínima pérdida en circunstancias extraordinarias, por lo que se puede considerar natural estimar el VaR a partir del máximo M_n de una muestra X_1, \dots, X_n de variables aleatorias independientes e idénticamente distribuidas que representan pérdidas.

Supongamos que X, X_1, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas con función de distribución F , tal que $F \in \mathcal{MDA}(H_\xi)$, para algún $\xi \in \mathbb{R}$. Dado un umbral alto u , por (3.13), sabemos que el estimador de q_α ($q_\alpha > u$) es

$$\hat{q}_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u} (1 - \alpha) \right)^{-\hat{\xi}} - 1 \right),$$

donde ξ, β son los parámetros de la distribución generalizada de Pareto $G_{\xi, \beta}$.

Luego, para $\alpha \geq F(u)$, tenemos que un estimador para el VaR viene dado por:

$$\widehat{VaR}_\alpha(X) = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u} (1 - \alpha) \right)^{-\hat{\xi}} - 1 \right). \quad (3.22)$$

Así, por la ecuación (3.15) y por la linealidad de la esperanza condicional, se tiene que para $\widehat{VaR}_\alpha(X) > u$:

$$\begin{aligned}
 \widehat{ES}_\alpha(X) &= \widehat{VaR}_\alpha(X) + E\left[X - \widehat{VaR}_\alpha(X) \mid X > \widehat{VaR}_\alpha(X)\right] \\
 &= \widehat{VaR}_\alpha(X) + \frac{\widehat{\beta} + \widehat{\xi} \left(\widehat{VaR}_\alpha(X) - u\right)}{1 - \widehat{\xi}} \\
 &= \widehat{VaR}_\alpha(X) \left(1 + \frac{\widehat{\xi}}{1 - \widehat{\xi}}\right) + \frac{\widehat{\beta} - \widehat{\xi} u}{1 - \widehat{\xi}} \\
 &= \widehat{VaR}_\alpha(X) \left(\frac{1}{1 - \widehat{\xi}} + \frac{\widehat{\beta} - \widehat{\xi} u}{(1 - \widehat{\xi})\widehat{VaR}_\alpha(X)}\right). \tag{3.23}
 \end{aligned}$$

Para facilitar la comprensión de este capítulo, se ilustrará el método POT ajustando la GPD a datos reales mediante el siguiente ejemplo:

Ejemplo 3.3. Consideremos como muestra los datos de grandes reclamos de seguros contra incendios en Dinamarca, comprendidos entre el 3 de enero de 1980 y el 31 de diciembre de 1990. Éstos datos se pueden encontrar en el paquete “*evir*” del programa *R-Project*, programa que será de apoyo para los cálculos y gráficos necesarios.

Los datos Daneses consisten en 2167 pérdidas mayores a 1 millón de coronas danesas (DKK). Estas cifras de pérdidas son totales e incluyen daños a edificios, daños a muebles e inmuebles, así como la pérdida de ganancias de negocios.

El gráfico de los datos daneses (figura 3.7) permite identificar las pérdidas más extremas y el tiempo de ocurrencia aproximado de las mismas. También se puede observar si hay evidencia de agrupamiento de grandes pérdidas, lo que podría generar dudas sobre la suposición de que los datos son independientes e idénticamente distribuidos. Éste no parece ser el caso con los datos daneses.

El histograma en escala logarítmica (figura 3.7) muestra el amplio rango de los datos. La mayor pérdida en el conjunto de datos es de 263 millones de coronas. También se puede observar, al comparar su histograma con la densidad exponencial, que la

distribución de los datos tiene cola pesada y es asimétrica a la derecha, sugiriendo que las pérdidas pequeñas (entre 0 y 1 millón DKK) ocurren con bastante frecuencia; sin embargo, también pueden ocurrir pérdidas muy grandes ocasionalmente.

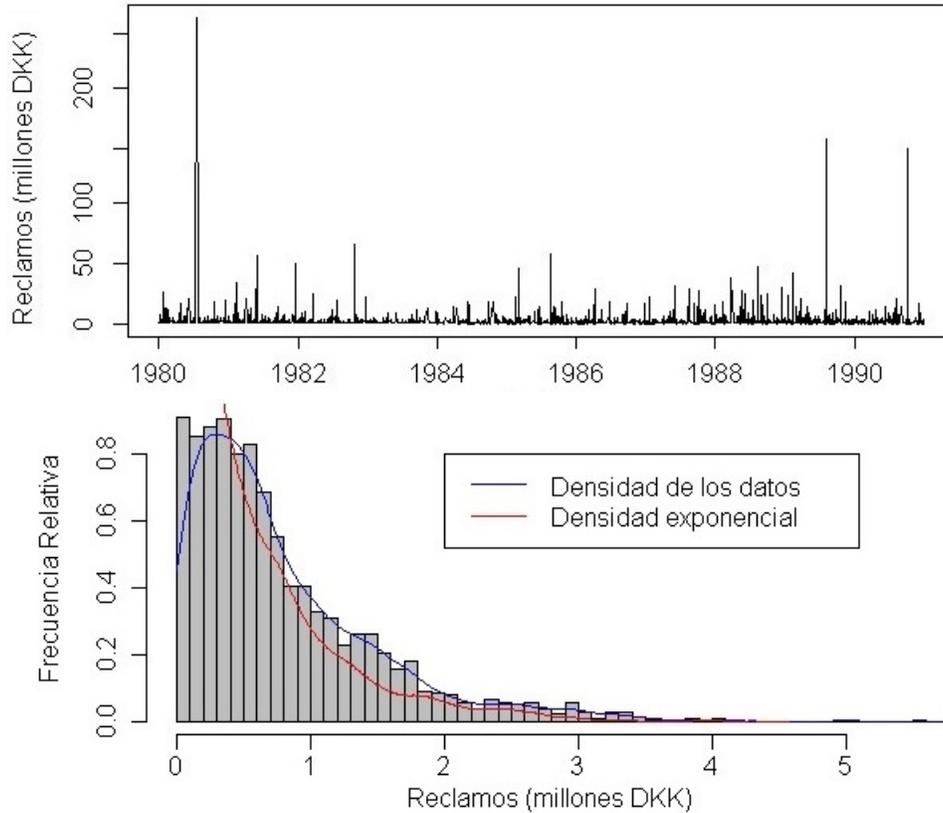


Figura 3.7: Grafica de los reclamos de seguros contra incendios de Dinamarca con respecto al tiempo e histograma en escala logarítmica.

