



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

TEORIA DE VALORES EXTREMOS CON APLICACIONES A MEDIDAS DE RIESGO

295 615
519 562

T E S I S

QUE PARA OBTENER EL TITULO DE:

M A T E M A T I C A

P R E S E N T A :

MARIA ANTONIETA CAMPA ROJAS



FACULTAD DE CIENCIAS UNAM

DIRECTORA DE TESIS:

DRA. MARIA ASUNCION BEGOÑA FERNANDEZ FERNANDEZ





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



REPUBLICA NACIONAL
ESTADOS UNIDOS MEXICANOS
MEXICO

M. EN C. ELENA DE OTEYZA DE OTEYZA
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis:

"Teoría de Valores Extremos con Aplicaciones a Medidas de Riesgo"

realizado por **María Antonieta Campa Rojas**

con número de cuenta **9653463-2**, pasante de la carrera de **Matemáticas**

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis **Dra. María Asunción Begoña Fernández Fernández.**

Propietario **Dr. Manuel Galán Medina.**

Propietario **Dr. José María González Barrios.**

Suplente **Mat. Margarita Elvira Chávez Cano.**

Suplente **N. en C. Beatriz Eugenia Rodríguez Fernández.**

Begoña Fernández
Manuel Galán Medina
José María González Barrios
Margarita Elvira Chávez Cano
Beatriz Eugenia Rodríguez Fernández

Consejo Departamental de Matemáticas

[Firma]
N. en C. **Alejandro Bravo Mojica.**

MATEMÁTICAS

*A Pelu, porque eres el mejor ejemplo que haya podido tener en la vida.
Por tu infinito amor, por tu comprensión, por tu apoyo, por tu constante
motivación y por tantas otras cosas... Mil gracias, mamá.*

Agradecimientos

A mi asesora, la Dra. Begoña Fernández Fernández, por su paciencia, dedicación y apoyo para la realización de este trabajo. Gracias por su tiempo, por compartir sus conocimientos y por su ayuda en todos los ámbitos.

A los sinodales, el Dr. Manuel Galán, el Dr. José María González Barrios, la M. en C. Beatriz Rodríguez y la Mat. Margarita Chávez Cano, por sus valiosos comentarios y correcciones en la revisión de este trabajo, no obstante la premura del tiempo.

Al Dr. Paul Embrechts, de ETH, Zurich, por su apoyo, conocimientos y por todo el material que hizo posible el desarrollo de este trabajo.

Al Dr. Alexander McNeil, de ETH, Zurich, por haberme dado el tiempo necesario para resolver mis dudas sobre la utilización de su programa *EVIS*, el cual fue una herramienta que facilitó enormemente la realización de este trabajo.

A mi abue, a mi papá, a Lorena, a Elvira y al resto de la familia, por su apoyo incondicional y por sus enseñanzas.

A Erick, por su gran cariño y comprensión, por haberme acompañado a través de los años y por su apoyo en la revisión final de este trabajo. Gracias, Epa, por todo.

A la Happy Corner, a Javier Herrera, a Melissa y a todos los amigos de la Facultad de Ciencias, por su amistad, comprensión y ayuda, que hicieron más fácil y agradable el logro de este objetivo.

A todas aquellas personas que en este momento escapan de mi mente, pero que, de una u otra manera, me han ayudado a alcanzar esta meta.

Índice General

1 Preliminares	1
1.1 Convergencia de Máximos.	4
1.2 Las Distribuciones de Valores Extremos y el Teorema de Fisher-Tippet.	14
1.3 Dominio de Atracción del Máximo de las Distribuciones de Valores Extremos.	29
1.3.1 El Dominio de Atracción del Máximo de la Distribución Fréchet	31
1.3.2 El Dominio de Atracción del Máximo de la Distribución Weibull	38
1.3.3 El Dominio de Atracción del Máximo de la Distribución Gumbel	43
1.4 La Distribución Pareto Generalizada (DPG) para Aproximar Excesos sobre un Umbral	57
2 Medidas de Riesgo	69
2.1 Propiedades Descendentes de Medidas de Riesgo	72
2.2 La Varianza	75
2.2.1 Ejemplos de Problemas al Usar la Varianza	77
2.3 Valor en Riesgo	87
2.3.1 Ejemplos de Problemas al Utilizar el VaR	91

2.4	Esperanza Condicional de la Cola	95
2.5	El Mundo Elíptico y la Medición de Riesgos	101
3	Medidas de Riesgo y Valores Extremos	111
3.1	La TVE puesta en acción	117
3.2	Modelo de Picos Sobre el Umbral (POT)	119
3.2.1	Estimador de Máxima Verosimilitud para los parámetros de la DPG	126
3.3	Modelo de Máximos de Bloque	128
3.4	Estimación de los Parámetros para la distribución VEG	134
3.4.1	El Estimador de Máxima Verosimilitud (EMV)	135
3.4.2	El Estimador de Hill (EH)	136
3.4.3	El Estimador Deckers-Einmahl-de Haan (EDEdH)	139
4	Análisis de Datos	145
4.1	Análisis Exploratorio de Datos	147
4.1.1	Para Empezar	147
4.1.2	Gráficas de Probabilidad y de Cuantiles (PP-Plot y QQ-Plot)	149
4.1.3	La Función Media de Exceso	152
4.1.4	Observaciones sobre el Análisis Exploratorio de Datos	154
4.2	Caso 1: Pérdidas Industriales por Incendios en Dinamarca	155
4.2.1	Análisis Preliminar de Datos de Pérdidas por Incendios en Di- namarca	157
4.2.2	Análisis de Máximos de Bloque para Datos Daneses	160

4.2.3	Análisis POT para Datos Daneses	164
4.2.4	Comparación de Cuantiles Obtenidos con el Método Máximos de Bloque y con el Método POT para Pérdidas por Incendios en Dinamarca	175
4.3	Caso 2: Datos acción BMW	177
4.3.1	Análisis Preliminar de Datos de Pérdidas en el Precio de la Acción BMW	177
4.3.2	TVE para Series Estacionarias	182
4.3.3	Análisis de Máximos de Bloque para Pérdidas en el Precio de la Acción BMW	188
4.3.4	Análisis POT para Pérdidas en el Precio de la Acción BMW .	194
4.3.5	Comparación de Cuantiles Obtenidos con el Método Máximos de Bloque y con el Método POT para Pérdidas en el Precio de la Acción BMW	200
4.4	Caso 3: Tipo de cambio FIX peso/dólar	201
4.4.1	Análisis Preliminar de Datos de Pérdidas en el Tipo de Cambio FIX	202
4.4.2	Análisis de Máximos de Bloque para el Tipo de Cambio FIX .	206
4.4.3	Análisis POT para Pérdidas en el Tipo de Cambio FIX	211
4.4.4	Comparación de Cuantiles Obtenidos con el Método Máximos de Bloque y con el Método POT para Pérdidas en el Tipo de Cambio FIX	217
4.5	Conclusión	219

Introducción

El constante desarrollo de los mercados, productos y competidores, aunado a cambios en las disposiciones legales, han dado lugar en nuestro país a la integración de grupos financieros con las características de la llamada Banca Universal. Por otra parte, el proceso de apertura que ha experimentado México y que puede observarse en la participación del capital extranjero en las instituciones financieras, es un reflejo de los grandes cambios que en fechas recientes ha experimentado la industria financiera, tanto en el área bancaria como en la de seguros. Estos cambios, que no son exclusivos de México sino que, por el contrario, están teniendo lugar en todo el mundo, obedecen al fenómeno de globalización que estamos viviendo, y se manifiestan también mediante el surgimiento de productos financieros más complejos. Al mismo tiempo, la industria de reaseguro ha sufrido un incremento tanto en la magnitud como en la intensidad de las pérdidas debidas a catástrofes naturales o causadas por el hombre. Ante esto, las instituciones financieras buscan cada vez más tener y poder ofrecer a sus clientes productos “securitizados” de manera apropiada que les permitan lidiar con eventos catastróficos. Por ello, la transferencia alternativa de riesgos y “securitización” se ha convertido en una importante área de investigación en las industrias bancaria y aseguradora.

Las grandes catástrofes, sean de índole natural —como tremendas inundaciones o fuertes terremotos— o causadas por el hombre —tales como *crashes* bursátiles o grandes pérdidas industriales, digamos, por incendios provocados— son manifestaciones de eventos extremos. La posibilidad de ocurrencia de estos eventos entraña riesgos que las instituciones financieras deben considerar.

La meta de este trabajo es la obtención de estimadores de medidas de riesgo extremo que estén basados en un modelo matemático sólido. Para ello, usaremos el enfoque matemático estándar para modelar riesgos, el cual utiliza el lenguaje de la Teoría de Probabilidad. El planteamiento probabilístico es el siguiente: tenemos la pareja (Ω, \mathcal{A}) , donde $\Omega \neq \emptyset$ es un conjunto arbitrario y \mathcal{A} es un σ -álgebra de Ω . Pensamos en cada $\omega \in \Omega$ como un estado del mundo en el futuro; suponemos que todos los participantes están de acuerdo en los posibles estados futuros del mundo, pero pueden no estar de acuerdo en la probabilidad de ocurrencia de cada uno de ellos.

Usaremos notación matemática sencilla: un riesgo X_i tiene distribución marginal $F_i(x) = P\{X_i \leq x\}$. Un vector de riesgos tiene la función de distribución conjunta $F(x_1, \dots, x_n) = P\{X_1 \leq x_1, \dots, X_n \leq x_n\}$. Puesto que F_i y F son las funciones de distribución de los valores potenciales de un riesgo, estas son distribuciones que nunca podremos observar exactamente.

Sabemos, por otra parte, que las instituciones financieras tienen clasificaciones para diferentes tipos de riesgo, de manera que si consideramos X_1, X_2, \dots variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.), con función de distribución F desconocida, X_i puede representar, por ejemplo, cualquiera de los si-

güentes conceptos:

- La pérdida en la transacción i -ésima.
- La ganancia en el portafolio i -ésimo.
- El i -ésimo valor absoluto del logaritmo del rendimiento de un instrumento financiero subyacente (como aproximación al rendimiento porcentual $\left| \frac{S_t - S_{t-1}}{S_{t-1}} \right|$).
- El i -ésimo reclamo relativo a una pérdida en seguros.
- La i -ésima pérdida crediticia.

Puesto que, matemáticamente, los eventos extremos ocurren cuando un riesgo toma valores de la cola de su distribución, una rama de la Teoría de la Probabilidad que cobra gran importancia cuando se trata de modelar eventos extremos es la Teoría de Valores Extremos (TVE). La TVE proporciona métodos que permiten cuantificar los eventos extremos y sus consecuencias de manera óptima desde el punto de vista estadístico. Esto hace que la TVE sea particularmente útil en el área de Administración de Riesgos, donde, de acuerdo con la definición de la Comisión Nacional Bancaria y de Valores, la Administración de Riesgos es *el conjunto de objetivos, políticas, procedimientos y acciones que se implementan para identificar, medir, monitorear, limitar, controlar, informar y revelar los distintos tipos de riesgo a que se encuentran expuestas las instituciones*.¹

El administrador de riesgos se encarga de estimar probabilidades de colas y altos cuantiles de distribuciones de pérdida y ganancia (P&L) y de datos financieros en

¹ Tomado de la Circular 1473 del 17 de julio de 2000, de la Comisión Nacional Bancaria y de Valores.

general. Es por ello que en Administración de Riesgos se busca modelar el riesgo de manera que se contemple la posibilidad de un resultado extremo. Al usar este modelo, se quiere medir el riesgo con una medida que proporcione información acerca del resultado extremo.

Además, puesto que, desde el punto de vista de un administrador de riesgos, el tamaño de las pérdidas resulta tan importante como el saber qué tan frecuentemente ocurren, la TVE se ha convertido en una herramienta de gran importancia para las instituciones financieras, ya que proporciona métodos para estimar probabilidades relacionadas con eventos-cola, es decir, dado que se ha incurrido en una pérdida mayor que cierta cantidad, nos indica qué tan grande esperamos que sea este exceso. Esta es la idea básica de la medida de riesgo extremo conocida como Esperanza Condicional de la Cola (ES) que, como veremos en el desarrollo de este trabajo, satisface las propiedades deseables de una medida de riesgo: monotonía, subaditividad, homogeneidad positiva e invariancia bajo translaciones; por su parte, el Valor en Riesgo (VaR) y la Varianza en general no cumplen con una de estas propiedades: la subaditividad. Así, el uso de la Varianza o del VaR puede ocasionar problemas en los casos en que un portafolio se subdivide en varios subportafolios, pues es posible que la suma del VaR (o de la varianza) correspondiente a los subportafolios sea menor que el VaR (o la varianza) del portafolio total. A pesar de este inconveniente, el VaR es una medida de riesgo muy usada en la práctica, por lo que en este trabajo expondremos métodos para estimar tanto la ES como el VaR, utilizando la TVE.

Otro punto a favor de la TVE es que permite cuantificar los riesgos extremos de una manera más adecuada que dos de los enfoques utilizados anteriormente: la

Simulación Histórica (SH) y los modelos que utilizan aproximaciones normales. En el enfoque de SH la distribución P&L estimada para el portafolio queda dada simplemente por la distribución empírica de las ganancias y pérdidas anteriores de este portafolio. El método es por tanto fácil de implementar y evita la suposición de una forma específica para la distribución P&L, pero tiene una seria desventaja: los cuantiles extremos son notoriamente difíciles de estimar, puesto que la extrapolación más allá de las observaciones anteriores es imposible y los estimadores de cuantiles extremos dentro de la muestra tienden a ser muy ineficientes, pues el estimador está sujeto a una alta varianza. Por su parte, la principal debilidad de los modelos que suponen normalidad es que esta suposición en general no es viable para datos reales, los cuales, particularmente en finanzas y seguros, tienden a mostrar colas más pesadas, que no pueden modelarse utilizando la distribución Normal. En cambio, los métodos basados en la TVE tienen dos características que los hacen atractivos para la estimación de colas: están basados en una fuerte teoría estadística y ofrecen una forma paramétrica para la cola de la distribución. Así, estos métodos permiten cierta extrapolación más allá del rango de los datos, lo que permite la obtención de estimadores de medidas de riesgo extremo.

La importancia de las medidas de riesgo radica en que, además de ser una base para la determinación del capital de riesgo, ayudan también a que la Dirección de la institución financiera pueda entender mejor el riesgo inherente a su portafolio. Y si bien es imposible que una institución financiera ajuste rápidamente su base de capital a las condiciones cambiantes del mercado, que quedarían reflejadas en el cálculo de una medida de riesgo, lo que la institución sí puede hacer es ajustar el tamaño de su

exposición.

Tenemos pues que la Administración de Riesgos es llevada a cabo por las instituciones financieras (Bancos, Casas de Bolsa y Aseguradoras), tanto con fines de control interno como por órdenes de sus reguladores. En este sentido, la Administración de Riesgos en México ha cobrado un papel mucho más importante en fechas recientes, puesto que por órdenes de la Comisión Nacional Bancaria y de Valores (CNBV), el Consejo Directivo de los Bancos -a partir del 1 de febrero de 1999 - y de las Casas de Bolsa -a partir del 4 de diciembre de 2000- deberá constituir un Comité de Riesgos, cuyo objeto será la administración de los riesgos a que se encuentra expuesta la institución. Este Comité de Riesgos estará encargado de aprobar:

a) La metodología para identificar, medir, monitorear, limitar, controlar, informar y revelar los distintos tipos de riesgos a que se encuentra expuesta la institución; b) los modelos, parámetros y escenarios que habrán de utilizarse para llevar a cabo la medición y el control de los riesgos; c) la realización de nuevas operaciones que por su propia naturaleza conlleven un riesgo.²

Para llevar a cabo la administración de riesgos, el Comité de Riesgos se apoyará en una Unidad para la Administración Integral de Riesgos. Una de las obligaciones de esta Unidad será la realización de pruebas bajo condiciones extremas, que permitan identificar el riesgo que enfrentaría la institución en dichas condiciones y reconocer las posiciones o estrategias que hacen más vulnerable a la propia institución.

La aplicación de modelos basados en la TVE podría ser de gran utilidad para estos Comités de Riesgos, particularmente en la realización de las pruebas extremas

² Tomado de las Circulares 1473, 10-247 y 1423 de la Comisión Nacional Bancaria y de Valores.

por parte de la Unidad para la Administración Integral de Riesgos. El hecho de que la CNBV requiera que se lleven a cabo estas pruebas es una muestra de que los extremos son tremendamente importantes en el mundo financiero. Cabe señalar que este tipo de requisitos no son exclusivos de México: algunos reguladores de la actividad financiera internacional, como el *Bank for International Settlement* (BIS), piden a las instituciones a las que regulan que obtengan estimadores del Valor en Riesgo (VaR) y de la Esperanza Condicional de la Cola (ES). Estos estimadores, para aproximarse más a la realidad, requieren la utilización de métodos que tomen en cuenta que los datos financieros muestran colas pesadas.

Otro hecho importante es que la globalización y las prácticas seguidas a partir del primer Acuerdo de Basilea (1988), han tendido a estandarizar en el mundo prácticas bancarias y también regulatorias. Hasta la fecha, más de 100 países, incluyendo a México, han adoptado las disposiciones de este Acuerdo. Debido a la continua revolución en las prácticas comerciales, en la informática y a la innovación financiera, el Comité de Basilea decidió revisar el Acuerdo de 1988, haciendo énfasis en estrechar la relación que guarda el riesgo que toma un Banco y el capital que requiere tener, proponiendo estímulos para que los Bancos mejoren sus capacidades de administración y de medidas de riesgo. Se busca lograr un acuerdo definitivo para diciembre de 2001 e instrumentarlo a partir del 2004. Este tipo de acuerdos pueden ayudar a mostrar la importancia y extender la práctica de los modelos basados en la TVE.

En palabras del Sr. William Routledge, Vicepresidente Ejecutivo del Banco de la Reserva Federal de Nueva York e integrante del Comité de Basilea: *La administración de riesgos debe ser un proceso dinámico y los reguladores tienen que estar atentos para*

asegurar que sus políticas alienten mayores avances.

En conclusión, sabemos que crear un modelo para un riesgo implica seleccionar una distribución particular. Para hacerlo apropiadamente es necesario, además de utilizar datos históricos, tener una base teórica que fundamente la elección de la función de distribución. En este sentido, la Teoría de Valores Extremos es una herramienta ideal, pues nos indica el tipo de distribución que teóricamente nos permite aproximar la cola de un riesgo y nos proporciona el mejor estimador posible basándose en la información histórica disponible. Para ello, la TVE se basa en los Teoremas de Fisher-Tippet y de Gnedenko-Balkema-Pickands-deHaan (TGBPdH). El primero afirma que las únicas tres posibles distribuciones límite (no-degeneradas) para máximos apropiadamente normalizados de n v.a.i.i.d. son las llamadas distribuciones de valores extremos, es decir, la Fréchet (Φ_α), la Weibull (Ψ_α) y la Gumbel (Λ); en este sentido, el Teorema de Fisher-Tippet es el equivalente —para el caso de máximos de n v.a.i.i.d.— del Teorema de Límite Central. Por su parte, el TGBPdH nos permite aproximar la función de distribución de excesos F_u por encima de un umbral alto u mediante una Distribución Pareto Generalizada (DPG); esto, a su vez, nos permite estimar la cola $\overline{F}(x)$ de la distribución, para valores grandes de x .

Hay que enfatizar que la elección de un modelo siempre conlleva el uso de suposiciones adicionales sin las cuales el modelo no sería válido. La motivación para el uso de la TVE es que la condición de que una función de distribución F esté en el Dominio de Atracción del Máximo de la distribución de Valores Extremos Generalizada H_ζ ($F \in DAM(H_\zeta)$) nos permite estimar cuantiles fuera del rango de los datos.

Estamos conscientes de que muchos de los temas que se tratarán a lo largo de este

trabajo —entre los que se mencionan, a manera de ejemplo, las distribuciones elípticas, la variación regular y algunos estimadores— podrían tratarse de manera mucho más extensa, pero no pretendemos que la naturaleza de este trabajo sea exhaustiva; de la misma manera, sabemos que otros métodos, otras medidas de riesgo y otros estimadores podrían también aplicarse para la obtención de estimados de medidas de riesgo extremo, pero nuestra intención aquí es la de dar un panorama general sobre la medición de riesgos extremos usando la Teoría de Valores Extremos (TVE) como base teórica y, como medidas de riesgo extremo, la Esperanza Condicional de la Cola (ES) y el Valor en Riesgo (VaR).

A continuación se da una breve descripción de la organización de este trabajo. En el Capítulo 1 se aborda la parte teórica, a través de los resultados más importantes de la TVE, entre los que destacan el Teorema de Fisher-Tippet y el Teorema de Gnedenko-Balkema-Pickands-de Haan; se dan, asimismo, las propiedades más importantes de la Distribución Pareto Generalizada (DPG) y las caracterizaciones de los Dominios de Atracción del Máximo de las tres distribuciones de Valores Extremos: la Fréchet (Φ_α), la Weibull (Ψ_α) y la Gumbel (Λ). En el Capítulo 2 se tratan las características deseables de una medida de riesgo y se analizan, a través de ejemplos, las ventajas y desventajas de tres conocidas medidas de riesgo: la varianza, el Valor en Riesgo (VaR) y la Esperanza Condicional de la Cola (ES). Se hace notar que la ES es una medida de riesgo coherente, en el sentido de Artzner *et al.* [4], es decir, satisface las características deseables de una medida de riesgo. El Capítulo 3 expone dos modelos que nos permiten obtener estimadores de dos medidas de riesgo extremo —el VaR y la ES. Estos modelos son el Modelo de Picos sobre el Umbral

(POT) y el Modelo de Máximos de Bloque, los cuales están fundamentados en la teoría presentada en el Capítulo 1. Finalmente, el Capítulo 4 comienza con una breve exposición de técnicas para el análisis gráfico de los datos; inmediatamente después se pone en práctica la teoría expuesta en los capítulos anteriores, mediante el análisis estadístico de tres conjuntos de datos, correspondientes, respectivamente, a pérdidas por incendios industriales en Dinamarca, a variaciones en el precio diario de la acción de BMW y a las pérdidas en el tipo de cambio FLX de pesos mexicanos a dólares americanos (MXN/USD). Cada uno de estos conjuntos de datos se analizan tanto con el Modelo de Máximos de Bloque como con el de Picos sobre el Umbral; el uso de ambos modelos está ampliamente ilustrado con gráficas y tablas. Para estos análisis se utiliza el *software* EVIS (*Extreme Values in S-Plus*), desarrollado en ETH-Zurich, disponible a través del *internet* en <http://www.math.ethz.ch/~mcneil>.

El lector puede elegir si desea comenzar por el Capítulo 1 y estudiar la teoría allí presentada para después pasar a los capítulos posteriores o, si lo prefiere, puede comenzar directamente por el Capítulo 2, continuar con el Capítulo 3 y sólo regresar al Capítulo 1 a medida que requiera —en el contexto de los Modelos expuestos en el Capítulo 3— los resultados teóricos pertinentes. Por su parte, los ejemplos del Capítulo 4 pueden ser muy útiles para aclarar el funcionamiento de los Modelos del Capítulo 3, ya que muestran posibles aplicaciones prácticas de estos.

Capítulo 1

Preliminares

En la teoría clásica de probabilidad, la mayoría de los resultados relevantes para las finanzas y los seguros se basan en sumas de la forma $S_n = \sum_{r=1}^n X_r$.

El resultado más importante en este sentido es el Teorema del Límite Central (ver Apéndice A), que nos dice que, para valores de n suficientemente grandes, tenemos que la suma S_n de observaciones mutuamente independientes X_1, X_2, \dots, X_n de una v.a. X (que a su vez pueden verse como n v.a.i.i.d.) tiene una distribución aproximadamente normal con media $n\mu_X$ y varianza $n\sigma_X^2$. Matemáticamente, si $Z_n = \frac{S_n - n\mu_X}{\sqrt{n}\sigma_X}$,

$$P \{S_n \leq s\} = P \left\{ Z_n \leq \frac{s - n\mu_X}{\sqrt{n}\sigma_X} \right\} \simeq F_Z \left(\frac{s - n\mu_X}{\sqrt{n}\sigma_X} \right).$$

Existen, además, generalizaciones del Teorema del Límite Central, diseñadas para incluir los casos de las distribuciones α -estables, con $0 < \alpha < 2$, los procesos α -estables, con $0 < \alpha < 2$, e incluso el movimiento Browniano.

Sin embargo, el utilizar modelos basados en sumas funciona de manera adecuada solamente cuando nos interesan los promedios de los fenómenos aleatorios que estu-

diamos. Pero no podemos pretender ampliar el campo de aplicación de las sumas si lo que nos interesa son los eventos extremos, como en nuestro caso. Para estudiar los eventos extremos existe una herramienta que, tanto desde el punto de vista intuitivo como desde el punto de vista teórico, es mucho más adecuada: esta herramienta se basa en el máximo M_n de n v.a.i.i.d. con función de distribución F :

$$M_n = \max \{X_1, X_2, \dots, X_n\}$$

Intuitivamente, el máximo M_n representa la observación más extrema dentro de un conjunto de n observaciones.

Puesto que la mayoría de las veces lo que nos interesa en finanzas es tratar de cuantificar las pérdidas, en este trabajo consideraremos a las pérdidas como positivas. Sin embargo, todos los resultados que veremos para máximos pueden formularse de manera equivalente para mínimos. Es más común encontrar los resultados de la Teoría de Valores Extremos para máximos puesto que ésta se desarrolla más fácilmente como una teoría de grandes pérdidas que como una teoría de pequeñas ganancias.

La organización de este Capítulo es la siguiente: en la Sección 1.1 se trata el tema de la convergencia de máximos M_n de v.a.i.i.d. Para este propósito, se ofrecen resultados de convergencia en probabilidad, convergencia casi segura y, más importantemente, convergencia en distribución. Se proporcionan, asimismo, algunos ejemplos en los que no existe un límite no-degenerado para M_n , lo que nos permite percibir la importancia que tiene el comportamiento de la cola de la distribución F en el caso de los máximos. En la Sección 1.2 se prueba un resultado fundamental, conocido como Teorema de Fisher-Tippett, que nos indica las tres posibles distribuciones límite para

los máximos M_n de v.a.i.i.d., bajo la normalización adecuada. Más específicamente, este resultado clásico nos dice que, si existen sucesiones de constantes normalizantes $c_n > 0$ y $d_n \in \mathbb{R}$ para las cuales $c_n^{-1}(M_n - d_n)$ converge a una función de distribución límite no-degenerada $H(x)$, entonces H debe ser del tipo de una de las tres distribuciones de valores extremos: la Fréchet (Φ_α), la Weibull (Ψ_α) y la Gumbel (Λ). Se verá, además, que estas tres distribuciones pueden identificarse con una clase que tiene cierta propiedad de estabilidad: las distribuciones max-estables. En la Sección 1.3 se introduce el concepto de Dominio de Atracción del Máximo (DAM) de una distribución y se analizan las condiciones que se requieren de la cola \bar{F} y del extremo derecho x_F de una f.d. F para poder pertenecer al DAM de cada una de las tres distribuciones de valores extremos. Se obtienen también fórmulas para las constantes normalizantes c_n y d_n en cada uno de los casos. Esta Sección se acompaña de numerosos ejemplos; además, al final se presenta un cuadro que resume los resultados más importantes acerca de los tres DAM, de manera que se facilite su consulta. Finalmente, la Sección 1.4 introduce el concepto de función de distribución de exceso F_u , así como otra importante distribución: la Distribución Pareto Generalizada (DPG). La relación entre ellas queda dada por el Teorema Gnedenko-Balkema-Pickands-de Haan (TGBPdH), que nos dice que, para una elección suficientemente alta del umbral u , la función de distribución de exceso F_u puede aproximarse por medio de una DPG. Asimismo, se exponen algunas propiedades de la DPG que resultarán de gran importancia para los Modelos que consideraremos en el Capítulo 3.

Este Capítulo está tomado principalmente de [6], [22] y [8]. Otras referencias de gran utilidad son [15] y [14].

1.1 Convergencia de Máximos.

Hemos dicho que la herramienta que utilizaremos para modelar los eventos extremos es el máximo muestral M_n para una sucesión de v.a.i.i.d. X, X_1, X_2, \dots con función de distribución F . A continuación, investigaremos las propiedades asintóticas de estos máximos. Nos interesa encontrar un resultado para máximos que nos proporcione información semejante a la que se obtiene para sumas con el Teorema del Límite Central. Como se mencionó anteriormente, la teoría se desarrollará en términos de máximos porque se concibe más fácilmente como una teoría de grandes pérdidas que como una teoría de ganancias pequeñas. Sin embargo, los resultados para mínimos pueden obtenerse fácilmente a partir de los correspondientes a máximos, usando la identidad

$$\min(X_1, X_2, \dots, X_n) = -\max(-X_1, -X_2, \dots, -X_n)$$

Recordemos que la distribución del máximo M_n de v.a.i.i.d. está dada por:

$$P(M_n \leq x) = P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) = F^n(x), \quad x \in \mathbb{R}, \quad n \in \mathbb{N}$$

Ahora bien, puesto que los extremos ocurren 'cerca' del extremo superior del soporte de la distribución, tendríamos que, intuitivamente, la conducta asintótica de M_n debe estar relacionada con la cola derecha de F cerca de su extremo derecho. Denotaremos por $\bar{F} = 1 - F$ a la cola derecha de la distribución F y por

$$x_F = \sup \{x \in \mathbb{R} : F(x) < 1\} \tag{1.1}$$

al extremo derecho de la distribución F .

De acuerdo a esta definición de x_F , es fácil notar que, para todo $x < x_F$, se tiene

$$P(M_n \leq x) = F^n(x) \rightarrow 0, \quad n \rightarrow \infty \quad (1.2)$$

puesto que para $x < x_F$ se da $F(x) < 1$.

Si además $x_F < \infty$, para toda $x \geq x_F$ se tendrá que $F(x) = 1$ y, por tanto,

$$P(M_n \leq x) = F^n(x) = 1 \quad (1.3)$$

Proposición 1.1.1 $M_n \xrightarrow{P} x_F$, cuando $n \rightarrow \infty$, para $x_F \leq \infty$.

Demostración:

$$P.D.: \forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} P(|M_n - x_F| < \epsilon) = 1$$

Sea $\epsilon > 0$.

$$P(|M_n - x_F| < \epsilon) = F^n[(x_F + \epsilon)_-] - F^n(x_F - \epsilon)$$

donde $F^n[(x_F + \epsilon)_-]$ es el límite por la izquierda de $F^n(m)$ en $m = (x_F + \epsilon)$.

Ahora bien, por (1.3), $\lim_{n \rightarrow \infty} F^n[(x_F + \epsilon)_-] = 1$ si $x_F < \infty$. Esto sucede aún si $x_F = \infty$, pues una propiedad de las funciones de distribución es $\lim_{x \rightarrow \infty} F(x) = 1$.

Por otra parte, (1.2) nos asegura que $\lim_{n \rightarrow \infty} F^n(x_F - \epsilon) = 0$.

De aquí que $\lim_{n \rightarrow \infty} P(|M_n - x_F| < \epsilon) = 1 - 0 = 1$, es decir, $M_n \xrightarrow{P} x_F$ cuando $n \rightarrow \infty$, para $x_F \leq \infty$. ■

De hecho, gracias a que la sucesión (M_n) es no-decreciente, se puede asegurar algo más fuerte en términos de convergencia:

Proposición 1.1.2 $M_n \xrightarrow{c.s.} x_F$, cuando $n \rightarrow \infty$.

Demostración:

P.D.: $P(\lim_{n \rightarrow \infty} M_n = x_F) = 1$

Supongamos que $P(\lim_{n \rightarrow \infty} M_n = x_F) < 1$.

Entonces, $\exists (M_{n_k})_{n_k \geq 1}$ subsucesión de (M_n) tal que para $x_G \neq x_F$,

$$P(\lim_{n \rightarrow \infty} M_n = x_G) > 0.$$

Puesto que (M_n) es no-decreciente, tenemos que $M_{n+1} \geq M_n \quad \forall n \in \mathbb{N}$.

Supongamos que $x_G > x_F$.

Entonces $\exists A \in (M_{n_k})$ tal que $A > M_n \quad \forall n \in \mathbb{N}$; esto es, p.a. $k \in \mathbb{N}, M_{n_k} > M_n \quad \forall n \in \mathbb{N}$.

Pero (M_{n_k}) es subsucesión de (M_n) , así que $\forall n_k \quad \exists s \in \mathbb{N}$ tal que $M_{n_k} = M_s$. Por tanto, $x_G \not> x_F$.

Supongamos ahora que $x_F > x_G$.

Entonces $\exists M_r \in (M_n)_{n \geq 1}$ tal que $M_r > M_{n_k} \quad \forall k \in \mathbb{N}$. Esto significa que, como (M_n) es no-decreciente, $\forall s > r, \quad \forall A \in (M_{s+n})_{n \geq 1}, \quad A \notin (M_{n_k})_{n_k \geq 1}$. Pero entonces, tendríamos que (M_{n_k}) es finita, lo cual es una contradicción. Por tanto, $x_F \not> x_G$.

Entonces se tiene $x_F = x_G$, de donde $P(\lim_{n \rightarrow \infty} M_n = x_F) = 1$, es decir, $M_n \xrightarrow{c.s.} x_F$,

cuando $n \rightarrow \infty$. ■

Desafortunadamente, este hecho tampoco nos proporciona suficiente información acerca de la magnitud de los máximos, pues para cualesquiera dos variables aleatorias con soporte infinito -por ejemplo, la Normal y la Poisson- obtendríamos exactamente el mismo resultado: que sus máximos convergen casi seguramente a infinito; así, para fines prácticos, no habría diferencia entre el comportamiento del máximo de una variable aleatoria Normal y el de una variable aleatoria Poisson. Para saber

un poco más al respecto, necesitamos considerar resultados de convergencia débil (convergencia en distribución) para máximos.

Antes de dar los resultados de manera general, consideraremos algunos ejemplos que pueden proporcionarnos una idea intuitiva sobre el comportamiento de los máximos de v.a.i.i.d. y su relación con la cola de la distribución.

Ejemplo 1.1.3 (Distribución Exponencial) Sea $F(x) = 1 - e^{-\lambda x}$, $\lambda > 0$. Tomando $u_n = \frac{1}{\lambda}(\log n - \log \tau)$, $\tau > 0$, se obtiene

$$1 - F(u_n) = 1 - \left(1 - e^{-\lambda \left[\frac{1}{\lambda}(\log n - \log \tau)\right]}\right) = e^{-(\log n - \log \tau)} = e^{\log(\tau/n)} = \tau/n.$$

Por otra parte,

$$P(M_n \leq u_n) = F^n(u_n) = \left(1 - e^{-\lambda \left[\frac{1}{\lambda}(\log n - \log \tau)\right]}\right)^n = \left(1 - \frac{\tau}{n}\right)^n \rightarrow e^{-\tau}.$$

Así, vemos que para v.a.i. con distribución exponencial, tomando la sucesión (u_n) antes mencionada se da

$$n\bar{F}(u_n) = \tau \quad \text{y} \quad P(M_n \leq u_n) \rightarrow e^{-\tau}.$$

Ejemplo 1.1.4 (Distribución Uniforme (0,1)) Sea $F(x) = x$, $0 \leq x \leq 1$. Sea $\tau > 0$ y $u_n = 1 - \tau/n$. Entonces

$$1 - F(u_n) = 1 - (1 - \tau/n) = \tau/n$$

para $n \geq \tau$. También se cumple

$$P(M_n \leq u_n) = F^n(u_n) = (1 - \tau/n)^n \rightarrow e^{-\tau},$$

por lo que para la distribución uniforme, tomando $u_n = 1 - \tau/n$, se satisface

$$n\bar{F}(u_n) = \tau \quad \text{y} \quad P(M_n \leq u_n) \rightarrow e^{-\tau}.$$

Ejemplo 1.1.5 (Distribución Pareto) Sea $F(x) = 1 - Kx^{-\alpha}$, $\alpha > 0$, $K > 0$, $x \geq K^{1/\alpha}$. Sea $u_n = (Kn/\tau)^{1/\alpha}$. Entonces, tenemos que

$$1 - F(u_n) = 1 - \left(1 - K \left((Kn/\tau)^{1/\alpha}\right)^{-\alpha}\right) = K \left((Kn/\tau)^{-1}\right) = \tau/n.$$

De la misma manera,

$$P(M_n \leq u_n) = F^n(u_n) = \left(1 - K \left((Kn/\tau)^{1/\alpha}\right)^{-\alpha}\right)^n = \left(1 - \frac{\tau}{n}\right)^n \rightarrow e^{-\tau}.$$

Vemos entonces que también la distribución Pareto satisface

$$n\bar{F}(u_n) = \tau \quad \text{y} \quad P(M_n \leq u_n) \rightarrow e^{-\tau},$$

para $u_n = (Kn/\tau)^{1/\alpha}$.

En los ejemplos anteriores hemos visto que se requiere que la función de distribución F satisfaga ciertas condiciones para que se tenga que el $\lim_{n \rightarrow \infty} P(M_n \leq u_n)$ existe para constantes u_n adecuadas. El siguiente resultado nos confirma esta suposición.

Proposición 1.1.6 (Aproximación de Poisson) Para $\tau \in [0, \infty]$ y una sucesión

(u_n) de números reales, son equivalentes:

$$n\bar{F}(u_n) \rightarrow \tau \tag{1.4}$$

$$P(M_n \leq u_n) \rightarrow e^{-\tau} \tag{1.5}$$

Demostración:

“(1.4) \implies (1.5)”

Sea $0 \leq \tau < \infty$. Supongamos que se satisface (1.4). Entonces

$$P(M_n \leq u_n) = F^n(u_n) = 1 - (1 - F^n(u_n)) = (1 - \bar{F}(u_n))^n \stackrel{(1.4)}{=} \left(1 - \frac{\tau}{n} + o\left(\frac{1}{n}\right)\right)^n$$

Y puesto que sabemos que $\lim_{n \rightarrow \infty} \left(1 + \frac{b}{n} + \frac{\vartheta(n)}{n}\right)^{cn} = e^{bc}$, donde b, c no dependen de n y $\lim_{n \rightarrow \infty} \vartheta(n) = 0$, tenemos que $\left(1 - \frac{\tau}{n} + o\left(\frac{1}{n}\right)\right)^n$ es equivalente a (1.5), de donde (1.4) \implies (1.5).

“(1.5) \implies (1.4)”

Supongamos que se da (1.5). Entonces,

$$P(M_n \leq u_n) = F^n(u_n) = 1 - (1 - F^n(u_n)) = (1 - \bar{F}(u_n))^n \rightarrow e^{-\tau}$$

Sacando logaritmos, se obtiene

$$-n \ln(1 - \bar{F}(u_n)) \rightarrow -\tau \tag{1.6}$$

Ahora bien, sabemos que $-\ln(1 - x) \sim x$ si $x \rightarrow 0$. En este caso, si tuviéramos $\bar{F}(u_n) \rightarrow 0$, obtendríamos de (1.6) que $n\bar{F}(u_n) = \tau + o(1)$, que es equivalente a (1.4). Por tanto, basta probar que se da $\bar{F}(u_n) \rightarrow 0$.