Para poder estimar a partir de los datos el parámetro de forma ξ , se debe escoger primero el valor del umbral u (de manera equivalente, el número de los excesos N_u). Para esto, se grafica la función media de excesos empírica $e_n(u)$, dada por la ecuación (3.16).

El comportamiento de pendiente positiva que se observa en la gráfica de $e_n(u)$ (figura 3.8) evidencia que la distribución es de cola pesada.

Se escoge el umbral u en una región sobre la cual el gráfico de $e_n(u)$ sea aproximadamente lineal, por lo que se sugiere $u = 10$ millones DKK ($N_u = 109$ excedentes).

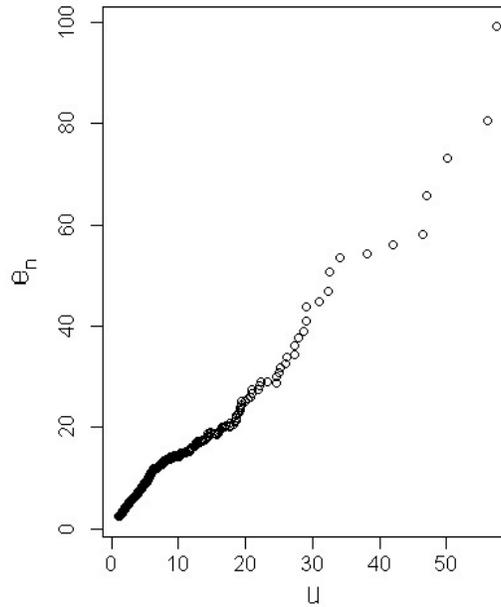


Figura 3.8: Gráfica de la función media de excesos empírica de los datos daneses.

En realidad, toda la gráfica es suficientemente recta como para sugerir que la GPD pueda representar un ajuste razonable para todo el conjunto de datos.

Una vez seleccionado el umbral u , se estiman los valores de los parámetros ξ y β de la GPD por el método de máxima verosimilitud. Mediante el programa R-Project, se tiene que $\hat{\xi} = 0,497$ y $\hat{\beta} = 6,974$ para $u = 10$ ($N_u = 109$ excedentes).

A partir de estos estimadores, y por la relación (3.11), una estimación para la distribución condicional de los excesos $F_u(y)$ dado que el umbral es excedido, viene dada por

$$\begin{aligned} \widehat{F}_u(y) &= 1 - \left(1 + \hat{\xi} \frac{y}{\hat{\beta}}\right)^{-\frac{1}{\hat{\xi}}} \\ &= 1 - (1 + 0,0712 y)^{-2,012}. \end{aligned}$$

En la figura 3.9 se muestra la gráfica de la distribución empírica de los excesos y la GPD ajustada a los datos, donde se puede observar que, en efecto, la distribución generalizada de Pareto es una muy buena aproximación de los datos.

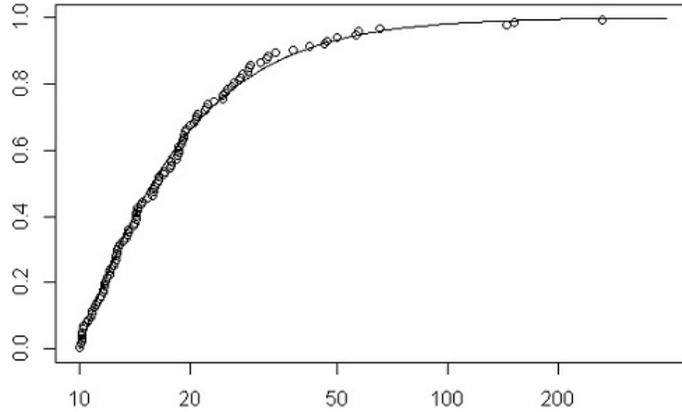


Figura 3.9: Gráfica de la distribución empírica de los excesos de los datos daneses y gráfica de la GPD ajustada a los datos daneses.

Ahora bien, por las ecuaciones (3.22) y (3.23), se tiene para $\alpha = 0,95$ que

$$\begin{aligned}\widehat{VaR}_{0,95}(X) &= 10 + \frac{6,974}{0,497} \left(\left(\frac{2167}{109} (1 - 0,95) \right)^{-0,497} - 1 \right) \\ &= 10,042.\end{aligned}$$

Y, así

$$\begin{aligned}\widehat{ES}_{0,95}(X) &= 10,042 \left(\frac{1}{1 - 0,497} + \frac{6,974 - (0,497)(10)}{(1 - 0,497)(10,042)} \right) \\ &= 23,947.\end{aligned}$$

Es decir, existe un 5 % de probabilidad de incurrir en una pérdida mayor a 10,042. En este caso, se tiene entonces que la pérdida esperada es 23,947.

Finalizando, es importante resaltar de este capítulo que la aplicación de este método requiere que las pérdidas sean independientes, que los excesos siguen una distribución generalizada de Pareto y que el número de excesos sigue una distribución de Poisson, de acuerdo con Leadbetter (1991, [20]).

Capítulo 4

Método POT adaptado al modelo no estacionario y covariables

En este capítulo se analizarán las pérdidas por riesgo operacional y se verá, entre otras cosas, que por lo general la cantidad de excesos de estas pérdidas depende del tiempo, por lo que será necesario adaptar el método POT para poder estudiar los extremos de pérdidas generadas por este tipo de riesgo.

4.1. Riesgo Operacional

Recordemos que el riesgo operacional se define como el riesgo de pérdidas financieras originadas por insuficiencias o fallas de procesos, personas y sistemas internos, o por eventos externos. Las siguientes situaciones ejemplifican consecuencias de riesgo operacional sufridas en varios ámbitos en años recientes:

- En 1995, el banco Barings en Reino Unido colapsó después de perder \$1.3 billones como resultado de transacciones especulativas, fraudulentas y no autorizadas, llevadas a cabo por un empleado llamado Nick Leeson.
- En 2001, el atentado terrorista el 11 de septiembre en USA.
- En 2001, la empresa energética Enron con sede en Texas, USA colapsó luego de revelarse que su condición financiera reportada fue sostenida por un fraude contable, institucionalizado y sistemático.