Supongamos que $\bar{F}(u_n) \not\rightarrow 0$. Entonces $\exists(u_{n_k})$ tal que para algún $\epsilon > 0$ y $\forall N \in \mathbb{N}$, $\exists n_k > N$ para el cual $\bar{F}(u_{n_k}) > \epsilon$. Pero entonces $(1 - \bar{F}(u_{n_k}))^{n_k} < (1)^{n_k}$, por lo que $P(M_{n_k} \leq u_{n_k}) = (1 - \bar{F}(u_{n_k}))^{n_k} \rightarrow 0$. Pero esto contradice a (1.5). De aquí que debe darse $\bar{F}(u_n) \rightarrow 0$.

Ahora, si $\tau = \infty$ y se da (1.4) pero no se da (1.5), entonces $\exists(u_{n_k})$ tal que para algún $\tau' < \infty$, se tiene $P(M_{n_k} \leq u_{n_k}) \rightarrow e^{-\tau'}$ cuando $n_k \rightarrow \infty$. Pero entonces, puesto que para $\tau \in (0, \infty)$ (1.5) implica (1.4), tendríamos $n_k \bar{F}(u_{n_k}) \rightarrow \tau' < \infty$, lo cual contradice que se dé (1.4) con $\tau = \infty$ como se había supuesto. De la misma manera,

si $\tau = \infty$ y se da (1.5) pero no se da (1.4), entonces $\exists(u_{n_k})$ tal que para algún $\tau' < \infty$, se tiene $n_k \bar{F}(u_{n_k}) \rightarrow \tau' < \infty$. Pero entonces, puesto que en este caso ya se probó que (1.4) implica (1.5), se tendría $P(M_{n_k} \leq u_{n_k}) \rightarrow e^{-\tau'}$ cuando $n_k \rightarrow \infty$, lo cual contradice el que se dé (1.5) con $\tau = \infty$ como se había supuesto. ■

Es importante notar que si existe una sucesión $(u_n^{(\tau)})$ que satisfaga (1.4) para algún $\tau > 0$ fijo, entonces podemos encontrar este tipo de sucesiones para cualquier $\tau > 0$. Por ejemplo, si $(u_n^{(1)})$ satisface (1.4) con $\tau = 1$, entonces $u_n^{(\tau)} = u_{\lfloor \frac{n}{\tau} \rfloor}^{(1)}$ satisface (1.4) para $\tau > 0$ arbitrario, donde $\lfloor x \rfloor$ representa al mayor entero menor o igual que x .

Antes de enunciar la Aproximación de Poisson se expusieron tres ejemplos de distribuciones (la Exponencial, la Uniforme (0,1) y la Pareto) que satisfacen las relaciones (1.4) y (1.5). El siguiente Corolario nos dice que no siempre podemos encontrar límites no-degenerados para máximos de v.a.i.i.d.

Corolario 1.1.7 *Supóngase que $x_F < \infty$ y que*

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

Entonces, para toda sucesión (u_n) tal que

$$P(M_n \leq u_n) \rightarrow \rho,$$

se tiene que $\rho = 0$ o $\rho = 1$.

Demostración:

Puesto ρ es el límite de una probabilidad, $0 \leq \rho \leq 1$. Entonces podemos expresar $\rho = e^{-\tau}$, con $0 \leq \tau \leq \infty$. Pero esto, por la Proposición anterior, significa que tenemos

$n\bar{F}(u_n) \rightarrow \tau$ conforme $n \rightarrow \infty$.

CASO I. Si $u_n < x_F$ para una infinidad de n 's, tenemos que, para esas n ,

$$F(u_n) \leq F(x_{F-}) \Leftrightarrow 1 - F(u_n) \geq 1 - F(x_{F-}) \geq F(x_F) - F(x_{F-}) \Leftrightarrow \bar{F}(u_n) \geq \bar{F}(x_{F-}) > 0$$

Y puesto que $n\bar{F}(u_n) \rightarrow \infty$, esto significa que $\tau = \infty$.

CASO II. Si $\exists K \in \mathbb{N}$ tal que $\forall n > K$ se tenga $u_n \geq x_F$, entonces $F(u_n) = 1$ $\forall u_n = u_{K+s}, s \in \mathbb{N}$. En este caso, se tiene $\bar{F}(u_n) = 1 - F(u_n) = 1 - 1 = 0$, lo cual hace $n\bar{F}(u_n) \rightarrow 0$ conforme $n \rightarrow \infty$. Así, tenemos $\tau = 0$.

Hemos visto entonces que las únicas dos opciones son tener $\tau = 0$ o $\tau = \infty$, que nos dan, respectivamente, $\rho = 1$ o $\rho = 0$. ■

Este resultado nos muestra que no existe una distribución límite no-degenerada (es decir, distinta de 0 y de 1) para el máximo M_n de una función de distribución con un salto en su extremo derecho finito, sin importar cuál sea la normalización de M_n que se utilice.

A continuación se enuncia un resultado similar que cubre también a ciertas distribuciones con extremo derecho infinito.

Teorema 1.1.8 *Sea F un f.d. con extremo derecho $x_F \leq \infty$ y sea $\tau \in (0, \infty)$. Si*

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x)}{\bar{F}(x-)} = 1, \tag{1.7}$$

entonces existe una sucesión (u_n) para la cual se satisface $n\bar{F}(u_n) \rightarrow \tau$ cuando $n \rightarrow \infty$.

Demostración:

Supongamos que se da (1.7) y sea $\{u_n\}$ cualquier sucesión para la cual $F(u_n-) \leq 1 - \tau/n \leq F(u_n)$ (por ejemplo, puede ser $u_n = F^{-1}(1 - \tau/n)$). De aquí se sigue que:

$$\begin{aligned} \frac{\tau}{\overline{F}(u_n-)} &\leq n, \\ \tau &\geq n\overline{F}(u_n), \end{aligned}$$

de donde se obtiene

$$\frac{\overline{F}(u_n)}{\overline{F}(u_n-)} \tau \leq n\overline{F}(u_n) \leq \tau.$$

El resultado se sigue de aquí puesto que $u_n \rightarrow x_F$ cuando $n \rightarrow \infty$. ■

Este resultado se aplica en particular a las distribuciones discretas con extremo derecho infinito. Si las alturas de los saltos de la f.d. no decaen suficientemente rápido, entonces no existe una distribución límite no-degenerada para el máximo de la distribución. En particular si X toma valores en los enteros y $x_F = \infty$, entonces (1.7) se traduce a $\overline{F}(n)/\overline{F}(n-1) \rightarrow 1$ cuando $n \rightarrow \infty$.

Vemos entonces que, en contraste con el caso de las sumas, para máximos siempre necesitamos condiciones de continuidad relativamente delicadas en la cola \overline{F} para asegurar la convergencia de $P(M_n \leq u_n)$ a un límite no trivial. Esto deja fuera a algunas distribuciones importantes, como la Poisson y la Geométrica.

Ejemplo 1.1.9 (Distribución Poisson) Tenemos que $P[X = k] = e^{-\lambda} \lambda^k / k!$, $k \in \mathbb{N} \cup \{0\}$, $\lambda > 0$.

Entonces

$$\frac{\overline{F}(k)}{\overline{F}(k-1)} = \frac{1 - F(k) + F(k-1) - F(k-1)}{\overline{F}(k-1)} = 1 - \frac{F(k) - F(k-1)}{\overline{F}(k-1)}$$

$$= 1 - \frac{e^{-\lambda} \lambda^k}{k!} \left(e^{-\lambda} \sum_{r=k}^{\infty} \frac{\lambda^r}{r!} \right)^{-1} = 1 - \frac{\lambda^k}{k!} \left(\sum_{r=k}^{\infty} \frac{\lambda^r}{r!} \right)^{-1} \quad (1.8)$$

$$= 1 - \left(1 + \sum_{r=k+1}^{\infty} \frac{k!}{r!} \lambda^{r-k} \right)^{-1}. \quad (1.9)$$

Esta última suma puede reescribirse como

$$\sum_{s=1}^{\infty} \frac{\lambda^s}{(k+1)(k+2)\dots(k+s)} \leq \sum_{s=1}^{\infty} \left(\frac{\lambda}{k} \right)^s = \frac{\lambda/k}{1 - \lambda/k}, \quad k > \lambda,$$

lo cual tiende a cero cuando $k \rightarrow \infty$, de donde $\bar{F}(k)/\bar{F}(k-1) \rightarrow 0$. Esto significa, por el Teorema 1.1.8 que no existe una distribución límite no-degenerada para el máximo de v.a. Poisson y, más aún, no existe ningún límite de la forma $P[M_n \leq u_n] \rightarrow \rho \in (0, 1)$, sin importar cuál sea la sucesión de constantes (u_n) que tomemos.

Ejemplo 1.1.10 (Distribución Geométrica) Se tiene $P[X = k] = p(1-p)^{k-1}$, $k \in \mathbb{N}$, $0 < p < 1$.

En este caso

$$\frac{\bar{F}(k)}{\bar{F}(k-1)} = 1 - (1-p)^{k-1} \left(\sum_{r=k}^{\infty} (1-p)^{r-1} \right)^{-1} = 1 - p \in (0, 1).$$

Nuevamente, esto muestra que no existe el límite $P[M_n \leq u_n] \rightarrow \rho$, salvo para $\rho = 0$ ó 1 ; también nos dice que no existe una distribución límite no-degenerada para el máximo de v.a.s Geométricas.

1.2 Las Distribuciones de Valores Extremos y el Teorema de Fisher-Tippet.

Esta sección se ocupa principalmente de uno de los resultados más importantes en la Teoría de Valores Extremos: el Teorema de Fisher-Tippet. Este teorema especifica las tres posibles formas de las distribuciones límite para máximos apropiadamente normalizados de sucesiones de v.a.i.i.d. Estos tres tipos de distribuciones se conocen, por este motivo, como distribuciones de valores extremos. Veremos también que podemos identificarlas con una clase de distribuciones que poseen una cierta condición de estabilidad: las llamadas distribuciones max-estables.

Comenzaremos pues dando las definiciones de tres distribuciones que resultarán muy importantes para el estudio de la convergencia de máximos apropiadamente normalizados.

Definición 1.2.1 *La función de distribución de una variable aleatoria Fréchet, para $\alpha > 0$, está dada por*

$$\Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{si } x > 0 \end{cases} \quad (1.10)$$

Definición 1.2.2 *La función de distribución de una variable aleatoria Gumbel está dada por*

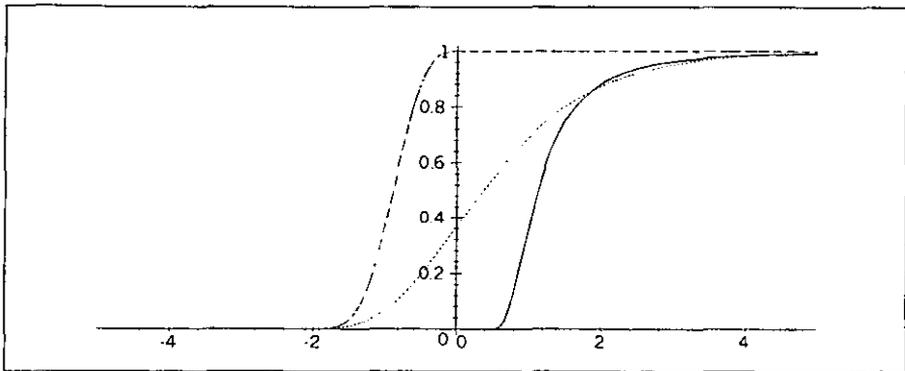
$$\Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R} \quad (1.11)$$

Definición 1.2.3 *La función de distribución de una variable aleatoria Weibull, para*

$\alpha > 0$, está dada por

$$\Psi_{\alpha}(x) = \begin{cases} \exp(-(-x)^{\alpha}) & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases} \quad (1.12)$$

De la definición de estas tres distribuciones de valores extremos, es fácil notar que la Fréchet corresponde a un modelo con cota inferior finita y cota superior infinita; la Weibull corresponde a un modelo con cota superior finita y cota inferior infinita; mientras que la Gumbel es no-acotada por ambos lados.



Distribuciones Fréchet (sólido) y Weibull (guiones) con $\alpha = 3$; distribución Gumbel (punteada).

El siguiente resultado nos muestra que existe una cercana relación entre las distribuciones Fréchet, Weibull y Gumbel.

Proposición 1.2.4 $X \sim \Phi_{\alpha}(x) \Leftrightarrow \log X^{\alpha} \sim \Lambda(x) \Leftrightarrow -\frac{1}{X} \sim \Psi_{\alpha}(x)$.

Demostración:

$$X \sim \Phi_{\alpha}(x) \Leftrightarrow P[X \leq x] = \exp(-x^{-\alpha}) \text{ para } x > 0 \text{ y } P[X \leq x] = 0 \text{ para } x \leq 0$$

$$\begin{aligned}
&\Leftrightarrow P\left[\frac{-1}{X} \leq x\right] = P\left[X \geq \frac{-1}{x}\right] = 1 - P\left[X \leq \frac{-1}{x}\right] = 1 - \exp\left(-\left(-\frac{1}{x}\right)^{-\alpha}\right) = 1 - \\
&\exp\{-(-x)^\alpha\} \text{ para } x < 0 \text{ y } P\left[\frac{-1}{X} \leq x\right] = 1 - P\left[X \leq \frac{-1}{x}\right] = 1 - 0 = 1 \text{ para } x \geq 0 \\
&\Leftrightarrow P[\log X^\alpha \leq x] = P[X^\alpha \leq \exp x] = P\left[X \leq (\exp x)^{1/\alpha}\right] = (\exp\{-\exp(x)\})^{-\alpha})^{1/\alpha} \\
&= \exp\left\{-\exp\left[x(-\alpha)\left(\frac{1}{\alpha}\right)\right]\right\} = \exp\{-\exp(-x)\} = \Lambda(x). \quad \blacksquare
\end{aligned}$$

Definición 1.2.5 *Dos variables aleatorias Z y Y son del mismo tipo si y sólo si $Z \stackrel{d}{=} aY + b$, para $a > 0$ y $b \in \mathbb{R}$. En términos de sus funciones de distribución F_Z y F_Y , esto significa que $F_Z(x) = F_Y\left(\frac{x-b}{a}\right)$.*

Esto es, dos variables del mismo tipo tienen la misma función de distribución, excepto por posibles cambios en los parámetros de escala y localización.

Una propiedad muy relevante de las distribuciones Fréchet, Weibull y Gumbel es que, cuando consideramos sus máximos, obtenemos que estos son del mismo tipo que la distribución original. Esta propiedad se precisa a continuación.

Definición 1.2.6 (Distribución max-estable) *Una v.a. X no-degenerada (la distribución o f.d. correspondiente) es max-estable si, para X, X_1, \dots, X_n i.i.d. y para constantes adecuadas $c_n > 0$ y $d_n \in \mathbb{R}$ satisface*

$$\max(X_1, \dots, X_n) \stackrel{d}{=} c_n X + d_n, \quad n \geq 2. \quad (1.13)$$

Ejemplo 1.2.7 (Propiedad max-estable de la f.d. Fréchet Φ_α) Sean X_1, \dots, X_n v.a.i. con distribución Fréchet Φ_α . Entonces

$$\begin{aligned}
P[M_n \leq x] &= \Phi_\alpha^n(x) = (\exp(-x^{-\alpha}))^n = (\exp(-x^{-\alpha})n) \\
&= (\exp(-xn^{-1/\alpha})^{-\alpha}) = \Phi_\alpha(n^{-1/\alpha}x).
\end{aligned}$$

De donde, la f.d. Fréchet es max-estable.

Ejemplo 1.2.8 (Propiedad max-estable de la f.d. Weibull Ψ_α) Consideremos X_1, \dots, X_n v.a.i. con distribución Weibull Ψ_α . El máximo de estas v.a.i.i.d. satisface

$$\begin{aligned} P[M_n \leq x] &= \Psi_\alpha^n(x) = (\exp(-(-x)^\alpha))^n = \exp(-(-x)^\alpha n) \\ &= \exp(-(-xn^{1/\alpha})^\alpha) = \Psi_\alpha(n^{1/\alpha}x), \end{aligned}$$

de donde es claro que la distribución Weibull es max-estable.

Ejemplo 1.2.9 (Propiedad max-estable de la f.d. Gumbel Λ) Supongamos que tenemos X_1, \dots, X_n v.a.i. con distribución Gumbel Λ . Entonces

$$\begin{aligned} P[M_n \leq x] &= \Lambda^n(x) = (\exp(-e^{-x}))^n = (\exp(-e^{-x}n)) \\ &= \exp(-e^{-(x+\ln n)}) = \Lambda(x - \ln n). \end{aligned}$$

Así, también la distribución Gumbel es max-estable.

Cuando introducimos las distribuciones Fréchet, Weibull y Gumbel mencionamos que éstas resultarían de gran importancia para el estudio de la convergencia de máximos de v.a.i.i.d. apropiadamente normalizados. Esta afirmación está parcialmente justificada por el hecho de que las tres distribuciones son max-estables. Pero, más allá de esto, las distribuciones Fréchet, Weibull y Gumbel aparecen como límites de máximos de v.a.i.i.d. apropiadamente normalizados. Los siguientes ejemplos ilustran este hecho:

Ejemplo 1.2.10 (Distribución Exponencial) Sean X_1, \dots, X_n v.a.i. con distribución exponencial con parámetro λ . En el Ejemplo 1.1.3 obtuvimos que

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

para $u_n = \frac{1}{\lambda}(\log n - \log \tau)$. Si tomamos ahora $\tau = e^{-x}$, se obtiene

$$\begin{aligned} P\left[M_n \leq \frac{1}{\lambda}(\log n - \log \tau)\right] &= P\left[M_n \leq \frac{1}{\lambda}(\log n - \log e^{-x})\right] \\ &= P\left[M_n \leq \frac{1}{\lambda}(x + \log n)\right] \rightarrow \exp(-e^{-x}). \end{aligned}$$

Otra forma de ver esto, sin necesidad de utilizar la Aproximación de Poisson, es tomando en cuenta que $P[M_n \leq x] = F^n(x) = (1 - e^{-\lambda x})^n$. De esta manera,

$$P\left[M_n \leq \frac{1}{\lambda}(x + \log n)\right] = (1 - e^{-\lambda(1/\lambda)(x + \log n)})^n = \left(1 - \frac{e^{-x}}{n}\right)^n \rightarrow \exp(-e^{-x}).$$

Obsérvese que la distribución límite es precisamente la distribución Gumbel Λ arriba definida.

Ejemplo 1.2.11 (Distribución Uniforme (0,1)) En el Ejemplo 1.1.4 vimos que, para $u_n = 1 - \tau/n$, se tiene

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

cuando X_1, \dots, X_n son v.a.i. con distribución uniforme en $(0,1)$. Si hacemos $\tau = -x$ para $x < 0$, obtenemos

$$P\left[M_n \leq 1 - \tau/n\right] = P\left[M_n \leq 1 - (-x)/n\right] = P\left[M_n \leq \frac{x}{n} + 1\right] \rightarrow \exp(-(-x)).$$

También en este caso se puede hacer el cálculo de manera directa: puesto que $P\{M_n \leq x\} = F^n(x) = x^n$, tenemos que

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

Nótese que aquí la distribución límite corresponde a una distribución Weibull con parámetro $\alpha = 1$, es decir, Ψ_1 .

Ejemplo 1.2.12 (Distribución Pareto) Para X_1, \dots, X_n v.a.i. con distribución Pareto, en el Ejemplo 1.1.5 se calculó que

$$P(M_n \leq u_n) \rightarrow e^{-\tau},$$

con $u_n = (Kn/\tau)^{1/\alpha}$. Tomando ahora $\tau = x^{-\alpha}$ para $x > 0$, esta expresión se convierte en

$$\begin{aligned} P\left[M_n \leq (Kn/\tau)^{1/\alpha}\right] &= P\left[M_n \leq \left(\frac{Kn}{x^{-\alpha}}\right)^{1/\alpha}\right] \\ &= P\left[M_n \leq (Kn)^{1/\alpha} x\right] \rightarrow \exp(-x^{-\alpha}). \end{aligned}$$

O bien, utilizando el hecho de que conocemos la f.d., podemos llegar al mismo resultado haciendo el cálculo directamente:

$$\begin{aligned} P\left[M_n \leq (Kn)^{1/\alpha} x\right] &= F^n\left((Kn)^{1/\alpha} x\right) = \left[1 - K\left((Kn)^{1/\alpha} x\right)^{-\alpha}\right]^n \\ &= \left(1 - \frac{1}{x^\alpha n}\right)^n \rightarrow \exp(-x^{-\alpha}). \end{aligned}$$

Es importante señalar que la distribución límite obtenida en este ejemplo es precisamente una distribución Fréchet con parámetro α la cual, como hemos visto, se denota Φ_α .

Estos tres ejemplos nos han mostrado que podemos obtener a las distribuciones Fréchet, Weibull y Gumbel como límites de máximos apropiadamente normalizados. Surge entonces una pregunta: ¿es posible obtener alguna otra distribución límite no-degenerada para máximos apropiadamente normalizados? El siguiente Teorema, que data de 1928, nos permite responder esta pregunta. Por esta razón es uno de los resultados más importantes de la Teoría de Valores Extremos.

Teorema 1.2.13 (FISHER-TIPPET) Sean X_1, X_2, \dots, X_n v.a.i.i.d. con función de distribución F , y sean $(a_n), (b_n)$ constantes tales que para alguna distribución límite H no degenerada se tiene

$$F^n(a_n x + b_n) = P\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{d} H(x), \quad x \in \mathbb{R} \quad (1.14)$$

Entonces H es del tipo de alguna de las tres distribuciones siguientes: la Fréchet Φ_α , la Weibull Ψ_α o la Gumbel Λ .

Demostración:

Se procede por pasos:

PASO I) De acuerdo con (1.14), $\forall t > 0$ se tiene

$$F^{\lfloor nt \rfloor}(a_{\lfloor nt \rfloor} x + b_{\lfloor nt \rfloor}) \xrightarrow{d} H(x)$$

donde $\lfloor y \rfloor$ representa el mayor entero menor o igual que y . Por otra parte

$$F^{\lfloor nt \rfloor}(a_n x + b_n) = (F^n(a_n x + b_n))^{\frac{\lfloor nt \rfloor}{n}} \xrightarrow{d} H^t(x)$$

Entonces, por el Teorema de Convergencia a Tipos (ver Apéndice A) H y H^t son

del mismo tipo, es decir, existen $\alpha(t) > 0$ y $\beta(t)$ tales que para toda $t > 0$, $\frac{a_n}{a_{\lfloor nt \rfloor}} \rightarrow$

$\alpha(t)$, $\frac{b_n - b_{\lfloor nt \rfloor}}{a_{\lfloor nt \rfloor}} \rightarrow \beta(t)$ con las cuales

$$H^t(x) = H(\alpha(t)x + \beta(t)). \quad (1.15)$$

PASO II) Observamos que las funciones $\alpha(t)$ y $\beta(t)$ son Lebesgue-medibles.

P.D.1: $\alpha(t)$ es Lebesgue-medible.

Puesto que los límites de funciones medibles son medibles, basta mostrar que, para

cada $n \in \mathbb{N}$, la función $t \mapsto \frac{a_n}{a_{\lfloor nt \rfloor}}$ es medible. Como a_n no depende de t , lo anterior

es cierto si $t \mapsto a_{[nt]}$ es medible; pero el rango de esta función es $\{a_j : j \geq 1\}$, que es numerable, por lo que basta probar que $\{t > 0 : a_{[nt]} = a_j\}$ es medible. Pero este conjunto es igual a $\cup_{k:a_k=a_j} [\frac{k}{n}, \frac{k+1}{n}]$ el cual, por ser unión de intervalos, es un conjunto medible.

P.D.₂: $\beta(t)$ es Lebesgue-medible.

Para ello basta mostrar que, para cada $n > 0$, la función $t \mapsto \frac{b_n - b_{[nt]}}{a_{[nt]}}$ es medible. Como b_n no depende de t , esto equivale a mostrar que $t \mapsto \frac{-b_{[nt]}}{a_{[nt]}}$ es medible. Acabamos de mostrar que $t \mapsto a_{[nt]}$ es medible; procediendo de la misma manera, puede verse que $t \mapsto -b_{[nt]}$ es medible. Y puesto que el cociente de funciones medibles es medible, $t \mapsto \frac{-b_{[nt]}}{a_{[nt]}}$ es medible.

PASO III) Utilizando (1.15) se tiene, para $t, s > 0$, por un lado

$$H^{ts}(x) = H(\alpha(ts)x + \beta(ts));$$

y por otra parte,

$$\begin{aligned} H^{ts}(x) &= (H^s(x))^t = H(\alpha(s)x + \beta(s))^t = H(\alpha(t)\{\alpha(s)x + \beta(s)\} + \beta(t)) \\ &= H(\alpha(t)\alpha(s)x + \alpha(t)\beta(s) + \beta(t)). \end{aligned}$$

Sabemos que si una distribución F no-degenerada satisface que $F(ax + b) = F(cx + d)$ para todo $x \in \mathbb{R}$ y para algunos $a > 0$, $c > 0$, entonces debe darse que $a = c$ y $b = d$. Puesto que suponemos que H es no degenerada, concluimos entonces que

$$\alpha(ts) = \alpha(t)\alpha(s) \tag{1.16}$$

$$\beta(ts) = \alpha(t)\beta(s) + \beta(t) = \alpha(s)\beta(t) + \beta(s) = \beta(st). \tag{1.17}$$

Nótese que (1.16) es precisamente la ecuación de Hamel (ver Apéndice A), cuya única solución finita, medible y no-negativa es de la forma

$$\alpha(t) = t^{-\theta}, \quad \theta \in \mathbb{R}.$$

PASO IV) Tenemos entonces tres casos: (a) $\theta = 0$; (b) $\theta > 0$; (c) $\theta < 0$.

Caso (a), $\theta = 0$.

En este caso $\alpha(t) \equiv 1$, así que (1.17) se convierte en

$$\beta(ts) = \beta(t) + \beta(s).$$

De manera que $\exp\{\beta(\cdot)\}$ satisface la ecuación de Hamel, lo que implica que $\exp\{\beta(t)\} = t^{-c}$ para alguna $c \in \mathbb{R}$; esto quiere decir que

$$\beta(t) = -c \log t, \quad t > 0, \quad c \in \mathbb{R},$$

y (1.15) se convierte en

$$H^t(x) = H(x - c \log t) \tag{1.18}$$

Ahora bien, si c fuera 0, no podría darse que H sea no-degenerada. De modo que $c \neq 0$. Además, para x fija, $H^t(x)$ es no-creciente en t , lo que nos indica que $c > 0$, pues de otra manera el lado derecho de la expresión (1.18) no sería decreciente. Si para algún $x_0 \in \mathbb{R}$ se tiene que $H(x_0) = 1$, entonces, de (1.18) se obtiene

$$1 = H(x_0 - c \log t), \quad \forall t > 0,$$

lo cual, mediante un cambio de variable, implica que $H(u) = 1$ para toda $u \in \mathbb{R}$, lo que contradice que H sea no-degenerada. De la misma manera, si para algún $x_0 \in \mathbb{R}$

se tiene que $H(x_0) = 0$, entonces, de (1.18) se obtiene

$$0 = H^t(x_0) = H(x_0 - c \log t), \quad \forall t > 0,$$

lo cual implica que $H(x) = 0$ para todo $x \in \mathbb{R}$, lo que contradice de nuevo el que H sea no-degenerada. Concluimos entonces que $0 < H(y) < 1$ para todo $y \in \mathbb{R}$. Basta ahora sustituir $x = 0$ en la expresión (1.18), para obtener, para $t > 0$,

$$H^t(0) = H(-c \log t). \quad (1.19)$$

Sean $\exp\{-e^{-p}\} = H(0) \in (0, 1)$ y $u = -c \log t$. Como el rango de t es $(0, \infty)$, el rango de u queda dado por $(-\infty, \infty)$, y aplicando el cambio de variable en (1.19) se obtiene

$$\begin{aligned} H(u) &= (\exp\{-e^{-p}\})^t = \exp\{-e^{-p}t\} = \exp\{-e^{-(u/c+p)}\} \\ &= \Lambda\left(\frac{u}{c} + p\right). \end{aligned}$$

Caso (b), $\theta > 0$.

Para $\theta \neq 0$ se tiene, de (1.17),

$$\alpha(t)\beta(s) + \beta(t) = \alpha(s)\beta(t) + \beta(s) = \beta(st).$$

Si fijamos $s_0 \neq 1$ (pues de lo contrario $\alpha(s_0) = s_0^{-\theta} = 1$) se obtiene

$$\alpha(t)\beta(s_0) + \beta(t) = \alpha(s_0)\beta(t) + \beta(s_0),$$

de manera que, para $s_0 \neq 1$, $t \neq 1$

$$\left(\frac{\beta(s_0)}{1 - \alpha(s_0)}\right) = \frac{\beta(t)}{1 - \alpha(t)},$$

es decir, la función $\beta(\cdot)/(1 - \alpha(\cdot))$ es igual a una constante c . Entonces, para $t \neq 1$

$$\beta(t) = \left(\frac{\beta(s_0)}{1 - \alpha(s_0)} \right) (1 - \alpha(t)) = c(1 - t^\theta).$$

De esta manera, (1.15) se convierte en este caso en

$$H^t(x) = H(t^{-\theta}x + c(1 - t^{-\theta})) = H(t^{-\theta}(x - c) + c);$$

haciendo el cambio de variable $y = x - c$ obtenemos

$$H^t(y + c) = H(t^{-\theta}y + c).$$

Sea $G(y) = H(y + c)$. Entonces G y H son del mismo tipo, por lo que basta resolver para G . La función G es no-degenerada y satisface

$$G^t(y) = G(t^{-\theta}y). \quad (1.20)$$

Si hacemos $y = 0$ y sacamos logaritmos en la expresión (1.20), tenemos, para $t > 0$, que $t \log G(0) = \log G(0)$, de donde $\log G(0) = 0$ ó $-\infty$; es decir, $G(0) = 0$ ó 1 . Sin embargo, $G(0) = 1$ es imposible, porque implicaría la existencia de $x < 0$ para el cual el lado izquierdo de (1.20) es decreciente en t , mientras que el lado derecho de (1.20) es creciente en t . Concluimos entonces que $G(0) = 0$.

También de la expresión (1.20) se obtiene, para $y = 1$, $G^t(1) = G(t^{-\theta})$. Con esto es fácil ver que si tuviéramos que $G(1) = 0$, entonces $G \equiv 0$; y si $G(1) = 1$, entonces $G \equiv 1$. En ambos casos se contradice que G sea no-degenerada, por lo que debe darse $G(1) \in (0, 1)$.

Sea $\theta^{-1} = \alpha$, $H(1) = \exp\{-p^{-\alpha}\}$, $u = t^{-\theta}$ (nótese que esto implica que $u^{-\alpha} = (t^{-\theta})^{-1/\theta} = t$). Al sustituir esto en (1.20) con $y = 1$ se obtiene, para $u > 0$

$$G(u) = (\exp\{-p^{-\alpha}\})^t = \exp\{-p^{-\alpha}t\} = \exp\{-(pu)^{-\alpha}\} = \Phi_\alpha(pu).$$

Caso (c), $\theta < 0$.

De la misma manera que en el caso (b), se obtiene (1.20), de modo que, al hacer $y = 0$ en esta expresión tenemos que también en este caso debe darse $G(0) = 0$ ó 1. Pero $G(0) = 0$ es imposible, pues implicaría la existencia de $x > 0$ tal que el lado izquierdo de (1.20) es decreciente en t , mientras que el lado derecho de (1.20) es creciente en t . Por lo tanto, $G(0) = 1$. Si ahora hacemos $y = -1$, tendremos que $G^t(-1) = G(-t^{-\theta})$. Pero

$$\text{Si } G(-1) = 0 \Rightarrow G(-t^{-\theta}) = 0^t \equiv 0 \quad \forall t \in \mathbb{R},$$

$$\text{Si } G(-1) = 1 \Rightarrow G(-t^{-\theta}) = 1^t \equiv 1 \quad \forall t \in \mathbb{R}.$$

En ambos casos se contradice que G sea no-degenerada, por lo que $G(-1) \in (0, 1)$.

Sea $\theta^{-1} = -\alpha$, $H(-1) = \exp\{-(-p)^\alpha\}$, $u = t^{-\theta}$ (nótese que esto implica que $u^\alpha = (t^{-\theta})^{-1/\theta} = t$). Al sustituir esto en (1.20) con $y = -1$ se obtiene, para $u > 0$

$$G(-u) = (\exp\{-(-p)^\alpha\})^t = \exp\{-(-p)^\alpha t\} = \exp\{-(-pu)^\alpha\} = \Psi_\alpha(pu). \blacksquare$$

Esencialmente, el teorema nos dice que si tenemos máximos apropiadamente normalizados que convergen en distribución a un límite no-degenerado, entonces este límite debe ser una distribución de valores extremos. En este sentido, el Teorema de Fisher-Tippet nos proporciona un resultado para máximos del mismo tipo del que ofrece el Teorema del Límite Central para sumas.

Mencionamos antes que las distribuciones Fréchet Φ_α , Weibull Ψ_α y Gumbel Λ —esto es, las tres posibles distribuciones límites no-degeneradas para máximos de

v.a.i.i.d apropiadamente normalizados— son max-estables, es decir, satisfacen que sus máximos son del mismo tipo que la distribución original. El siguiente Teorema nos dice que no existe ningún otro tipo de distribución max-estable que sea no-degenerada.

Teorema 1.2.14 (Propiedad límite de las distribuciones max-estables) *La clase de las distribuciones max-estables coincide con la clase de las posibles distribuciones límite (no-degeneradas) para máximos de v.a.i.i.d., con la normalización apropiada.*

Demostración:

Supongamos que (X_n) es una sucesión de v.a.i.i.d max-estables. Entonces podemos reescribir (1.13) de la siguiente manera:

$$c_n^{-1} (M_n - d_n) \stackrel{d}{=} X.$$

De manera que toda distribución max-estable es una distribución límite para máximos de v.a.i.i.d.

Falta demostrar que la distribución límite de máximos apropiadamente normalizados es max-estable. Para ello, supongamos que, para constantes adecuadas y para alguna f.d. H no degenerada, se tiene que

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

Puesto que, de acuerdo con el Teorema de Fisher-Tippett, las posibles f.d. límite H son funciones continuas en \mathbb{R} , tenemos que, para cada $k \in \mathbb{N}$

$$\lim_{n \rightarrow \infty} F^{nk}(c_n x + d_n) = \left(\lim_{n \rightarrow \infty} F^n(c_n x + d_n) \right)^k = H^k(x), \quad x \in \mathbb{R}.$$

Por otra parte,

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

Esto significa, por el Teorema de Convergencia a Tipos (ver Apéndice A), que existen constantes $\tilde{c}_k > 0$ y $\tilde{d}_k \in \mathbb{R}$ tales que

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

y, para v.a.i.i.d. Y_1, \dots, Y_k con f.d. H ,

$$\max(Y_1, \dots, Y_k) = \tilde{c}_k Y_1 + \tilde{d}_k. \quad \blacksquare$$

Introduciremos ahora una f.d. que nos permitirá reescribir en una sola expresión a las f.d. Fréchet, Weibull y Gumbel.

Definición 1.2.15 (Distribución VEG) La función de distribución de Valores Extremos Generalizada (VEG) está dada por

$$H_\xi(x) = \begin{cases} \exp\left(-(1 + \xi x)^{-\frac{1}{\xi}}\right) & \text{si } \xi \neq 0 \\ \exp(-e^{-x}) & \text{si } \xi = 0 \end{cases} \quad (1.21)$$

donde $(1 + \xi x) > 0$ y ξ es el parámetro de forma.

El nombre de la VEG proviene del hecho de que Fréchet, la Weibull y la Gumbel son, de acuerdo con el Teorema de Fisher-Tippet, los tres tipos de distribuciones de valores extremos, las cuales pueden obtenerse de la expresión (1.21) para la distribución VEG de la siguiente manera:

$$\begin{aligned} \xi > 0 \quad \text{Fréchet} \quad H_\xi\left(\frac{x-1}{\xi}\right) &= \Phi_{\frac{1}{\xi}}(x) \\ \xi = 0 \quad \text{Gumbel} \quad H_0(x) &= \Lambda(x) \\ \xi < 0 \quad \text{Weibull} \quad H_\xi\left(\frac{-(x+1)}{\xi}\right) &= \Psi_{-\frac{1}{\xi}}(x) \end{aligned}$$

Observación 1.2.16 *La parametrización dada por la VEG es continua en ξ .*

Demostración:

En efecto, $\lim_{\xi \uparrow 0} \exp\left(- (1 + \xi x)^{-\frac{1}{\xi}}\right) = \exp(-e^{-x}) = \lim_{\xi \downarrow 0} \exp\left(- (1 + \xi x)^{-\frac{1}{\xi}}\right)$. ■

También pueden introducirse parámetros de localización μ y de escala $\sigma > 0$ en la VEG. Para ello, se define

$$H_{\xi, \mu, \sigma}(x) := H_{\xi}\left(\frac{x - \mu}{\sigma}\right)$$

Así, se tiene que $H_{\xi, \mu, \sigma}$ es del tipo H_{ξ} .

Utilizando la definición de la VEG y los resultados de distribuciones max-estables podemos reenunciar el Teorema de Fisher-Tippett de dos maneras alternativas: Sean X_1, X_2, \dots, X_n v.a.i.i.d. con función de distribución F , y sean $(a_n), (b_n)$ constantes tales que para alguna distribución límite H no degenerada se tiene la relación (1.14). Entonces:

- H es una distribución max-estable.
- H es del tipo de una distribución VEG H_{ξ} para algún $\xi \in \mathbb{R}$.

Tenemos pues que, de la misma manera que en el Teorema del Límite Central la distribución Normal aparece como límite asintótico para la suma de v.a.i.i.d., sin importar cuál era su distribución original, en el Teorema de Fisher-Tippett las tres distribuciones de valores extremos o, equivalentemente, las tres distribuciones max-estables —a saber, la Fréchet Φ_{α} , la Weibull Ψ_{α} y la Gumbel Λ — aparecen como únicos límites no-degenerados posibles para los máximos M_n de v.a.i.i.d., bajo la normalización adecuada. En la siguiente sección veremos que no necesitamos conocer

a detalle a la f.d. F para saber a cuál de las tres posibles distribuciones límites dará lugar; este hecho quedará determinado por las propiedades de la cola \overline{F} para valores grandes de x .

1.3 Dominio de Atracción del Máximo de las Distribuciones de Valores Extremos.

Ahora que sabemos que las distribuciones de valores extremos son las únicas distribuciones límite para máximos apropiadamente normalizados de v.a.i.d (por el Teorema de Fisher-Tippett), queremos saber bajo qué condiciones sobre F los máximos M_n convergen débilmente a una distribución de valores extremos H . Además, necesitamos un método para elegir las constantes $c_n > 0$, $d_n \in \mathbb{R}$, tales que $c_n^{-1}(M_n - d_n) \xrightarrow{d} H$.