- En 2010, las erupciones de un volcán en Islandia arrojaron ceniza volcánica a la atmósfera, lo que llevó al cierre del espacio aéreo sobre la mayor parte del norte de Europa, afectando a 10 millones de pasajeros. Este evento es la mayor interrupción de viajes aéreos desde la segunda guerra mundial.

Dirigido a las entidades bancarias, en junio de 2004, el Comité de Supervisión Bancaria de Basilea publicó su segundo acuerdo, conocido como Basilea II. En este acuerdo se introducen las regulaciones para la gestión del riesgo operacional y se definen los 7 tipos de pérdidas por riesgo operacional:

1. Fraude interno: evasión de impuestos, soborno.
2. Fraude externo: robo de información, piratería.
3. Relaciones laborales y seguridad en el trabajo: discriminación, la salud y seguridad de los empleados.
4. Clientes, productos y práctica empresarial: manipulación del mercado, comercio inadecuado, defectos del producto.
5. Daños a activos físicos: desastres naturales, vandalismo, terrorismo.
6. Interrupción en los negocios y fallas en los sistemas: fallas de software, fallas de hardware, interrupción de servicios públicos.
7. Ejecución, entrega y gestión de procesos: errores de entrada de datos, errores de contabilidad, pérdida negligente de los activos de los clientes.

El acuerdo Basilea II se estructura en tres pilares. En cuanto a la medición del riesgo, en el Pilar I se establece el cálculo de requerimientos mínimos de capital para afrontar los distintos riesgos.

Sobre el riesgo operacional, se distinguen tres metodologías para calcular el capital requerido:

- Método del indicador básico: está basado en los ingresos anuales de la entidad financiera.
- Método estándar: está basado en los ingresos anuales de cada uno de los grandes sectores de actividad de la entidad financiera, conocidos como líneas de negocios.

- Métodos de medición avanzada: se basa en el marco de medición de riesgos del banco desarrollados internamente (métodos de medición interna, de distribución de pérdidas, entre otros).

En estudios de varias pérdidas provenientes de distintos bancos se han observado las siguientes características sobre las pérdidas por riesgo operacional:

- La gran magnitud de las pérdidas sugiere que éstas están relacionadas con extremos.
- Las pérdidas están espaciadas irregularmente en el tiempo.

En general, se puede decir que *la frecuencia y la severidad de las pérdidas por riesgo operacional son no estacionarias, es decir, la ocurrencia de las pérdidas y la magnitud de éstas son factores que dependen del tiempo, al igual que sus propiedades estadísticas, como la esperanza.*

Otro aspecto de importancia es que los datos de pérdidas por factores de riesgo operacional son bastante escasos. Una de las razones es el relativamente corto período durante el cual los datos históricos se han recogido de forma coherente. Otra razón es la confidencialidad.

A medida que los bancos recolectan los datos, además de reportar las pérdidas actuales, se hace un esfuerzo para construir las bases de datos que, en la actualidad, se remontan a unos 20 años. Esto último, sin duda, limita el alcance y la precisión de los resultados que se pueden obtener utilizando métodos estadísticos.

Una posibilidad de obtener una base de datos sustancial puede ser agrupar datos de pérdidas entre bancos (o pérdidas de distintos tipos) con el fin de encontrar las principales características de las distribuciones de pérdidas subyacentes, que pueden ser calibradas en contraste con la propia experiencia de pérdida de un banco en particular.

Para ilustrar lo anterior, consideremos la figura 4.1 tomada de Chavez-Demoulin y Embrechts (2004, [7]). Los datos representados en esta figura abarcan un período de 10 años (1992-2001) para tres diferentes tipos de pérdida por riesgo operacional, en adelante, señalados como: Tipo 1, Tipo 2 y Tipo 3. Estas pérdidas provienen de datos reales de un banco que fueron transformados para conservar el anonimato del mismo.

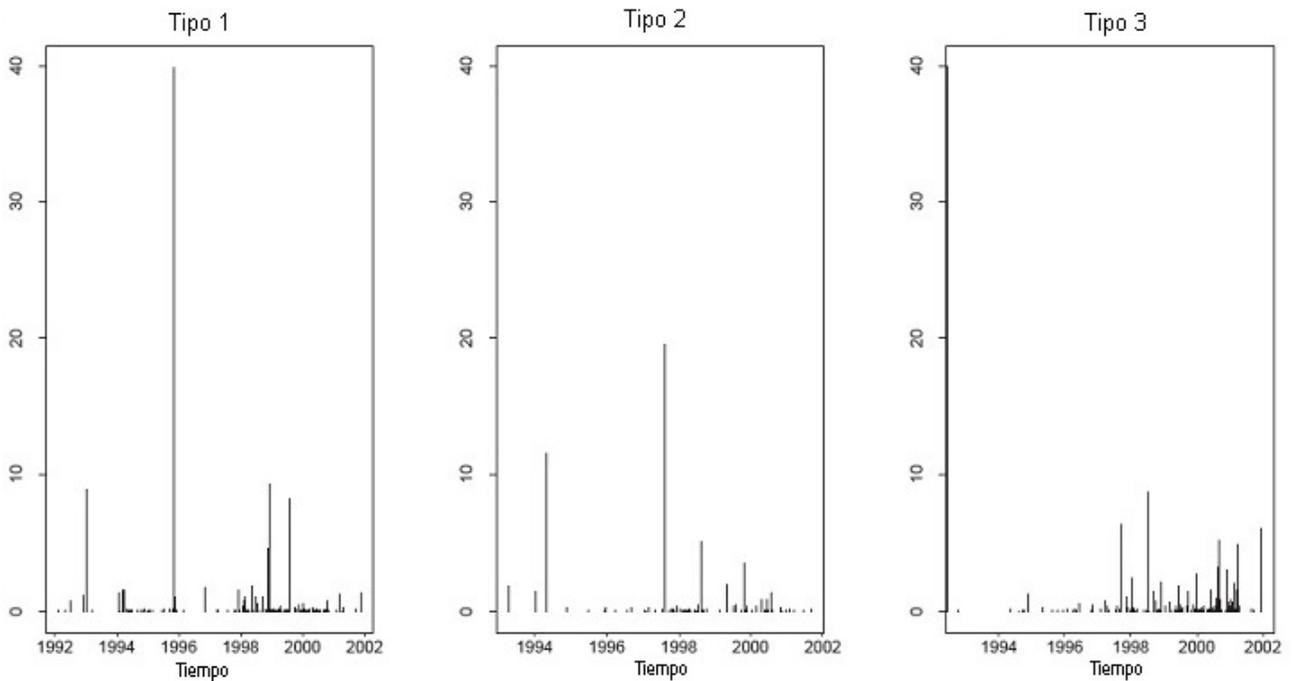


Figura 4.1: Pérdidas por riesgo operacional. De izquierda a derecha: Tipo 1 ($n = 162$), Tipo 2 ($n = 80$), Tipo 3 ($n = 175$).

De la figura 4.1, se pueden observar los siguientes hechos:

- el período histórico es relativamente pequeño, sólo 10 años de datos;
- las magnitudes de las pérdidas muestran extremos;
- las veces que ocurren pérdidas están espaciadas irregularmente en el tiempo, y
- la frecuencia incrementa con respecto al tiempo con un cambio radical alrededor de 1998.

El último punto destaca que los datos son no estacionarios. La discontinuidad en 1998 puede deberse a que la cuantificación del riesgo operacional tomó importancia a finales de la década de los noventa. Estos cambios también pudieron deberse a cambios internos o a cambios en el entorno económico, político o regulatorio en el que el banco opera.

De las observaciones anteriores, se sigue que, al considerar el análisis y cuantificación del riesgo operacional, se deberían tener en consideración el tipo de riesgo y la no estacionariedad en la ocurrencia de las pérdidas. Debido a esto, para la utilización de métodos de valores extremos en el estudio del riesgo operacional, se debe adaptar el método POT, permitiendo que los parámetros del modelo dependan del tiempo.

Con el objetivo de obtener una muestra más grande, se agrupan en una muestra los tres tipos de pérdidas, como se muestra en la siguiente figura

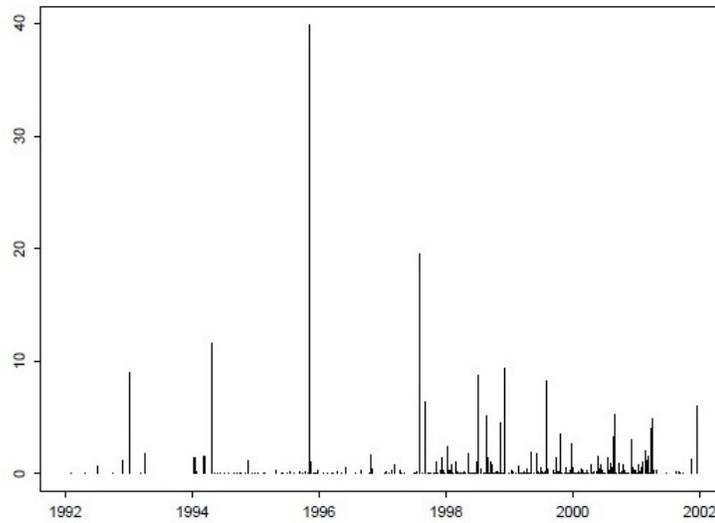


Figura 4.2: Los tres tipos de pérdidas por riesgo operacional agrupados ($n = 417$).

Ahora bien, recordemos del capítulo 3 el método POT.

Denotemos las pérdidas por X_1, \dots, X_n , u un umbral (que permite destacar los valores inusualmente altos de la variable) y los excesos correspondientes por Y_1, \dots, Y_{N_u} , donde $Y_k = X_{i_k} - u$ para $k = 1, \dots, N_u$, $i_k = 1, \dots, n$ y $X_{i_k} > u$.

Para pérdidas i.i.d. y u dado (lo suficientemente grande), se tiene que:

- los excesos Y_1, \dots, Y_{N_u} siguen la distribución generalizada de Pareto $G_{\xi, \beta}$:

$$G_{\xi, \beta}(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\beta}\right)^{-\frac{1}{\xi}}, & \text{si } \xi \neq 0, \\ 1 - e^{-\frac{y}{\beta}}, & \text{si } \xi = 0, \end{cases} \quad (4.1)$$

- el número de excesos sobre el umbral u , N_u , sigue aproximadamente un proceso de Poisson homogéneo con intensidad $\lambda > 0$, es decir, $N_u \sim \mathcal{P}(\lambda)$ con $\lambda = E[N_u]$, de acuerdo con Leadbetter (1991, [20]);
- las magnitudes de los excesos sobre el umbral u son iid e independientes de las veces que ocurren los excedentes.

Luego, se pueden estimar los parámetros λ, ξ y β , a partir de la siguiente función de log-verosimilitud:

$$\begin{aligned} l(\lambda, \xi, \beta) &= \ln \left(\frac{\lambda^{N_u} e^{-\lambda}}{N_u!} \prod_{i=1}^{N_u} g_{\xi, \beta}(y_i) \right) \\ &= N_u \ln(\lambda) - \lambda - N_u \ln \beta - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^{N_u} \ln \left(1 + \frac{\xi}{\beta} y_i \right) \\ &= l(\lambda) + l(\xi, \beta). \end{aligned}$$

La última igualdad sugiere que se pueden estimar los parámetros de la GPD y la intensidad de Poisson por separado.

En la siguiente sección se verá que el parámetro λ modela el incremento en la intensidad de las pérdidas que se puede observar en la figura 4.2, permitiendo que el parámetro dependa del tiempo.

4.2. Método POT avanzado

El método POT clásico se puede extender a un modelo más dinámico permitiendo que los parámetros λ, ξ, β dependan del tiempo y de otras variables con el fin de tomar en consideración el carácter no estacionario.