Sabemos, por el Teorema de Convergencia a Tipos (ver Apéndice A), que la elección de constantes $c_n \neq c'_n$, $d_n \neq d'_n$ no puede ocasionar que $c_n^{-1}(M_n - d_n) \xrightarrow{d} H_1$ y $(c'_n)^{-1}(M_n - d'_n) \xrightarrow{d} H_2$ para distribuciones de valores extremos H_1 y H_2 de distintos tipos, ya que la distribución límite está determinada de manera única, salvo por transformaciones afines.

Definición 1.3.1 *Decimos que una v.a. X [su distribución F] pertenece al Dominio de Atracción del Máximo (DAM) de la distribución de valores extremos H si existen constantes $c_n > 0$, $d_n \in \mathbb{R}$ tales que $c_n^{-1}(M_n - d_n) \xrightarrow{d} H$. Denotaremos esto $X \in \text{DAM}(H)$ [$F \in \text{DAM}(H)$].*

Esto quiere decir que reunimos en una sola clase a todas las distribuciones F cuyos

máximos normalizados tienen el mismo tipo de distribución límite.

La siguiente proposición es una consecuencia de la aproximación de Poisson, dada en la Proposición 1.1.6.

Proposición 1.3.2 *La f.d. $F \in DAM(H)$, con H una distribución de valores extremos y constantes $c_n > 0$, $d_n \in \mathbb{R}$ si y sólo si*

$$\lim_{n \rightarrow \infty} n\overline{F}(c_n x + d_n) = -\ln H(x), \quad x \in \mathbb{R} \quad (1.22)$$

(Cuando $H(x) = 0$, el límite se interpreta como ∞).

A continuación se presentan dos definiciones que resultan útiles para la caracterización del Dominio de Atracción del Máximo de las tres distribuciones de valores extremos y para la elección de las constantes $c_n > 0$, $d_n \in \mathbb{R}$, tales que $c_n^{-1}(M_n - d_n) \xrightarrow{d} H_\xi$.

Definición 1.3.3 (Equivalencia de Colas) *Dos f.d. F y G se llaman de cola equivalente si tienen el mismo extremo derecho (es decir. $x_F = x_G$) y, para alguna constante $c \in (0, \infty)$*

$$\lim_{x \uparrow x_F} \frac{\overline{F}(x)}{\overline{G}(x)} = c \quad (1.23)$$

Definición 1.3.4 (Función cuantil o inverso generalizado) *El inverso generalizado de la f.d. F es*

$$F^{-}(t) = \inf \{x \in \mathbb{R} : F(x) \geq t\}, \quad 0 < t < 1 \quad (1.24)$$

La cantidad $x_t = F^{-}(t)$ define el t -cuantil de F .

1.3.1 El Dominio de Atracción del Máximo de la Distribución

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

En esta sección daremos una caracterización del Dominio de Atracción del Máximo de la distribución Fréchet ($DAM(\Phi_\alpha)$); veremos que, para pertenecer al $DAM(\Phi_\alpha)$, las f.d. deben satisfacer que su extremo derecho x_F sea infinito; es precisamente por esto que la Fréchet resulta ser un modelo apropiado para distribuciones de cola pesada. Veremos también que la condición más importante para pertenecer al $DAM(\Phi_\alpha)$ es que la f.d. F sea de variación regular (ver Apéndice B). Daremos, asimismo, la forma de las constantes normantes c_n y d_n bajo las cuales se satisface el Teorema de Fisher-Tippett (ver 1.14) en el caso Fréchet.

Puesto que M_n es la versión empírica del $(1 - n^{-1})$ -cuantil de la distribución subyacente F , esto nos sugiere que este cuantil podría utilizarse para normalizar M_n . El siguiente teorema nos confirma esta observación. Para caracterizar el $DAM(\Phi_\alpha)$, usaremos las nociones de variación lenta y variación regular, que se denotarán \mathcal{R}_0 y \mathcal{R}_α , respectivamente (ver Apéndice B).

Teorema 1.3.5 ($DAM(\Phi_\alpha)$) *La f.d. $F \in DAM(\Phi_\alpha)$, $\alpha > 0$, si y sólo si $\bar{F}(x) = x^{-\alpha}L(x)$ para alguna función $L \in \mathcal{R}_0$.*

Si $F \in DAM(\Phi_\alpha)$, entonces $c_n^{-1}M_n \xrightarrow{d} \Phi_\alpha$, donde $c_n = F^{-1}(1 - n^{-1})$.

Antes de dar la demostración de este Teorema haremos algunas observaciones.

Observación 1.3.6

$$c_n = F^{-1}(1 - n^{-1}) = \inf \{x \in \mathbb{R} : F(x) \geq 1 - n^{-1}\}$$

$$\begin{aligned}
&= \inf \{x \in \mathbb{R} : n^{-1} \geq 1 - F(x) = \overline{F}(x)\} \\
&= \inf \left\{x \in \mathbb{R} : \left(\frac{1}{\overline{F}}\right)(x) \geq n\right\} = \left(\frac{1}{\overline{F}}\right)^{\leftarrow}(n)
\end{aligned}$$

Observación 1.3.7 $\overline{F}(x) = x^{-\alpha}L(x)$ es equivalente a $\overline{F} \in \mathcal{R}_{-\alpha}$.

En efecto,

$$\begin{aligned}
\overline{F}(x) = x^{-\alpha}L(x) &\Leftrightarrow \lim_{x \rightarrow \infty} \frac{\overline{F}(xt)}{\overline{F}(x)} = \lim_{x \rightarrow \infty} \frac{(xt)^{-\alpha}L(xt)}{x^{-\alpha}L(x)} = \lim_{x \rightarrow \infty} \frac{x^{-\alpha}t^{-\alpha}L(xt)}{x^{-\alpha}L(x)} \\
&= \lim_{x \rightarrow \infty} \frac{t^{-\alpha}L(xt)}{L(x)} = t^{-\alpha} \lim_{x \rightarrow \infty} \frac{L(xt)}{L(x)} = t^{-\alpha} \Leftrightarrow \overline{F} \in \mathcal{R}_{-\alpha}
\end{aligned}$$

Observación 1.3.8 $d_n = 0$, es decir, no es necesario centrar a los máximos cuando su distribución subyacente cumple $F \in \text{DAM}(\Phi_\alpha)$.

Observación 1.3.9 Este resultado implica en particular que toda $F \in \text{DAM}(\Phi_\alpha)$ tiene extremo derecho infinito.

Si suponemos que existe $G \in \text{DAM}(\Phi_\alpha)$, $\alpha > 0$, tal que $x_G < \infty$, entonces, para toda $z > x_G$, $G(z) = 1$, y $\overline{G}(z) = 0$. Pero entonces $\lim_{x \rightarrow \infty} \frac{\overline{G}(xt)}{\overline{G}(x)} = 1$, lo que nos diría que $\overline{G} \in \mathcal{R}_0$, contradiciendo el que $G \in \text{DAM}(\Phi_\alpha)$.

Procedemos ahora a dar la Demostración del Teorema 1.3.5.

Demostración:

“ \Leftarrow ” Sea $\overline{F} \in \mathcal{R}_{-\alpha}$ para algún $\alpha > 0$ y c_n como se pide en el Teorema. Entonces

$$\begin{aligned}
\overline{F}(c_n) &= 1 - F(c_n) = 1 - F(F^{\leftarrow}(1 - n^{-1})) \geq 1 - (1 - n^{-1}) = n^{-1} \\
&\Rightarrow \overline{F}(c_n) \sim n^{-1}, \quad n \rightarrow \infty
\end{aligned}$$

Así, $\overline{F}(c_n) \rightarrow 0$, lo que nos dice que $c_n \rightarrow \infty$.

Para $x > 0$, tenemos, debido a que $\bar{F} \in \mathcal{R}_{-\alpha}$, lo siguiente:

$$\bar{F}(c_n) \sim \frac{1}{n} \Rightarrow n \sim \frac{1}{\bar{F}(c_n)} \Rightarrow n \sim \frac{\bar{F}(c_n x)}{\bar{F}(c_n) \bar{F}(c_n)} \Rightarrow n \bar{F}(c_n x) \sim \frac{\bar{F}(c_n x)}{\bar{F}(c_n)} \xrightarrow{n \rightarrow \infty} x^{-\alpha}$$

Para $x < 0$,

$$F^n(c_n x) \leq F^n(0) \rightarrow 0, \quad (1.25)$$

porque la variación regular requiere que $F(0) < 1$. (De lo contrario se tendría $\bar{F}(0) = 0$ y $\bar{F}(y) = 0$ para todo $y \geq 0$, lo que querría decir que $\bar{F} \in \mathcal{R}_0$).

Como las distribuciones de valores extremos son continuas en \mathbb{R} , entonces $c_n^{-1}(M_n - d_n) \xrightarrow{d} H$ es equivalente a tener

$$\lim_{n \rightarrow \infty} P(M_n \leq c_n x + d_n) = \lim_{n \rightarrow \infty} F^n(c_n x + d_n) = H(x), \quad x \in \mathbb{R}$$

Así, (1.25) nos dice que, para $x < 0$, debe darse $H(x) = 0$. Y de (1.22) y el caso $x > 0$, tenemos que $\lim_{n \rightarrow \infty} n \bar{F}(c_n x) = -\ln H(x)$ y que $\lim_{n \rightarrow \infty} n \bar{F}(c_n x) = x^{-\alpha}$, de donde

$$-\ln H(x) = x^{-\alpha} \Leftrightarrow H(x) = \exp\{-x^{-\alpha}\}.$$

Por lo tanto, $F \in DAM(\Phi_\alpha)$.

" \Rightarrow " Supongamos que $\lim_{n \rightarrow \infty} F^n(c_n x + d_n) = \Phi_\alpha(x)$ para todo $x > 0$ y $c_n > 0$, $d_n \in \mathbb{R}$ apropiadas. Entonces, para $s > 0$, $x > 0$ se tiene

$$\begin{aligned} \lim_{n \rightarrow \infty} F^n(c_{[ns]} x + d_{[ns]}) &= \Phi_\alpha^{1/s}(x) = (\exp\{-x^{-\alpha}\})^{1/s} = \exp\left\{\frac{1}{s}(-x^{-\alpha})\right\} \\ &= \exp\left\{-(xs^{1/\alpha})^{-\alpha}\right\} = \Phi_\alpha(s^{1/\alpha} x). \end{aligned}$$

Y, por el Teorema de Convergencia a Tipos (ver Apéndice A), esto significa

$$\frac{c_{[ns]}}{c_n} \rightarrow s^{1/\alpha} \quad \text{y} \quad \frac{d_{[ns]} - d_n}{c_n} \rightarrow 0, \quad s > 0. \quad (1.26)$$

Nótese que (1.26) es justamente la definición de que (c_n) es una sucesión de variación regular con índice $1/\alpha$. En particular, esto significa que (ver Apéndice B) $c_n \rightarrow \infty$.

Supondremos que $d_n = 0$.

Por (1.22), $\lim_{n \rightarrow \infty} F^n(c_n x) = \Phi_\alpha(x) \Leftrightarrow \lim_{n \rightarrow \infty} n \bar{F}(c_n x) = -\ln \Phi_\alpha(x) = x^{-\alpha}$.

Esto nos dice (ver Apéndice B) que $\bar{F} \in \mathcal{R}_{-\alpha}$.

Si suponemos que $d_n \neq 0$, hay que demostrar primero que $d_n/c_n \rightarrow 0$. La demostración en este caso es más complicada y puede encontrarse en [22]. ■

Con esto se ha demostrado que

$$F \in DAM(\Phi_\alpha) \Leftrightarrow \bar{F} \in \mathcal{R}_{-\alpha}$$

Nótese que, por propiedades de $\bar{F} \in \mathcal{R}_{-\alpha}$ (ver Apéndice B), el Dominio de Atracción del Máximo de la Fréchet contiene distribuciones de cola muy pesada, en el sentido de que $E[X^+]^\delta = \infty$ para $\delta > \alpha$, donde $X^+ = \max\{X, 0\}$, . Por ello, esta clase de distribuciones puede ser apropiada para modelar grandes reclamaciones de seguros y fluctuaciones de precios, *log-returns*, etc.

El teorema anterior y las propiedades de $\bar{F} \in \mathcal{R}_{-\alpha}$ (ver Apéndice B) dan lugar al siguiente resultado:

Corolario 1.3.10 (Condición de von Mises) *Sea F una f.d. absolutamente continua cuya densidad f satisface*

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

Entonces $F \in DAM(\Phi_\alpha)$.

El siguiente resultado afirma que $DAM(\Phi_\alpha)$ es cerrado respecto a equivalencia de colas, así que si F y G son de cola equivalente, tendremos $F \in DAM(\Phi_\alpha) \Leftrightarrow G \in DAM(\Phi_\alpha)$. Además, podemos utilizar las mismas constantes normantes para F y para G .

Proposición 1.3.11 (Cerradura del $DAM(\Phi_\alpha)$) Sean F y G f.d. y supongamos que $F \in DAM(\Phi_\alpha)$ con constantes normantes $c_n > 0$, es decir

$$\lim_{n \rightarrow \infty} F^n(c_n x) = \Phi_\alpha(x), \quad x > 0. \quad (1.27)$$

Entonces

$$\lim_{n \rightarrow \infty} G^n(c_n x) = \Phi_\alpha(cx), \quad x > 0 \quad (1.28)$$

para algún $c > 0$ si y sólo si F y G tienen cola equivalente con

$$\lim_{x \rightarrow \infty} \frac{\overline{F}(x)}{\overline{G}(x)} = c^\alpha.$$

Demostración:

“ \Rightarrow ” Supongamos que sabemos que se dan (1.27) y (1.28) y queremos mostrar que F y G tienen cola equivalente. Sean

$$V_F(t) = \left(\frac{1}{1-F} \right)^{\leftarrow}(t), \quad V_G(t) = \left(\frac{1}{1-G} \right)^{\leftarrow}(t).$$

Así, (1.27) y (1.28) pueden reescribirse en términos de inversos de la siguiente manera:

$$\begin{aligned} \lim_{t \rightarrow \infty} (V_F(ty)/c(t)) &= \left(\frac{1}{-\log \Phi_\alpha(x)} \right)^{\leftarrow}(y) = y^{1/\alpha}, \\ \lim_{t \rightarrow \infty} (V_G(ty)/c(t)) &= \left(\frac{1}{-\log \Phi_\alpha(cx)} \right)^{\leftarrow}(y) = c^{-1}y^{1/\alpha}, \end{aligned}$$

lo cual significa que $V_F(t) \sim cV_G(t)$. Ahora, puesto que $V_F \in \mathcal{R}_{1/\alpha}$, al invertir se obtiene

$$\frac{1}{1-F(t)} \sim c^{-\alpha} \left(\frac{1}{1-G(t)} \right),$$

es decir, $\lim_{t \rightarrow \infty} (1-F(t))/(1-G(t)) = c^\alpha$.

" \Leftarrow " Supongamos que $\overline{F}(x) \sim q\overline{G}(x)$ si $x \rightarrow \infty$ para alguna $q > 0$. Por (1.22), el límite (1.27) es equivalente a

$$\lim_{n \rightarrow \infty} n\overline{F}(c_n x) = x^{-\alpha}, \quad x > 0.$$

Para tales x , $c_n x \rightarrow \infty$ si $n \rightarrow \infty$ y, por equivalencia de colas, $n\overline{G}(c_n x) \sim nq^{-1}\overline{F}(c_n x) \rightarrow q^{-1}x^{-\alpha}$. Lo que, significa, por (1.22), que

$$\lim_{n \rightarrow \infty} G^n(c_n x) = \exp \left\{ - (q^{1/\alpha} x)^{-\alpha} \right\} = \Phi_\alpha (q^{1/\alpha} x).$$

Basta ahora tomar $c = q^{1/\alpha}$. ■

Ejemplo 1.3.12 (Distribución Fréchet) Consideremos la f.d. Fréchet

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

con $\alpha > 0$.

Notemos que

$$\Phi_\alpha^n(n^{1/\alpha}x) = \left[\exp \left\{ (-xn^{1/\alpha})^{-\alpha} \right\} \right]^n = \left[\exp \left\{ \frac{1}{n} (-x)^{-\alpha} \right\} \right]^n = \exp(-x^{-\alpha}) = \Phi_\alpha(x).$$

Esto nos indica que $\Phi_\alpha \in DAM(\Phi_\alpha)$ y, más aún,

$$P \{ M_n \leq n^{1/\alpha}x \} = \Phi_\alpha(x),$$

para todo n , de manera que las constantes normantes para los máximos de la f.d. Fréchet son

$$c_n = n^{1/\alpha}, \quad d_n = 0.$$

Ejemplo 1.3.13 (Distribuciones Pareto y similares) Como ya se ha mencionado, la f.d. Pareto está dada por $F(x) = 1 - Kx^{-\alpha}$, $\alpha, K > 0$, $x \geq K^{1/\alpha}$, de manera que su cola es $\bar{F}(x) = Kx^{-\alpha}$.

Por su parte, la cola de la f.d. Burr está dada por $\bar{F}(x) = \left(\frac{k}{k+x^\tau}\right)^\alpha$, $\alpha, k, \tau > 0$.

En ambos casos, tenemos que las colas derechas de estas f.d. son de la forma $\bar{F}(x) \sim Kx^{-\alpha}$, $x \rightarrow \infty$, para alguna $K, \alpha > 0$. Es en este sentido que decimos que la Burr es una distribución similar a la Pareto.

Tenemos

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(xt)}{\bar{F}(x)} = \frac{K(xt)^{-\alpha}}{Kx^{-\alpha}} = t^{-\alpha}, \quad t > 0.$$

Por lo tanto $\bar{F} \in \mathcal{R}_{-\alpha}$, lo que implica $F \in DAM(\Phi_\alpha)$. Podemos escoger como constantes normantes

$$\begin{aligned} c_n &= \left(\frac{1}{\bar{F}}\right)^{\leftarrow}(n) = \inf \left\{ x \in \mathbb{R} : \left(\frac{1}{\bar{F}}\right)(x) \geq n \right\} = \inf \left\{ x \in \mathbb{R} : \frac{1}{n} \geq \bar{F}(x) \right\} \\ &= \inf \left\{ x \in \mathbb{R} : \frac{1}{n} \geq Kx^{-\alpha} \right\} = \inf \left\{ x \in \mathbb{R} : x^\alpha \geq nK \right\} \\ &= \inf \left\{ x \in \mathbb{R} : x \geq (nK)^{1/\alpha} \right\} = (nK)^{1/\alpha}. \end{aligned}$$

De donde $(nK)^{1/\alpha} M_n \xrightarrow{d} \Phi_\alpha$.

Ejemplo 1.3.14 (Máximo de v.a.s Cauchy) Sea (X_i) una sucesión de v.a.i.i.d. Cauchy. La distribución Cauchy es absolutamente continua y su densidad está dada

por

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

Usando la regla de l'Hôpital se obtiene

$$\lim_{x \rightarrow \infty} \frac{1 - F(x)}{(\pi x)^{-1}} = \lim_{x \rightarrow \infty} \frac{f(x)}{\pi^{-1}x^{-2}} = \lim_{x \rightarrow \infty} \frac{\pi x^2}{\pi(1+x^2)} = 1,$$

lo que quiere decir que $\bar{F}(x) \sim (\pi x)^{-1}$. Claramente, $\bar{F} \in \mathcal{R}_{-1}$, lo que significa que $F \in DAM(\Phi_1)$. Basta entonces resolver $\bar{F}(x) = n^{-1}$, es decir, resolver $(\pi x)^{-1} = n^{-1}$ para así obtener $c_n = n/\pi$. Esto implica que

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

1.3.2 El Dominio de Atracción del Máximo de la Distribución

Weibull $\Psi_\alpha(x) = \exp \{ -(-x)^\alpha \}$, $\alpha > 0$.