Considerando los datos representados por la figura 4.2, se fija un umbral $u = 0,4$, de acuerdo con Chavez-Demoulin y Embrechts (2004,[8]). Luego, siguiendo una metodología no paramétrica, se ajustan diferentes modelos para λ, ξ y β , permitiendo:

- dependencia funcional del tiempo $g(t)$, donde t se refiere al año sobre el dominio de estudio;

- dependencia de T , donde T define el tipo de pérdida de datos, modelado mediante el indicador I_T :

$$I_T = \begin{cases} 1, & \text{si } Tipo = T, \\ 0, & \text{en otro caso,} \end{cases}$$

donde $T = 1, 2, 3$, y

- discontinuidad modelada mediante el indicador $I_{(t>t_c)}$, donde $t_c = 1998$ es el año de punto de cambio o cambio de régimen e

$$I_{(t>t_c)} = \begin{cases} 1, & \text{si } t > t_c, \\ 0, & \text{si } t \leq t_c. \end{cases}$$

Siguiendo un modelo generalizado aditivo, se ajusta a cada parámetro λ, ξ y β un modelo general semiparamétrico como sigue

$$\begin{aligned} \lambda(t) &= e^{\mathbf{x}'\boldsymbol{\alpha}_\lambda + g_\lambda(t)}, \\ \xi(t) &= \mathbf{x}'\boldsymbol{\alpha}_\xi + g_\xi(t), \\ \nu(t) &= e^{\mathbf{x}'\boldsymbol{\alpha}_\nu + g_\nu(t)}, \end{aligned} \tag{4.2}$$

donde \mathbf{x}' es un vector fila de variables independientes, $\boldsymbol{\alpha}_\lambda, \boldsymbol{\alpha}_\xi, \boldsymbol{\alpha}_\beta$ son vectores columna de coeficientes y $g_\lambda, g_\xi, g_\beta$ son funciones suaves.

En (4.2), $\nu(t)$ está dada por la reparametrización $\nu(\xi, \beta) = \beta(1+\xi)$, para así evitar dificultades computacionales. [6]

Ahora bien, una directa maximización de la función de log-verosimilitud ya no es apropiada como un método de estimación, pues conduce a un fenómeno conocido como *sobreajuste del modelo*, en el cual el modelo se ajusta muy bien a los datos existentes pero tiene un pobre rendimiento para predecir nuevos resultados. Ante esto, se puede maximizar en su lugar la log-verosimilitud menos un término que regula el grado de suavidad de las funciones g_λ, g_ξ, g_ν . Esta nueva función a maximizar es conocida como *log-verosimilitud penalizada*.

La log-verosimilitud penalizada de λ es

$$l(\lambda) - \frac{1}{2} \gamma_\lambda \int g_\lambda''(t)^2 dt$$

y la log-verosimilitud penalizada de ξ, ν es

$$l(\xi, \nu) - \frac{1}{2} \gamma_\xi \int g_\xi''(t)^2 dt - \frac{1}{2} \gamma_\nu \int g_\nu''(t)^2 dt,$$

donde $\gamma_\lambda, \gamma_\xi, \gamma_\nu$ son parámetros que controlan el grado de suavidad de las funciones g_λ, g_ξ, g_ν .

Aún cuando la implementación de este método de estimación involucra algoritmos y técnicas estadísticas cuyo estudio está fuera del alcance de esta tesis, se presenta un resumen de los resultados obtenidos en Chavez-Demoulin (2004, [7]), empleando los datos señalados en Chavez-Demoulin y Embrechts (2004, [8]).

Aplicación del método POT avanzado

Chavez-Demoulin y Embrechts (2004, [8]) aplican diferentes modelos a cada parámetro λ, ξ y β con la forma de las relaciones dadas en (4.2) y los comparan usando pruebas basadas en el estadístico de la razón de verosimilitud y un criterio conocido como criterio de información de Akaike (para más información sobre este criterio ver McQuarrie y Tsai [23]).

Finalmente, los modelos que Chavez-Demoulin y Embrechts (2004, [8]) seleccionan para ξ y β dependen solamente del tipo de pérdida T . Mientras que el modelo que seleccionan para λ depende del tiempo t y del tipo T y está dado por:

$$\log \hat{\lambda}(t, T) = \hat{\gamma}_T I_T + \hat{\sigma} I_{(t > t_c)} + \hat{g}(t). \tag{4.3}$$

Es importante resaltar que, si bien se consideran los datos agrupados, los modelos seleccionados permiten separar los tipos de pérdidas.

La inclusión del primer componente $\hat{\gamma}_T I_T$ en (4.3) indica que el tipo de pérdida T es importante para modelar la intensidad de Poisson, esto es, el número de excedentes sobre el umbral difiere significativamente para cada tipo de pérdida 1, 2 o 3, como se verá mas adelante.

El modelo seleccionado para λ también contiene el indicador de discontinuidad $I_{(t > t_c)}$, se obtuvo que $\hat{\sigma} = 0,47$ para $t > t_c$ y que la intensidad en promedio es bastante diferente

antes y después de 1998.

Por otra parte, la presencia de la función $g_\lambda(t)$ indica que el parámetro de intensidad λ depende del tiempo.

La figura 4.3, tomada de Chavez-Demoulin (2004, [7]), representa la intensidad estimada resultante, $\hat{\lambda}$, para cada tipo de pérdida y su intervalo de confianza del 95 %. Se puede observar que la curva resultante refleja bastante bien el comportamiento del número de excedentes (los puntos de las gráficas) para cada tipo.

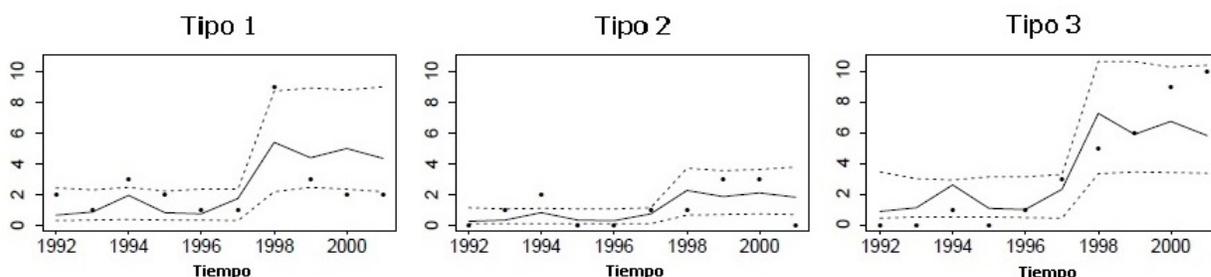


Figura 4.3: Intesidad de Poisson estimada $\hat{\lambda}$ y su intervalo de confianza del 95 % para pérdidas de tipo 1, 2 y 3. Los puntos son los números de excedentes anuales sobre $u = 0,4$.

En la figura 4.4, se representan los estimadores $\hat{\xi}(T)$, $\hat{\beta}(T)$ y sus intervalos de confianza del 95 %. Para los tipos 1 y 2, el estimador del parámetro de forma $\hat{\xi}$ es aproximadamente 0,7 y es significativamente más pequeño para el tipo 3 (aproximadamente 0,3), esto sugiere una distribución de pérdida para el tipo 3 de cola menos pesada que para los tipos 1 y 2; pues mientras más grande sea el parámetro de forma, más pesada será la cola.

El efecto debido al punto de cambio en 1998 no se conserva en los modelos de ξ y β , es decir, la distribución del tamaño de la pérdida no cambia alrededor de 1998.

Un posible método general para evaluar lo bien que se ajusta el modelo para los parámetros ξ y β de la GPD está basado en el resultado de que, dado que el modelo es correcto, los siguientes *residuos*

$$R_i = \frac{1}{\hat{\xi}} \log \left(1 + \frac{\hat{\xi}}{\hat{\beta}} Y_i \right), \quad i = 1, \dots, N_u,$$

se distribuyen aproximadamente como variables aleatorias exponenciales independientes.

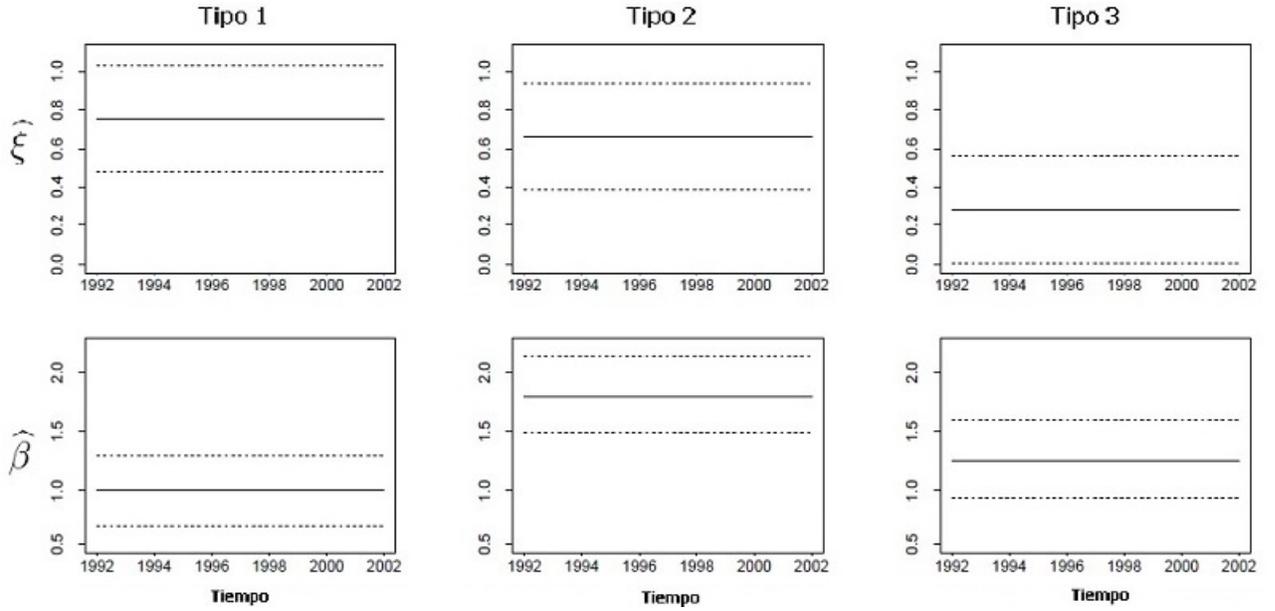


Figura 4.4: Parámetros de la GPD estimados $\hat{\xi}, \hat{\beta}$ y sus intervalos de confianza del 95 % para pérdidas de tipo 1, 2 y 3.

En efecto, si los excesos Y_i siguen una GPD de parámetros $\hat{\xi}, \hat{\beta}$ para $i = 1, \dots, N_u$, entonces

$$\begin{aligned}
 F(r_i) &= \mathbb{P}(R_i \leq r_i) \\
 &= \mathbb{P}\left(\frac{1}{\hat{\xi}} \log\left(1 + \frac{\hat{\xi}}{\hat{\beta}} Y_i\right) \leq r_i\right) \\
 &= \mathbb{P}\left(1 + \frac{\hat{\xi}}{\hat{\beta}} Y_i \leq e^{\hat{\xi} r_i}\right) \\
 &= \mathbb{P}\left(Y_i \leq \frac{\hat{\beta}}{\hat{\xi}} (e^{\hat{\xi} r_i} - 1)\right) \\
 &= 1 - \left(1 + \frac{\hat{\xi}}{\hat{\beta}} \left[\frac{\hat{\beta}}{\hat{\xi}} (e^{\hat{\xi} r_i} - 1)\right]\right)^{-1/\hat{\xi}}, \quad \text{pues } Y_i \sim GPD(\hat{\xi}, \hat{\beta}) \\
 &= 1 - \left(e^{\hat{\xi} r_i}\right)^{-1/\hat{\xi}} \\
 &= 1 - e^{-r_i},
 \end{aligned}$$

por lo tanto, R_i se distribuye exponencialmente con parámetro $\lambda = 1$.