Un hecho importante, que no resulta en absoluto obvio, es que todas las f.d. F en el $DAM(\Psi_\alpha)$ tienen extremo derecho x_F finito. Recordemos que, como se afirmó en la

Proposición 1.2.4, Ψ_α y Φ_α están cercanamente relacionadas; en efecto,

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

Así, es de esperarse que haya también una relación cercana entre el $DAM(\Psi_\alpha)$ y el $DAM(\Phi_\alpha)$. El siguiente teorema confirma esta idea intuitiva.

Teorema 1.3.15 ($DAM(\Psi_\alpha)$) *La f.d. F pertenece al dominio de atracción del máxi-*

~~*mo de Ψ_α , $\alpha > 0$, si y sólo si $x_F < \infty$ y $\bar{F}(x_F - x^{-1}) = x^{-\alpha}L(x)$ para alguna función*~~

L de variación lenta.

Si $F \in DAM(\Psi_\alpha)$, entonces

$$c_n^{-1} (M_n - d_n) \xrightarrow{d} \Psi_\alpha,$$

donde las constantes normantes c_n pueden elegirse como $c_n = x_F - F^{-(1-n^{-1})}$ y $d_n = x_F$.

Demostración:

" \Rightarrow " Esta parte es larga y complicada, pero puede encontrarse en [22].

" \Leftarrow " Supongamos que $x_F < \infty$ y $\bar{F}(x_F - x^{-1}) = x^{-\alpha}L(x)$. Definamos

$$F_\bullet(x) = F(x_F - x^{-1}), \quad x > 0. \quad (1.29)$$

Entonces $\bar{F}_\bullet(x) = x^{-\alpha}L(x)$, es decir, $\bar{F}_\bullet \in \mathcal{R}_{-\alpha}$. Así, por el Teorema 1.3.5, se tiene $F_\bullet \in DAM(\Phi_\alpha)$, con constantes normantes $c_n^\bullet = F_\bullet^{-(1-n^{-1})}$ y $d_n^\bullet = 0$. Ahora bien, $F_\bullet \in DAM(\Phi_\alpha)$ significa que

$$F_\bullet^n(c_n^\bullet x) \rightarrow \Phi_\alpha(x), \quad x > 0,$$

es decir,

$$F^n(x_F - (c_n^\bullet x)^{-1}) \rightarrow \exp\{-x^{-\alpha}\}, \quad x > 0.$$

Si hacemos la substitución $x = y^{-1}$, entonces

$$F^n\left(x_F + \frac{y}{c_n^\bullet}\right) \rightarrow \exp\{-(-y)^\alpha\}, \quad y < 0. \quad (1.30)$$

De aquí es claro que $d_n = x_F$. Finalmente,

$$\begin{aligned} c_n^\bullet &= F_\bullet^{-(1-n^{-1})} = \inf\{x \in \mathbb{R} : F(x_F - x^{-1}) \geq 1 - n^{-1}\} \\ &= \inf\{(x_F - u)^{-1} : F(u) \geq 1 - n^{-1}\} = (x_F - \inf\{u : F(u) \geq 1 - n^{-1}\})^{-1} \\ &= (x_F - F^{-(1-n^{-1})})^{-1}. \end{aligned}$$

Al sustituir este valor de c_n^* en (1.30), se obtiene

$$F^n \left(x_F + \frac{y}{(x_F - F^{-(1-n^{-1})})^{-1}} \right) \longrightarrow \exp \{ -(-y)^\alpha \}, \quad y < 0.$$

$$F^n (x_F + y (x_F - F^{-(1-n^{-1})})) \longrightarrow \exp \{ -(-y)^\alpha \}, \quad y < 0.$$

Pero esto puede escribirse como

$$(x_F - F^{-(1-n^{-1})})^{-1} (M_n - x_F) \xrightarrow{d} \Psi_\alpha,$$

lo que prueba el Teorema. ■

En consecuencia, hemos visto que

$$F \in \text{DAM}(\Psi_\alpha) \iff x_F < \infty, \quad \bar{F}(x_F - x^{-1}) \in \mathcal{R}_{-\alpha}.$$

Así, el $\text{DAM}(\Psi_\alpha)$ está conformado por f.d. F cuyo soporte está acotado por la derecha.

Precisamente porque $x_F < \infty$, pudieran no ser la mejor elección para modelar eventos

extremos en seguros y en finanzas. Si bien es claro que en la práctica siempre hay

un límite superior (aunque éste puede ser excesivamente alto), es posible que no

queramos incorporar el parámetro adicional x_F a nuestro modelo. Además, casi

siempre es preferible utilizar distribuciones con $x_F = \infty$, porque así se permiten

valores arbitrariamente grandes para una muestra. Tales distribuciones pertenecen

típicamente al $\text{DAM}(\Phi_\alpha)$ o al $\text{DAM}(\Lambda)$.

También podemos caracterizar a las distribuciones que pertenecen al $\text{DAM}(\Psi_\alpha)$

~~a través de sus funciones de densidad. Para ello, basta recordar la transformación~~

(1.29) y aplicarla al Corolario 1.3.10 para obtener el siguiente Corolario.

Corolario 1.3.16 (Cond. de von Mises) Sea F una f.d. absolutamente continua, cuya densidad f es positiva en algún intervalo finito (z, x_F) . Si

$$\lim_{x \uparrow x_F} \frac{(x_F - x)f(x)}{\overline{F}(x)} = \alpha > 0, \quad (1.31)$$

entonces $F \in \text{DAM}(\Psi_\alpha)$.

Demostración:

Recordemos que $F \in \text{DAM}(\Psi_\alpha)$ si y sólo si $F_\bullet(x) = F(x_F - x^{-1}) \in \text{DAM}(\Phi_\alpha)$.

La Condición de von Mises para F_\bullet nos dice que

$$\alpha = \lim_{x \rightarrow \infty} \frac{x f_\bullet(x)}{\overline{F}_\bullet(x)}. \quad (1.32)$$

Al aplicar la transformación de F_\bullet a F en (1.32) se obtiene

$$\alpha = \lim_{x \rightarrow \infty} \frac{x f(x_F - x^{-1}) x^{-2}}{\overline{F}(x_F - x^{-1})}.$$

Pero esto sucede si y sólo si

$$\lim_{s \uparrow x_F} \frac{(x_F - s)f(s)}{\overline{F}(s)} = \alpha.$$

■

También aplicando la transformación (1.29), podemos reformular la Proposición 1.3.11 de la siguiente manera:

Proposición 1.3.17 (Cerradura del $\text{DAM}(\Psi_\alpha)$) Sean F y G distribuciones con extremos derechos $x_F = x_G < \infty$ y supongamos que $F \in \text{DAM}(\Psi_\alpha)$ con constantes normantes $c_n > 0$; esto es,

$$\lim_{n \rightarrow \infty} F^n(c_n x + x_F) = \Psi_\alpha(x), \quad x < 0.$$

Entonces

$$\lim_{n \rightarrow \infty} G^n(c_n x + x_F) = \Psi_\alpha(cx), \quad x < 0,$$

para alguna $c > 0$ si y sólo si F y G son de cola equivalente y satisfacen

$$\lim_{x \uparrow x_F} \overline{F}(x) / \overline{G}(x) = c^{-\alpha}.$$

Por otra parte, el Teorema de Representación para Funciones de Variación Regular (ver Apéndice B) implica que toda $F \in DAM(\Psi_\alpha)$ tiene cola equivalente a una f.d. absolutamente continua que satisface (1.31). Esta es una manera de caracterizar al $DAM(\Psi_\alpha)$.

Ejemplo 1.3.18 (Distribución Weibull) Como hemos visto antes, la f.d. Weibull Ψ_α , $\alpha > 0$ está dada por

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

Nótese que, para cada n ,

$$\Psi_\alpha^n(n^{-1/\alpha}x) = \left[\exp\left(-(-n^{-1/\alpha}x)^\alpha\right) \right]^n = \exp\left(-\frac{1}{n}(-x)^\alpha\right)^n = \exp(-(-x)^\alpha) = \Psi_\alpha(x),$$

de manera que $\Psi_\alpha \in DAM(\Psi_\alpha)$ y, además,

$$P[M_n \leq n^{-1/\alpha}x] = \Psi_\alpha(x),$$

lo que nos indica que las constantes normantes son, en este caso,

$$cn = n^{-1/\alpha}, \quad d_n = 0.$$

Ejemplo 1.3.19 (Distribución Uniforme en $(0, 1)$) Claramente, se tiene $x_F = 1$ y $\bar{F}(1 - x^{-1}) = 1 - F(1 - x^{-1}) = 1 - (1 - x^{-1}) = x^{-1} \in \mathcal{R}_{-1}$. Esto implica, por el Teorema 1.3.15, que $F \in DAM(\Psi_1)$. Ya que $\bar{F}(1 - n^{-1}) = n^{-1}$, se tiene $c_n = n^{-1}$. Entonces, se tiene

$$n(M_n - 1) \xrightarrow{d} \Psi_1.$$

1.3.3 El Dominio de Atracción del Máximo de la Distribución Gumbel $\Lambda(x) = \exp\{-e^{-x}\}$.

El Dominio de Atracción del Máximo de la distribución Gumbel Λ contiene f.d. con colas muy diferentes, que van desde las moderadamente pesadas (como la distribución lognormal) hasta ligeras (como la distribución normal). Además, se dan tanto $x_F = \infty$ como $x_F < \infty$. Los siguientes dos ejemplos nos dejan ver este hecho.

Ejemplo 1.3.20 En el Ejemplo 1.2.10 vimos que, para la distribución exponencial $F(x) = 1 - e^{-x}$, $x > 0$, se tiene que

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

es decir, $F \in DAM(\Lambda)$ con $c_n = 1$ y $d_n = \log n$. Este es un ejemplo de una f.d. en el $DAM(\Lambda)$ para la cual $x_F = \infty$.

Ejemplo 1.3.21 Un ejemplo de una f.d. $F \in DAM(\Lambda)$ donde $x_F < \infty$ es

$$F(x) = \begin{cases} 1 - e^{1/x} & \text{si } x < 0, \\ 1 & \text{si } x \geq 0. \end{cases}$$

Aquí, por la Proposición 1.1.6, para sucesiones $\{u_n\}$ tales que $ne^{1/u_n} \rightarrow \tau > 0$ se tiene que

$$P[M_n \leq u_n] \rightarrow e^{-\tau}.$$

Si tomamos $\tau = e^{-x}$ ($-\infty < x < \infty$) y $u_n = (\log \tau - \log n)^{-1}$, se sigue que

$$P(M_n \leq (\log \tau - \log n)^{-1}) = P(M_n \leq -(x + \log n)^{-1}) \rightarrow \exp\{-e^{-x}\} = \Lambda(x).$$

Empezaremos a estudiar el DAM(Λ) considerando un caso especial.

Definición 1.3.22 (Función de von Mises) Una f.d. F con extremo derecho x_F se llama una función de von Mises si existe $z_0 < x_F$ tal que para $z_0 < x < x_F$ y $c > 0$ se tiene la representación

$$\bar{F}(x) = c \exp \left\{ - \int_{z_0}^x \frac{1}{a(u)} du \right\}, \quad (1.33)$$

donde $a(u) > 0$, $z_0 < u < x_F$ y a es absolutamente continua en (z_0, x_F) con densidad $a'(u)$ y $\lim_{u \uparrow x_F} a'(u) = 0$. En este caso llamamos a a la función auxiliar de F .

Lema 1.3.23 Supóngase que $a(u)$ es una función auxiliar absolutamente continua con $\lim_{u \uparrow x_F} a'(u) = 0$.

(a) Si $x_F = \infty$, entonces $\lim_{t \rightarrow \infty} t^{-1} a(t) = 0$.

(b) Si $x_F < \infty$, entonces $a(x_F) = \lim_{t \uparrow x_F} a(t) = 0$ y $\lim_{t \uparrow x_F} (x_F - t)^{-1} a(t) = 0$.

En cualquier caso, para todo $x \in \mathbb{R}$ se tiene

$$\lim_{t \uparrow x_F} (t + xa(t)) = x_F.$$

Demostración:

(a) Cuando $t \rightarrow \infty$ se tiene

$$t^{-1}a(t) \sim t^{-1} \int_{z_0}^t a'(u) du$$

y, puesto que de acuerdo con la definición anterior el integrando tiende a cero, también $t^{-1}a(t)$ tiende a cero. De modo que

$$t + xa(t) = t(1 + xa(t)/t) \sim t, \quad t \rightarrow \infty.$$

(b) Si $x_F < \infty$, entonces $\bar{F}(x_F) = 1 - F(x_F) = 0$, así que, de (1.33),

$$\int_x^{x_F} \frac{1}{a(u)} du = \infty, \quad z_0 < x < x_F.$$

De aquí se deduce que para toda $x \in (z_0, x_F)$

$$\sup_{x \leq u \leq x_F} \frac{1}{a(u)} = \infty,$$

así que

$$\inf_{x \leq u \leq x_F} a(u) = 0.$$

Entonces, por continuidad existe una sucesión $u_n \uparrow x_F$ con $a(u_n) = 0$, de donde $a(x_F) = 0$.

Ahora, puesto que $a(x_F) = 0$,

$$\lim_{t \uparrow x_F} \frac{a(t)}{(x_F - t)} = \lim_{t \uparrow x_F} - \int_t^{x_F} \left(\frac{a'(u)}{(x_F - t)} \right) du. \quad (1.34)$$

Si hacemos el cambio de variables $y = x_F - u$ y $s = x_F - t$, la ecuación (1.34) se convierte en

$$\lim_{s \downarrow 0} -s^{-1} \int_0^s a'(x_F - y) dy,$$

el cual es cero, puesto que $a'(x_F - y) \rightarrow 0$ cuando $y \rightarrow 0$. Finalmente, puesto que $a(t) \rightarrow 0$ cuando $t \rightarrow x_F$, se tiene que

$$\lim_{t \uparrow x_F} t + xa(t) = x_F. \blacksquare$$

Lema 1.3.24 *Si a satisface las condiciones del Lema 1.3.23, entonces*

$$\lim_{t \uparrow x_F} \frac{a(t + xa(t))}{a(t)} = 1$$

localmente uniformemente para $x \in \mathbb{R}$.

Demostración:

Sea $x(t)$ una función tal que $\lim_{t \rightarrow x_F} x(t) = x \in \mathbb{R}$. Entonces

$$|a(t + x(t)a(t)) - a(t)| \leq \left| \int_t^{t+x(t)a(t)} a'(u) du \right|.$$

Puesto que por el Lema 1.3.23, cuando $t \rightarrow \infty$ se tiene $t + x(t)a(t) \rightarrow x_F$, y del hecho de que $a'(u) \rightarrow 0$ cuando $u \rightarrow x_F$, se sigue que, dado $\varepsilon > 0$

$$\left| \int_t^{t+x(t)a(t)} a'(u) du \right| \leq \varepsilon |x(t)a(t)|, \quad t \geq t_0(\varepsilon).$$

Así que

$$\left| \frac{a(t + x(t)a(t))}{a(t)} - 1 \right| \leq \varepsilon |x(t)|.$$

Pero $\varepsilon > 0$ es arbitrario y $|x(t)|$ está acotado, de donde se sigue el resultado. \blacksquare

Proposición 1.3.25 (a) *Si F es una función de von Mises con representación (1.33), entonces $F \in DAM(\Lambda)$. Las constantes normantes pueden escogerse como*

$$d_n = (1/(1-F))^{-(n)}$$

$$c_n = a(d_n).$$

(b) Supóngase que F es absolutamente continua y que su segunda derivada F'' es negativa para todo $x \in (z_0, x_F)$. Si

$$\lim_{x \rightarrow x_F} \frac{F''(x)(1-F(x))}{(F'(x))^2} = -1 \quad (1.35)$$

entonces F es una función de von Mises y $F \in DAM(\Lambda)$. Puede escogerse $a = (1-F)/F'$.

Demostración:

(a) De la caracterización (1.33) de una función de von Mises se tiene que, para $x \in \mathbb{R}$ y t suficientemente grande

$$\frac{1-F(t+xa(t))}{1-F(t)} = \exp \left\{ - \int_t^{t+xa(t)} \frac{1}{a(u)} du \right\}.$$

Si hacemos el cambio de variable $s = (u-t)/a(t)$ se obtiene

$$\frac{1-F(t+xa(t))}{1-F(t)} = \exp \left\{ - \int_0^x \frac{a(t)}{a(t+sa(t))} ds \right\}$$

y, puesto que el Lema 1.3.24 nos dice que el integrando converge a 1 uniformemente en $(0, x)$, se obtiene

$$\lim_{t \rightarrow x_F} \frac{1-F(t+xa(t))}{1-F(t)} = e^{-x}.$$

Sea $U = 1/(1-F)$. Entonces la expresión anterior puede reescribirse como

$$\frac{U(t+xa(t))}{U(t)} \rightarrow e^{-x}.$$

Como U es continua por la derecha, se tiene que $U(U^{-1}(t)) \sim t$, de donde

$$\frac{U(U^{-1}(n) + xa(U^{-1}(n)))}{n} \rightarrow e^x.$$

Haciendo $d_n = U^{-}(n) = (1/(1-F))^{-}(n)$ y $c_n = a(U^{-}(n)) = a(d_n)$, se obtiene

$$\frac{U(c_n x + d_n)}{n} \rightarrow e^x.$$

De manera que

$$n(1 - F(c_n x + d_n)) \rightarrow e^{-x},$$

lo cual, por la Proposición 1.1.6, es equivalente a $F \in DAM(\Lambda)$.

(b) Sea $1 - F = \exp\{-R\}$. Entonces, la representación dada en (1.33) es posible con $f = 1/R'$ y $f' \rightarrow 0$ si y sólo si $(1/R')' \rightarrow 0$. Pero $R = -\log(1 - F)$, de manera que $R' = F'/(1 - F)$ y $(1/R')' = (1 - F)/F'$. Así

$$\left(\frac{1}{R'}\right)' = \frac{-(F')^2 - (1 - F)F''}{(F')^2} = -1 - \frac{(1 - F)F''}{(F')^2},$$

de donde se sigue que

$$\lim_{x \uparrow x_F} \frac{F''(x)(1 - F(x))}{(F'(x))^2} = -1.$$

■

Ejemplo 1.3.26 Sea $F(x) = \Phi(x)$ la distribución normal estándar. Se tiene que

$$F'(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

$$F''(x) = \frac{-1}{\sqrt{2\pi}} x e^{-x^2/2} = -x\phi(x).$$

La razón de Mill nos dice que $1 - \Phi(x) \sim x^{-1}\phi(x)$, de donde

$$\lim_{x \rightarrow \infty} \frac{(1 - F(x))F''(x)}{(F'(x))^2} = \lim_{x \rightarrow \infty} \frac{-x^{-1}\phi(x)x\phi(x)}{(\phi(x))^2} = -1,$$

es decir, $\Phi(x)$ es una función de Von Mises (y por tanto $\Phi \in DAM(\Lambda)$) con función auxiliar

$$a(x) = \frac{(1 - F(x))}{(F'(x))} \sim \frac{x^{-1}\phi(x)}{\phi(x)} = x^{-1}.$$

Lema 1.3.27 *Supóngase que $F \in DAM(\Lambda)$ y sea $V := (1/(1-F))^-$. Sean $V_1(t) = t \int_t^\infty V(u)u^{-2}du$ y $V_2(t) = t \int_t^\infty V_1(u)u^{-2}du$. Definimos también*

$$(1/(1-F_i))^- = V_i^-, \quad i = 1, 2.$$

Entonces

$$1 - F(x) \sim c_i(1 - F_i(x)), \quad x \rightarrow x_F,$$

donde $c_1 = e^{-1}$, $c_2 = e^{-2}$.

Demostración:

Tenemos que $V \in \Pi$ (ver Apéndice B), de modo que, para $x > 0$, cuando $t \rightarrow \infty$,

$$\frac{V(tx) - V_1(t)}{a(t)} = \frac{V(tx) - V(t)}{a(t)} + \frac{V(t) - V_1(t)}{a(t)} \rightarrow (\log x) - 1.$$

Al invertir esta ecuación se obtiene, para $y \in \mathbb{R}$,

$$\frac{V^-(ya(t) + V_1(t))}{t} \rightarrow \exp\{y + 1\}.$$

Si hacemos $y = 0$

$$\lim_{t \rightarrow \infty} \frac{V^-(V_1(t))}{t} \rightarrow e,$$

de donde, notando que V_1 es continua y estrictamente creciente (ver Apéndice B), se

obtiene que

$$\lim_{s \rightarrow \infty} \frac{V^-(s)}{V_1^-(s)} = e,$$

lo que nos dice que

$$1 - F(x) \sim e^{-1}(1 - F_1(x)), \quad x \rightarrow x_F.$$

El resultado para F_2 se obtiene de la misma manera. ■

Teorema 1.3.28 (*Caracterización del DAM(Λ)*) $F \in DAM(\Lambda)$ si y sólo si existe una función de von Mises F^* tal que para $x \in (z_0, x_F)$

$$1 - F(x) = c(x) (1 - F^*(x)) = c(x) \exp \left\{ - \int_{z_0}^x \frac{1}{a(u)} du \right\} \quad (1.36)$$

y

$$\lim_{t \rightarrow x_F} c(t) = c > 0.$$

Demostración:

“ \Leftarrow ” Supongamos que se da (1.36). Entonces, por la Proposición 1.3.25, existen $a_n > 0$, $b_n \in \mathbb{R}$ tales que

$$n(1 - F^*(a_n x + b_n)) \rightarrow e^{-x},$$

de manera que

$$n(1 - F(a_n x + b_n)) \rightarrow ce^{-x},$$

lo cual, por (1.22), significa que

$$F^n(a_n x + b_n) \rightarrow \exp \{ ce^{-x} \},$$

es decir, $F \in DAM(\Lambda)$.

“ \Rightarrow ” Supongamos que $F \in DAM(\Lambda)$ y sea $F^* = F_2$ del Lema 1.3.27, de manera que $1 - F^* = 1/V_2^-$, así que basta probar que F^* es una función de von Mises. Sea $R = -\log(1 - F^*)$; debemos verificar, al igual que en (b) de la Proposición 1.3.25

que

$$(1/R)' \rightarrow 0.$$

Sin embargo

$$1/R' = \frac{1 - F^*}{F^{*\prime}} = \frac{1/V_2^-}{(V_2^-)' / (V_2^-)^2} = \frac{V_2^-}{(V_2^-)'} = V_2^- \cdot V_2'(V_2^-),$$

de manera que

$$(1/R')' = V_2^- \cdot \{V_2''(V_2^-)/V_2'(V_2^-)\} + V_2'(V_2^-)/V_2''(V_2^-).$$

De aquí que

$$\lim_{t \rightarrow x_F} (1/R'(t))' = \lim_{y \rightarrow \infty} (yV_2''(y)/V_2'(y)) + 1 = -1 + 1 = 0,$$

por la relación entre Π -variación y el $\overline{\text{DAM}}(\Lambda)$ (ver Apéndice B). Esto prueba que F^* es una función de von Mises, de lo cual se sigue el resultado. ■

La siguiente caracterización del $\overline{\text{DAM}}(\Lambda)$ en ocasiones es más útil que la del Teorema 1.3.28.

Corolario 1.3.29 *La f.d. F con extremo derecho $x_F \leq \infty$ pertenece al dominio de atracción del máximo de Λ si y sólo si existe alguna $z_0 < x_F$ tal que F puede representarse mediante*

$$\overline{F}(x) = c(x) \exp \left\{ - \int_{z_0}^x \frac{g(t)}{a(t)} dt \right\}, \quad z_0 < x < x_F, \quad (1.37)$$

donde c y g son funciones medibles que satisfacen $c(x) \rightarrow c > 0$, $g(x) \rightarrow 1$ cuando $x \uparrow x_F$ y $a(x)$ es una función auxiliar positiva en (z_0, x_F) , a es absolutamente continua y $a'(x)$ satisface $\lim_{x \uparrow x_F} a'(x) = 0$.

Una posible elección para la función auxiliar a en (1.37) es

$$a(x) = \int_x^{x_F} \frac{\overline{F}(t)}{\overline{F}(x)} dt, \quad x < x_F. \quad (1.38)$$

Demostración:

" \Rightarrow " Si $F \in DAM(\Lambda)$, aplicar el Teorema 1.3.28 con $g \equiv 1$.

" \Leftarrow " Si se da (1.37), entonces para cualquier $x \in \mathbb{R}$,

$$\begin{aligned} \lim_{t \rightarrow x_F} \frac{1 - F(t + xa(t))}{1 - F(t)} &= \lim_{t \rightarrow x_F} \exp \left\{ - \int_t^{t+xa(t)} \frac{g(s)}{a(s)} ds \right\} \\ &= \lim_{t \rightarrow x_F} \exp \left\{ - \int_0^x g(t + sa(t)) \frac{a(t)}{a(t + sa(t))} ds \right\}. \end{aligned}$$

Por los Lemas 1.3.23 y 1.3.24 es claro que el integrando converge a 1 uniformemente para $s \in (0, x)$, de manera que

$$\lim_{t \rightarrow x_F} \frac{1 - F(t + xa(t))}{1 - F(t)} = e^{-x},$$

lo cual, por la Proposición 1.3.25, es equivalente a decir $F \in DAM(\Lambda)$. ■

Observación 1.3.30 Para una v.a. X , la función $a(x)$ definida en (1.38) es precisamente la función media de exceso

$$a(x) = E[X - x \mid X > x], \quad x < x_F$$

que, como veremos más adelante, resultará ser una herramienta importante para ajustar estadísticamente una función a los datos de eventos extremos.

Al igual que para los DAM de las distribuciones Fréchet y Weibull la equivalencia de colas es una herramienta que puede ayudarnos a decidir si una distribución particular pertenece al dominio de atracción del máximo de Λ y a calcular sus constantes normantes. Debido a la gran variedad de colas \bar{F} que podemos encontrar en $DAM(\Lambda)$, esta herramienta es aún más importante.

Proposición 1.3.31 (*Cerradura del DAM(Λ) bajo Equivalencia de Colas*)

Sean F y G f.d. con el mismo extremo derecho $x_F = x_G$ y supóngase que $F \in \text{DAM}(\Lambda)$ con constantes normantes $c_n > 0$ y $d_n \in \mathbb{R}$; es decir,

$$\lim_{n \rightarrow \infty} F^n(c_n x + d_n) = \Lambda(x), \quad x \in \mathbb{R}. \quad (1.39)$$

Entonces

$$\lim_{n \rightarrow \infty} G^n(c_n x + d_n) = \Lambda(x + b), \quad x \in \mathbb{R}, \quad (1.40)$$

si y sólo si F y G tienen cola equivalente con

$$\lim_{x \uparrow x_F} \bar{F}(x) / \bar{G}(x) = e^b.$$

Demostración:

“ \Leftarrow ” Esta parte es complicada; puede consultarse en [22], Proposición 1.19.

“ \Rightarrow ” Supongamos que $\bar{F}(x) \sim c\bar{G}(x)$ cuando $x \uparrow x_F$ para alguna $c > 0$. La expresión (1.22) nos dice que la relación (1.39) es equivalente a

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

Para tales x , $c_n x + d_n \rightarrow x_F$; de manera que, por equivalencia de colas,

$$n\bar{G}(c_n x + d_n) \sim nc^{-1}\bar{F}(c_n x + d_n) \rightarrow c^{-1}e^{-x}, \quad x \in \mathbb{R}.$$

Así, por (1.22)

$$\lim_{n \rightarrow \infty} G^n(c_n x + d_n) = \exp\{-e^{-(x+\ln c)}\} = \Lambda(x + \ln c), \quad x \in \mathbb{R}.$$

Basta ahora hacer $\ln c = b$ para obtener (1.40). ■

Tenemos entonces que el $DAM(\Lambda)$ está formado por una gran variedad de distribuciones cuyas colas pueden ser muy distintas. Las colas pueden ser desde moderadamente pesadas (como la lognormal) a muy ligeras (exponencial y f.d. con soporte acotado por la derecha). Una consecuencia natural de la gran variedad de colas en el $DAM(\Lambda)$ es que las constantes normantes varían mucho también. El siguiente ejemplo ilustra la utilidad del resultado de la Proposición 1.3.31.

Ejemplo 1.3.32 (Distribución Gumbel) La f.d. Gumbel está dada por $\Lambda(x) = \exp(-e^{-x})$. Al igual que en los casos de la Fréchet y la Weibull, esta distribución de valores extremos pertenece a su propio dominio de atracción del máximo (DAM).

Observemos que

$$\Lambda^n(x) = [\exp(-e^{-x})]^n = \exp(-e^{-x}n) = \exp(-e^{(\log n - x)}),$$

de manera que, para toda n ,

$$P[M_n \leq x + \log n] = \Lambda^n(x + \log n) = \Lambda(x).$$

Es claro entonces que las constantes normantes son, en este caso

$$c_n = 1, \quad d_n = \log n.$$

Ejemplo 1.3.33 (Distribución Lognormal) Es claro que, si f es una función monótona creciente y $X'_i = f(X_i)$, entonces

$$M'_n = \max(X'_1, \dots, X'_n) = f(M_n).$$

Si (X_i) es una sucesión de v.a.i.i.d. que satisface

$$P[M_n \leq c_n x + d_n] \rightarrow H(x),$$

entonces se tendrá

$$P [M'_n \leq f(c_n x + d_n)] \rightarrow H(x).$$

La distribución lognormal es el resultado de aplicar una transformación monótona (el logaritmo) a la distribución normal.

Así, si tenemos (X_i) v.a.i. con distribución Normal y tomamos $f(x) = e^x$, obtenemos (X'_i) lognormales y

$$P [M'_n \leq \exp(c_n x + d_n)] \rightarrow H(x) = \exp(-e^{-x}).$$

pues en el Ejemplo 1.3.26 vimos que la distribución Normal $(0,1)$ pertenece al dominio de atracción del máximo de la distribución Gumbel. Esto se traduce en

$$P [e^{-d_n} M'_n \leq 1 + c_n x + o(c_n)] \rightarrow \exp(-e^{-x})$$

y, puesto que $c_n \rightarrow \infty$, se sigue que

$$P [c_n^{-1} e^{-d_n} (M'_n - e^{d_n}) \leq x] \rightarrow \exp(-e^{-x}),$$

de manera que la distribución lognormal está en el DAM(Λ) con constantes normantes

$$c'_n = c_n e^{-d_n}, \quad d'_n = e^{d_n},$$

donde c_n y d_n son las constantes normantes de la distribución Normal. En el siguiente Capítulo veremos que los valores explícitos de c_n y d_n son

$$\begin{aligned} c_n &= (2 \log n)^{-1/2}, \\ d_n &= (2 \log n)^{1/2} - \frac{1}{2} (2 \log n)^{-1/2} (\log \log n + \log 4\pi). \end{aligned}$$

Hemos visto hasta ahora las características de los dominios de atracción del máximo (DAM) de las tres distribuciones de valores extremos, dando algunos resultados que nos permiten verificar si una distribución determinada pertenece o no a cada uno de estos DAMs. Al mismo tiempo, se proporcionaron métodos para la determinación de las constantes normantes en cada caso. La siguiente tabla es un pequeño resumen de los resultados más importantes presentados en esta sección. En todos los casos el resultado límite es

$$c_n^{-1} (M_n - d_n) \xrightarrow{d} H,$$

con H una distribución de valores extremos.

Tipo	Fréchet $\Phi_\alpha(x)$, $\alpha > 0$	Weibull $\Psi_\alpha(x)$, $\alpha > 0$	Gumbel $\Lambda(x)$
F. de distr.	$\exp(-x^{-\alpha})$, $x > 0$	$\exp(-(-x)^\alpha)$, $x \leq 0$	$\exp(-e^{-x})$
x_F	$= \infty$	$< \infty$	$\leq \infty$
FEDAM(H)	$\lim_{x \rightarrow \infty} \frac{1-F(tx)}{1-F(x)}$ $= x^{-\alpha}$, $t > 0, \alpha > 0.$	$\lim_{h \downarrow 0} \frac{1-F(x_F-th)}{1-F(x_F-h)}$ $= x^\alpha$, $t > 0, \alpha > 0.$	$\lim_{t \uparrow x_F} \frac{1-F(t+xa(t))}{1-F(t)}$ $= e^{-x}$, $x \in \mathbb{R}$, para alguna $a(t)$ posit. medible.
c_n	$F^{-1}(1 - n^{-1})$	$x_F - F^{-1}(1 - n^{-1})$	$a(d_n)$
d_n	0	x_F	$F^{-1}(1 - n^{-1})$
Ejemplos	Fréchet, Cauchy, Pareto.	Weibull, Uniforme, Beta.	Gumbel, Normal Exp., Lognormal.

Resumen de las caracterizaciones de los DAMs de las distribuciones de valores extremos.

1.4 La Distribución Pareto Generalizada (DPG) para Aproximar Excesos sobre un Umbral

En esta Sección daremos una breve introducción a la Distribución Pareto Generalizada (DPG), la cual resultará ser un modelo natural para los excesos por encima de un umbral alto u . Además, veremos que la DPG está íntimamente relacionada con las condiciones del Dominio de Atracción del Máximo de la distribución de Valores Extremos Generalizada ($DAM(H_\xi)$) que analizamos en la Sección anterior.

Precisemos, antes que nada, la noción de exceso.

Definición 1.4.1 (*Función de Distribución de Exceso*) Sea X una v.a. con f.d. F y extremo derecho x_F . Para u fijo, $u < x_F$, decimos que

$$F_u(x) = P\{X - u \leq x \mid X > u\}, \quad x \geq 0 \quad (1.41)$$

es la función de distribución de exceso sobre el umbral u de la v.a. X .

Definición 1.4.2 (*Función Media de Exceso*) La función media de exceso de una v.a. X con f.d. F está dada por

$$e(u) = E[X - u \mid X > u]. \quad (1.42)$$

Los excesos por encima de un umbral u son fundamentales en muchos campos. Por ejemplo, en el lenguaje utilizado en las estadísticas médicas F_u se conoce como la *f.d. de exceso de vida o de vida residual*; en el contexto de seguros, se utiliza el término *f.d. de exceso de pérdida* para referirse a F_u .

El cálculo de la función media de exceso $e(u)$ puede realizarse utilizando la definición de $e(u)$ e integración por partes. Supongamos que X es una v.a. positiva con f.d. F y esperanza finita. Entonces

$$\begin{aligned} e(u) &= \int_u^{x_F} (x-u) dF(x) / \bar{F}(u) \\ &= \frac{1}{\bar{F}(u)} \int_u^{x_F} \bar{F}(x) dx, \quad 0 < u < x_F. \end{aligned} \quad (1.43)$$

Si F es continua,

$$\bar{F}(x) = \frac{e(0)}{e(x)} \exp \left\{ - \int_0^x \frac{1}{e(u)} du \right\}, \quad x > 0. \quad (1.44)$$

De la expresión (1.44) se sigue que una f.d. continua queda determinada de manera única por su función media de exceso. Si, además, $\bar{F} \in \mathcal{R}_{-\alpha}$ para algún $\alpha > 1$, entonces aplicando el Teorema de Karamata (ver Apéndice B) se obtiene que $e(u) \sim u/(\alpha - 1)$ cuando $u \rightarrow \infty$.

Presentaremos ahora una nueva familia de f.d. la cual, como veremos más adelante, está muy relacionada con los excesos por encima de un umbral alto.

Definición 1.4.3 (Distribución Pareto Generalizada (DPG)) La f.d. G_ξ definida como

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

donde

$$x \geq 0 \quad \text{si } \xi \geq 0$$

$$0 \leq x \leq -1/\xi \quad \text{si } \xi < 0$$

se conoce como la Distribución Pareto Generalizada (DPG). Si se reemplaza el ar-

gumento x por $(x - \nu)/\beta$ para $\nu \in \mathbb{R}$ y $\beta > 0$ se obtiene -haciendo los cambios de

soporte pertinentes- la familia $G_{\xi,\nu,\beta}$ con parámetros de localización y escala; también nos referimos a $G_{\xi,\nu,\beta}$ como DPG.

Igual que en el caso de la distribución VEG H_0 , podemos interpretar a G_0 como el límite de G_ξ cuando $\xi \rightarrow 0$.

En el Capítulo 3 veremos que la f.d. $G_{\xi,0,\beta}$ juega una papel importante en la práctica, puesto que el ajuste de DPGs es uno de los conceptos más importantes en la estadística de eventos extremos. Abusando un poco de la notación la denotaremos

$$G_{\xi,\beta}(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}, \quad x \in D(\xi, \beta),$$

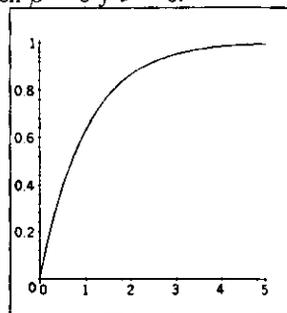
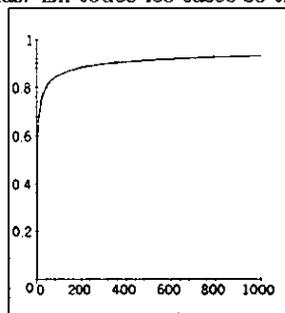
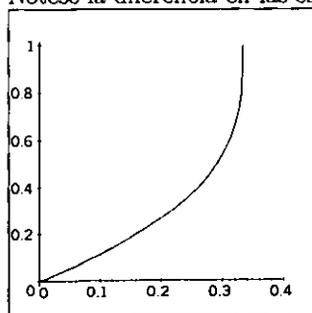
donde

$$x \in D(\xi, \beta) = \begin{cases} [0, \infty) & \text{si } \xi \geq 0 \\ [0, -\beta/\xi] & \text{si } \xi < 0. \end{cases}$$

Así, cuando digamos que X tiene una DPG con parámetros ξ y β , deberá entenderse que X tiene f.d. $G_{\xi,\beta}$.

A continuación se presentan gráficas de la DPG para tres valores distintos de ξ .

Nótese la diferencia en las escalas. En todos los casos se tienen $\beta = 0$ y $\nu = 0$.



DPG con $\xi = -3$ ($G_{-3,0,0}$). DPG con $\xi = 3$ ($G_{3,0,0}$). DPG con $\xi = 0$ ($G_{0,0,0}$).

Los siguientes ejemplos nos darán una idea intuitiva de la importancia de la DPG en el contexto de los excesos sobre un umbral alto u .

Ejemplo 1.4.4 (Distribución Exponencial) Tenemos $\bar{F}(x) = e^{-\lambda x}$, $x, \lambda > 0$.

Consideremos la siguiente expresión:

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \frac{e^{-\lambda(u+xa(u))}}{e^{-\lambda u}} = e^{-\lambda xa(u)}.$$

Si ahora tomamos $a(u) = \lambda^{-1}$, entonces se obtiene

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = e^{-x}.$$

Obsérvese que este es precisamente el valor de la cola de la DPG en el caso en que $\xi = 0$.