En la figura 4.5 se comparan los residuos contra los cuantiles de una distribución exponencial, usando los estimadores $\hat{\xi}(T)$ y $\hat{\beta}(T)$ para los tres tipos de pérdida agrupados. Este gráfico sugiere que el modelo es razonable pues, efectivamente, los residuos parecen aproximarse a una distribución exponencial.

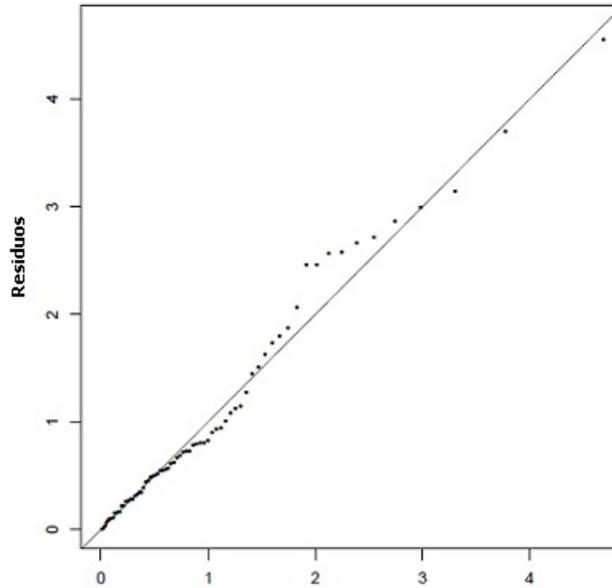


Figura 4.5: Gráfico de los residuos contra los cuantiles de una distribución exponencial.

Ahora bien, usando los modelos seleccionados para $\lambda(t, T)$, $\xi(t, T)$ y $\beta(t, T)$, se pueden predecir los valores $\lambda(t + 1, T)$, $\xi(t + 1, T)$ y $\beta(t + 1, T)$ para cada tipo T y, así, poder estimar la esperanza condicional de la pérdida total para el año 2002, dado que la pérdida es mayor al 99% VaR , es decir, el déficit esperado al 99%.

Así, si los estimadores de VaR_α y ES_α para los parámetros λ, ξ y β vienen dados por:

$$\widehat{VaR}_\alpha(X) = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{1 - \alpha}{\hat{\lambda}} \right)^{-\hat{\xi}} - 1 \right), \tag{4.4}$$

$$\widehat{ES}_\alpha(X) = \left(\frac{1}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} u}{(1 - \hat{\xi}) \widehat{VaR}_\alpha(X)} \right) \widehat{VaR}_\alpha(X). \tag{4.5}$$

entonces, sustituyendo $\hat{\lambda}, \hat{\xi}$ y $\hat{\beta}$ por $\hat{\lambda}(t + 1, T), \hat{\xi}(T)$ y $\hat{\beta}(T)$ con $t = 2001$, es posible estimar el $VaR_{0,99}$ y $ES_{0,99}$ para el año 2002.

Dado que el modelo seleccionado para λ depende del tiempo t , las medidas de riesgo se consideran “dinámicas” y, dado que los modelos seleccionados también dependen del tipo de pérdida T , entonces las medidas de riesgo para el tiempo t y tipo T se denotan por

$$dVaR_{0,99}^T(t) \quad y \quad dES_{0,99}^T(t).$$

En la tabla 4.1 se pueden observar los valores estimados del 99% VaR y del 99% ES para cada tipo de pérdida 1, 2 y 3 en el año 2002.

	$\widehat{dVaR}_{0,99}^T(2002)$	$\widehat{dES}_{0,99}^T(2002)$
T=1	40.4	166.4
T=2	48.4	148.5
T=3	11.9	18.8

Tabla 4.1: $VaR_{0,99}$ dinámico y $ES_{0,99}$ dinámico estimados para cada tipo de pérdida en el año 2002.

El valor $\widehat{dVaR}_{0,99}^{T=1}(2002) = 40,4$ indica que la estimación de la pérdida total de tipo 1 en el año 2002 es de 40,4 a un nivel de confianza del 99%. Así mismo, se puede observar que la estimación de la pérdida total de tipo 3 es aproximadamente 12, lo cual es significativamente más pequeño que las pérdidas estimadas para los tipos 1 y 2.

La gran diferencia que se puede observar entre $\widehat{dVaR}_{0,99}^{T=1}(2002)$ y $\widehat{dES}_{0,99}^{T=1}(2002)$ hace difícil la decisión de cuál medida será considerada para medir el capital mínimo requerido que debe tener la entidad bancaria para cubrir posibles pérdidas por riesgo operacional. Se pudiera pensar que considerar el VaR es más conveniente por requerirse menor capital adicional, sin embargo, aunque el capital requerido sea mucho menor en este caso, hay que tener en cuenta que no se cubriría la pérdida si ésta supera el VaR.

También se puede observar que el 99% ES para el tipo 3 ($\widehat{dES}_{0,99}^{T=1}(2002) = 18,8$) es significativamente más pequeño que el 99% ES para los tipos 1 y 2, lo cual indica que el capital requerido para las pérdidas de tipo 3 es mucho menor que para las pérdidas de tipo 1 y 2.

Si no se hace distinción entre los tipos de pérdida se tendría un único valor estimado del VaR (o ES) para toda la muestra, el cual podría ser menor que el valor estimado para pérdidas de tipo 1 y, por lo tanto, el banco sólo estaría cubriéndose ante posibles pérdidas del tipo 1 o 2.

Esta es una razón por la cual usar modelos que incluyen covariables es importante.

En cierto sentido, este modelo adaptado hace un mejor uso de los datos disponibles, lo cual es importante, en especial en el contexto del riesgo operacional.

Conclusión

El riesgo operacional ha venido tomando especial importancia desde hace algunos años. Con el acuerdo Basilea II, se incrementa el interés por un manejo explícito de este riesgo. Ante esto, viene siendo necesario el estudio de modelos flexibles de eventos extremos adaptados al riesgo operacional

Se presentan evidencias de que el ajuste de la distribución generalizada de Pareto a las pérdidas que superen los umbrales altos es un método útil para la estimación de las colas de las distribuciones de las pérdidas. Sin embargo, se ha visto que las características de datos históricos de pérdidas por riesgo operacional generalmente no concuerdan con la hipótesis de independencia, pues están espaciadas irregularmente en el tiempo.

También se ha visto que los datos de pérdidas por riesgo operacional son escasos y que una solución puede ser agrupar los datos de distintos tipos de riesgo operacional para tener una muestra más grande, permitiendo el uso de covariables, es decir, variables que inciden en los resultados, para representar estos tipos de riesgo. La agrupación de los datos tiene, además, la ventaja de permitir pruebas de interacciones entre los tipos de pérdida.

Como se ha explicado, es posible adaptar el método POT para tener en consideración la no estacionariedad y las covariables, permitiendo que los parámetros del modelo dependan del tiempo y del tipo de riesgo.

El considerar los tipos de riesgo en la aplicación de este método adaptado sugiere que se usaron eficientemente los datos disponibles. También se confirmó que la intensidad de las pérdidas presenta una discontinuidad en 1998 y que el número de los

excesos de las pérdidas dependen del tiempo, mientras que la magnitud de las pérdidas depende únicamente del tipo de pérdida.

El uso de un método adaptado de valores extremos considerando la no estacionariedad (dependencia del tiempo) y otras variables (tipos de riesgo) provee una técnica de exploración que es conveniente y flexible que podrá ser mejorada con el incremento de bases de datos.

Se estimaron, para cada tipo de riesgo, el déficit esperado y el valor en riesgo para predecir el valor esperado de la pérdida para cierto año. Los resultados observados resaltan la importancia de distinguir los tipos de riesgo. Utilizando los valores históricos estimados del VaR, es posible comprobar si se sostiene la hipótesis de que el método calcula correctamente las medidas de riesgo. Sin embargo, esta prueba necesitaría muchos más datos históricos.

Bibliografía

- [1] ACERBI, C. Y TASCHE, D. *On the coherence of Expected Shortfall*, vol. 26. Journal of Banking and Finance, 2002.
- [2] ARRIOJAS, M. *Teoría de las Probabilidades*. Escuela de Matemática, UCV, 2015. http://www.matematica.ciens.ucv.ve/materias/files_materias/Probabilidades.
- [3] ARTZNER, P., DELBAEN, F., EBER, J.-M. Y HEATH, D. *Coherent measures of risk*, vol. 9. Mathematical finance, 1999.
- [4] BALKEMA, A. Y DE HAAN, L. *Residual life time at great age*. Annals of Probability, 1974.
- [5] CEA. *Solvency II Glossary*. Comité Europeo de Seguros, 2007.
- [6] CHAVEZ-DEMOULIN, V. Y DAVISON, A. *Generalized additive modelling of sample extremes*, vol. 54. Journal of the Royal Statistical Society: Series C (Applied Statistics), 2005.
- [7] CHAVEZ-DEMOULIN, V. Y EMBRECHTS, P. *Advanced extremal models for operational risk*. Department of Mathematics, ETH, Zurich, 2004.
- [8] CHAVEZ-DEMOULIN, V. Y EMBRECHTS, P. *Smooth extremal models in finance and insurance*. No. 2. Journal of Risk and Insurance, 2004.
- [9] COLES, S. *An introduction to statistical modeling of extreme values*, vol. 208. Springer, 2001.

- [10] EMBRECHTS, P. Y HOFERT, M. *A note on generalized inverses*, vol. 77. Mathematical Methods of Operations Research, 2013.
- [11] EMBRECHTS, P., KAUFMANN, R. Y SAMORODNITSKY, G. *Ruin theory revisited: stochastic models for operational risk*. 2004.
- [12] EMBRECHTS, P., KLÜPPELBERG, C. Y MIKOSCH, T. *Modelling extremal events for insurance and finance*, vol. 33. Springer Verlag, 1997.
- [13] EMBRECHTS, P., RESNICK, S. I. Y SAMORODNITSKY, G. *Extreme value theory as a risk management tool*, vol. 3. North American Actuarial Journal, 1999.
- [14] GÓMEZ, J. G. M. *La teoría de valor extremo y el riesgo operacional: una aplicación en una entidad financiera*, vol. 8. Revista Ingenierías, Universidad de Medellín, 2009.
- [15] GREEN, P. J. Y YANDELL, B. S. *Semi-parametric generalized linear models*. Springer, 1985.
- [16] GRINSTEAD, C. M. Y SNELL, J. L. *Introduction to probability*. American Mathematical Society, 1997.
- [17] GUMBEL, E. J. *Statistics of Extremes*. Columbia University Press, New York, 1958.
- [18] HARDY, M. R. *An introduction to risk measures for actuarial applications*. Society of Actuaries, 2006.
- [19] KOTZ, S. Y NADARAJAH, S. *Extreme Value Distributions: Theory and Applications*. Imperial College Press, 2000.
- [20] LEADBETTER, M. R. *On a basis for 'Peaks over Threshold' modeling*, vol. 12. Statistics & Probability Letters, 1991.
- [21] MCNEIL, A. J., FREY, R. Y EMBRECHTS, P. *Quantitative risk management: Concepts, techniques and tools*. Princeton university press, 2005.

- [22] MCNEIL, A. J. Y SALADIN, T. *The peaks over thresholds method for estimating high quantiles of loss distributions*. 1997.
- [23] MCQUARRIE, A. D. R. Y TSAI, C. L. *Regression and time series model selection*. World Scientific, 1998.
- [24] MELO, L. F. Y BECERRA, S. R. *Medidas de riesgo, características y técnicas de medición*. Centro Editorial Universidad del Rosario. Colombia, 2006.
- [25] MORA, A. *Consideraciones para la estimación de cuantiles altos en el riesgo operativo*. No. 1. Análisis–Revista del Mercado de Valores, 2010.
- [26] OLIVARES, M. M. *Curso de Probabilidades*. Escuela de Matemática, UCV, 2002.
- [27] ORTEGA, J. *Introducción a la Teoría de Valores Extremos*. Octavo Congreso Latinoamericano de Sociedades de Estadística, Montevideo, 2008.
- [28] PICKANDS, J. *Statistical inference using extreme order statistics*. Annals of Statistics, 1975.