Ejemplo 1.4.5 (Distribución Uniforme (0,1)) Se tiene $\bar{F}(x) = 1-x$, $0 < x < 1$.

Entonces

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \frac{1 - (u + xa(u))}{1 - u} = 1 - \frac{a(u)}{1 - u}x.$$

Sea $a(u) = 1 - u$; así

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = 1 - x.$$

Este resultado corresponde a la cola de una DPG con $\xi = -1$.

Ejemplo 1.4.6 (Distribución Pareto) En este caso, $\bar{F}(x) = Kx^{-\alpha}$, $K, \alpha > 0$,

$x \geq K^{1/\alpha}$. Evaluamos

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \frac{K(u + xa(u))^{-\alpha}}{Ku^{-\alpha}} = \left(\frac{u + xa(u)}{u}\right)^{-\alpha} = \left(1 + x\frac{a(u)}{u}\right)^{-\alpha}.$$

Si hacemos $a(u) = u/\alpha$ y $\alpha = 1/\xi$, se obtiene

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \left(1 + \frac{x}{\alpha}\right)^{-\alpha} = (1 + \xi x)^{-1/\xi}.$$

Nótese que esto es la cola de una DPG donde $\xi > 0$.

Los resultados obtenidos en estos ejemplos no son casualidad. Habíamos visto antes (ver Ejemplo 1.2.10) que la f.d. exponencial está en el $DAM(\Lambda)$, lo cual, hablando en términos de la distribución VEG H_ξ , puede expresarse diciendo que la exponencial está en el $DAM(H_0)$. También vimos (Ejemplo 1.2.12) que la f.d. Pareto está en el $DAM(\Phi_\alpha)$ para $\alpha > 0$, lo cual puede reformularse diciendo que la f.d. Pareto está en el $DAM(H_\xi)$ para $\xi > 0$. Por último, de acuerdo con el Ejemplo 1.2.11, la f.d. Uniforme (0,1) está en el $DAM(\Psi_1)$, es decir, en el $DAM(H_{-1})$.

El siguiente teorema, que es uno de los resultados básicos en la TVE, nos confirma la relación existente entre el $DAM(H_\xi)$ y la DPG.

Teorema 1.4.7 (Gnedenko-Balkema-Pickands-de Haan) Sea $\xi \in \mathbb{R}$. Son equivalentes:

(a) $F \in DAM(H_\xi)$.

(b) Existe una función positiva y medible $a(\cdot)$ tal que para $1 + \xi x > 0$ se tiene que

$$\lim_{u \uparrow x_F} \frac{\overline{F}(u + xa(u))}{\overline{F}(u)} = \overline{G}_\xi(x) = \begin{cases} (1 + \xi x)^{-1/\xi} & \text{si } \xi \neq 0 \\ e^{-x} & \text{si } \xi = 0. \end{cases} \quad (1.45)$$

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

$$\lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1} & \text{si } \xi \neq 0 \\ \frac{\ln x}{\ln y} & \text{si } \xi = 0 \end{cases}, \quad (1.46)$$

donde $U(t) = F^{-1}(1 - t^{-1})$, $t > 0$ para F^{-1} la función cuantil de una f.d. F .

Demostración:

(a) \iff (b) Para $\xi = 0$, esto es equivalente a lo dicho en el Corolario 1.3.29.

Para $\xi > 0$ se tiene que $H_\xi(x) = \Phi_\alpha(a^{-1}(x+a))$ para $\alpha = 1/\xi$. De acuerdo al Teorema 1.3.5, vemos que (a) es equivalente a tener $\bar{F} \in \mathcal{R}_{-\alpha}$. Por el Teorema de Representación para Funciones de Variación Regular (ver Apéndice B) sabemos que, para alguna $z > 0$,

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

donde $c(x) \rightarrow c > 0$ y $(a(x)/x) \rightarrow \alpha^{-1}$ cuando $x \rightarrow \infty$ localmente uniformemente.

De donde

$$\lim_{u \rightarrow \infty} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \left(1 + \frac{x}{\alpha}\right)^{-\alpha} = (1 + \xi x)^{-1/\xi},$$

que es la expresión (1.45). Si se da (b), basta tomar $d_n = (1/\bar{F})^{-1}(n) = U(n)$, de manera que $1/\bar{F}(d_n) \sim n$. Haciendo $u = d_n$ en (1.45) se obtiene

$$\left(1 + \frac{x}{\alpha}\right)^{-\alpha} = \lim_{n \rightarrow \infty} \frac{\bar{F}(d_n + xa(d_n))}{\bar{F}(d_n)} = \lim_{n \rightarrow \infty} n \bar{F}(d_n + xa(d_n)),$$

lo que, por la Proposición 1.3.2, implica que $F \in DAM(H_\xi)$ para $\xi = \alpha^{-1}$.

(b) \iff (c) Restringimos nuestra atención al caso $\xi \neq 0$, siendo la demostración análoga para el caso $\xi = 0$. Por simplicidad, supondremos que F es continua y creciente en $(-\infty, x_F)$. Sea $s = 1/\bar{F}(u)$; entonces podemos reescribir (1.45) como

$$A_s(x) = (s\bar{F}(U(s) + xa(U(s))))^{-1} \rightarrow (1 + \xi x)^{1/\xi}, \quad s \rightarrow \infty.$$

Ahora, para cada $s > 0$, $A_s(x)$ es decreciente y, cuando $s \rightarrow \infty$, $A_s(x)$ converge a una función continua. Esto significa que también $A_s^-(t)$ converge puntualmente a la inversa de la función límite correspondiente, es decir,

$$\lim_{s \rightarrow \infty} \frac{U(st) - U(s)}{a(U(s))} = \frac{t^\xi - 1}{\xi}. \quad (1.47)$$

Para obtener (1.46) basta ahora usar la relación (1.47) para $t = x$ y $t = y$ y tomar el cociente. El recíproco puede probarse de manera similar. ■

Observación 1.4.8 *La condición expresada en (1.45) tiene una interpretación probabilística muy útil e interesante: sea X una v.a. con f.d. $F \in DAM(H_\xi)$, entonces (1.45) puede reformularse de la siguiente manera*

$$\lim_{u \uparrow z_F} P \left[\frac{X - u}{a(u)} > x \mid X > u \right] = \begin{cases} (1 + \xi x)^{-1/\xi} & \text{si } \xi \neq 0 \\ e^{-x} & \text{si } \xi = 0. \end{cases} \quad (1.48)$$

De modo que (1.48) da una aproximación (en escala) a la distribución de los excesos sobre un umbral alto u , donde el factor de escala es $a(u)$. Esta interpretación es crucial para muchas de las aplicaciones de la TVE que veremos en los siguientes capítulos. Esta importante relación puede reescribirse de la siguiente manera:

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

Observación 1.4.9 *En el Capítulo 3, dentro de los Modelos de Máximos de Bloque, veremos cómo al reformular ligeramente la relación dada en (1.46) se obtiene un método para estimar cuantiles fuera del rango de los datos.*

El Teorema anterior nos ha hecho ver la importancia de la DPG. El siguiente Teorema menciona algunas de sus propiedades básicas, las cuales serán ampliamente utilizadas en el Capítulo 3.

Teorema 1.4.10 (Propiedades de la DPG) *Supongamos que X tiene una DPG con parámetros ξ y β . Entonces:*

(a) Para cada $\xi \in \mathbb{R}$, $F \in DAM(H_\xi)$ si y sólo si

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

para alguna función positiva β .

(b) Supongamos que $x_i \in D(\xi, \beta)$, $i = 1, 2$; entonces

$$\frac{\overline{G}_{\xi, \beta}(x_1 + x_2)}{\overline{G}_{\xi, \beta}(x_1)} = \overline{G}_{\xi, \beta + \xi x_1}(x_2). \quad (1.49)$$

(c) Supongamos que X tiene DPG con parámetros $\xi < 1$ y β . Entonces, para

$u < x_F$,

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

Demostración:

(a) En el Teorema 1.4.7 se probó que $F \in DAM(H_\xi)$ si y sólo si

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

donde $\beta(u) = a(u)$. Puesto que la DPG es continua, se sigue el que la convergencia sea uniforme (ver Apéndice A).

(b) Basta hacer el cálculo directamente:

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

(c) Se sigue inmediatamente de la representación (1.43). ■

Observación 1.4.11 *La propiedad (b) arriba mencionada puede reformularse como sigue: la clase de las DPGs es cerrada con respecto a cambios en el umbral. Esto puede verse de la siguiente manera: el lado izquierdo de la expresión (1.49) es la probabilidad condicional de que, dado que nuestra v.a. subyacente excede x_1 , también exceda el umbral $x_1 + x_2$. El lado derecho de (1.49) nos dice que esta probabilidad es nuevamente de tipo Pareto Generalizada. Esta propiedad de cerradura es muy importante en reaseguro, donde las DPGs se utilizan en los contratos de exceso de pérdida.*

Observación 1.4.12 *La propiedad (a) nos sugiere a la DPG como una aproximación adecuada de la f.d. de exceso F_u para u grande. Este resultado puede reformularse de la siguiente manera: para alguna función β que deberá estimarse de los datos,*

$$\bar{F}_u(x) = P[X - u > x \mid X > u] \approx \bar{G}_{\xi, \beta(u)}(x), \quad x > 0.$$

De forma alternativa, puede considerarse, para $x > u$,

$$P[X > x \mid X > u] \approx \bar{G}_{\xi, u, \beta(u)}(x).$$

En ambos casos u debe tomarse suficientemente grande.

Observación 1.4.13 *Juntas, las propiedades (a) y (c) nos proporcionan un buen método gráfico para elegir el umbral u suficientemente alto como para justificar una aproximación de la f.d. de exceso F_u mediante una DPG: dada una muestra i.i.d. X_1, \dots, X_n , construimos la función media de exceso empírica $e_n(u)$ como una versión muestral de la función media de exceso $e(u)$. Por (c) sabemos que la función media de exceso de una DPG es lineal, de manera que buscamos una región de valores de u*

donde la gráfica de $e_n(u)$ sea aproximadamente lineal. Para una u en esa región parece razonable aproximar F_u mediante una DPG. Este enfoque se usará en el Capítulo 3 para ajustar excesos por encima de umbrales altos.

Hemos visto que la DPG es de cola pesada cuando $\xi > 0$, por lo cual éste es el caso más relevante para administración de riesgos. Mientras que la distribución normal tiene momentos de todos los órdenes, una distribución de cola pesada no tiene un conjunto completo de momentos. En el caso de la DPG con $\xi > 0$, tenemos que $E[X^k]$ es infinita para $k \geq 1/\xi$. Cuando $\xi = 1/2$, la DPG es una distribución con segundo momento (varianza) infinito; cuando $\xi = 1/4$, la DPG tiene cuarto momento infinito. Algunos tipos de datos de grandes reclamaciones de seguros sugieren un segundo momento infinito; de la misma manera, ciertos rendimientos de mercado indican una distribución con cuarto momento infinito. La distribución normal no puede modelar estos fenómenos, pero la DPG se usa para capturar este tipo de conducta precisamente.

En este Capítulo vimos que la distribución VEG H_ξ , $\xi \in \mathbb{R}$, describe las distribuciones límite para máximos apropiadamente normalizados. Por su parte, la DPG $G_{\xi,\beta}$, $\xi \in \mathbb{R}$, $\beta > 0$ se obtiene como la distribución límite de los excesos (bajo escala) por encima de un umbral. Se mencionaron también algunas de las propiedades probabilísticas de ambos tipos de distribuciones generalizadas; estas propiedades serán de gran importancia para el análisis estadístico de eventos extremos, tal como se verá en el Capítulo 3. Así, usaremos la distribución VEG para el análisis de datos que se nos presentan como máximos i.i.d. de una serie de tiempo, como por ejemplo máximos

anuales de niveles de ríos, velocidad del viento, etc. El Teorema 1.4.7 nos da las bases para poder estimar la cola y altos cuantiles para este tipo de datos. La parte (b) de ese mismo Teorema nos guía directamente a la definición de la DPG, en la cual se basa el método llamado Picos Sobre el Umbral (POT, *Peaks Over Thresholds*) para aproximar la f.d. de exceso. El método POT nos permite asimismo estimar dos importantes medidas de riesgo: el Valor en Riesgo (VaR) y la Esperanza Condicional de la Cola (ES).

Capítulo 2

Medidas de Riesgo

Antes de considerar las medidas de riesgo, debemos *definir* matemáticamente lo que es un riesgo. Desde esta perspectiva, diremos que *un riesgo es una variable aleatoria que mapea estados futuros del mundo a valores que representan pérdidas o ganancias*.

Los objetos básicos de nuestro estudio serán entonces las variables aleatorias del conjunto de estados de la naturaleza (o del mundo) en una fecha futura, interpretados como posibles valores futuros de las posiciones o portafolios que se tienen actualmente.

Estos riesgos pueden considerarse ya sea de manera individual o como parte de un proceso estocástico en el que los riesgos actuales dependen de los riesgos previos.

En este estudio, las variables aleatorias que consideremos (los riesgos) representarán pérdidas. Así, al pensar en la parte derecha del soporte de una distribución, estaremos pensando en pérdidas muy grandes, mientras que al pensar en la parte izquierda del soporte de la distribución hablaremos de pérdidas pequeñas (que, en su caso, pueden ser ganancias).

Los valores potenciales de un riesgo tienen una distribución F que nunca podremos

observar exactamente, si bien es posible que las pérdidas ocurridas anteriormente y debidas a riesgos similares, en caso de existir, nos den información parcial acerca de la distribución.

Una primera aproximación a la noción de medida de riesgo es pensar en una medida de riesgo como un número que “resume” la distribución de un riesgo. Esto proporciona una especie de idea intuitiva sobre el concepto, si bien resulta un tanto ambigua. Para precisar un poco esta noción, diremos que una medida de riesgo ρ es una función de valores reales sobre el espacio de variables aleatorias de valores reales.

Si es positivo, el número $\rho(X)$ asignado por la medida ρ al riesgo X puede interpretarse como el capital necesario (requerido por un regulador, por ejemplo) para mantener la posición. Si $\rho(X)$ es negativa, la cantidad $-\rho(X)$ puede retirarse de la posición.

Si bien es cierto que describir un riesgo mediante un solo número puede conllevar una pérdida enorme de información, la decisión básica sobre tomar un riesgo (permitir a alguien tomar un riesgo, respectivamente) es fundamentalmente del tipo “sí o no” y, de hecho, este es el origen de las medidas de riesgo. Si la medida de riesgo nos indica que nos encontramos ante un riesgo inaceptable (es decir, una posición con valor futuro inaceptable), tenemos como remedio el alterar la posición.

Es claro que debe existir un balance entre el rigor de la medida de riesgo y el nivel de actividades en el área supervisada. En efecto, existe una gran variedad de medidas de riesgo y la elección de una medida de riesgo específica debe tomar en cuenta qué tan bien se adaptan sus características a los requerimientos del área en que ésta será aplicada. Por ejemplo, es posible que un regulador que debe tomar en

cuenta los estados desfavorables cuando permite que se tome una posición riesgosa cuya garantía de último recurso está dada por el gobierno utilice una medida de riesgo distinta a la utilizada por la compañía liquidadora de una Bolsa, que debe cumplir las promesas hechas a cada una de las partes, sobre completar todas las transacciones de manera segura.

La organización del resto de este Capítulo se da a continuación. En la Sección 2.1 se mencionan las propiedades deseables de una medida de riesgo y se justifica el que se pida cada una de ellas para las medidas de riesgo coherentes; asimismo, se presentan las definiciones de tres medidas de riesgo muy comunes: la varianza σ^2 , el Valor en Riesgo (VaR, por sus siglas en inglés: *Value at Risk*) y la Esperanza Condicional de la Cola (ES, *Expected Shortfall*), también llamada Déficit Esperado. En la Sección 2.2 se describen las propiedades y características de la varianza; en esta Sección se muestra, a través de ejemplos, que la varianza en general *no* es una medida de riesgo coherente. La Sección 2.3 está dedicada al Valor en Riesgo (VaR). En ella se muestra la utilidad del VaR como medida de riesgo, pero también se subraya que el VaR no es subaditivo (y por tanto no es una medida de riesgo coherente), lo cual se hace evidente a través de diversos ejemplos. En la Sección 2.4 se presenta la Esperanza Condicional de la Cola (ES) y se demuestra que esta medida de riesgo sí es coherente, por lo que, en este sentido, es una medida de riesgo mucho más confiable que la varianza o el VaR. Finalmente, en la Sección 2.5 se describen las distribuciones esféricas y elípticas, que son generalizaciones de la distribución Gaussiana multivariada. Allí se explica que la utilización del VaR y de la varianza como medidas de riesgo fue planeada para llevarse a cabo con distribuciones Gaussianas multivariadas, que es un escenario donde tanto

la varianza como el VaR resultan coherentes. De hecho, en esta Sección se demuestra que dentro del mundo elíptico el VaR sí es subaditivo, y que, cuando se trabaja con distribuciones elípticas, la varianza, el VaR y la ES pueden utilizarse indistintamente para elegir el portafolio de Markowitz que minimiza el riesgo.

Los ejemplos presentados en este capítulo están tomados de [4], [2], [9] y [10]. Otras referencias son [3], [8], [7], [6] y [21].

A continuación se describen las propiedades que debería satisfacer una medida de riesgo para no incurrir en contradicciones en su utilización.

2.1 Propiedades Deseables de Medidas de Riesgo

Sea ρ una medida de riesgo, es decir, una función de valores reales sobre el espacio de variables aleatorias de valores reales. Pediremos que ρ satisfaga las siguientes propiedades:

1. **Monotonía:** Para cualesquiera dos v.a. con $X \geq Y$, se tiene que $\rho(X) \geq \rho(Y)$.

Esto significa que, si hay un riesgo mayor que otro en todos los sentidos, la medida de riesgo debe indicárnoslo.

2. **Subaditividad:** Para cualesquiera dos v.a. X y Y , se tiene que

$$\rho(X + Y) \leq \rho(X) + \rho(Y). \quad (2.1)$$

Esta propiedad puede expresarse mediante la frase "Una fusión no crea riesgo adicional". Pedir esta propiedad a una medida de riesgo resulta natural, pues de lo contrario:

- Si un individuo deseara tomar el riesgo $X + Y$ en una Bolsa, y la medida de riesgo utilizada para fijar los márgenes en la Bolsa no fuera subaditiva, este individuo podría decidir abrir dos cuentas por separado, una para el riesgo X y otra para el riesgo Y , para que su margen requerido fuera menor, a saber $\rho(X) + \rho(Y)$.
- Si un regulador impone a una compañía un requerimiento de capital extra que no satisface la propiedad de subaditividad, la compañía podría decidir dividirse en dos afiliadas separadas.
- Supongamos que dos gerencias de una compañía calculan, de manera descentralizada, las medidas $\rho(X)$ y $\rho(Y)$ de los riesgos que han tomado. Si la medida ρ que utilizan es subaditiva, entonces el supervisor de estas gerencias puede estar seguro de que $\rho(X) + \rho(Y)$ es una garantía aceptable relativa al riesgo global $X + Y$. Así, si dispone de una cantidad m para sus negocios conjuntos, puede imponer límites m_1 y m_2 , con $m = m_1 + m_2$ a sus gerencias, y así descentralizar su restricción de capital.

3. **Homogeneidad Positiva:** Para $\lambda \geq 0$ se tiene que

$$\rho(\lambda X) = \lambda \rho(X). \quad (2.2)$$

Significa que el tamaño de la posición influencia directamente el riesgo.

Obsérvese que la condición de subaditividad implica que $\rho(nX) \leq n\rho(X)$ para $n = 1, 2, \dots$. En la condición de Homogeneidad positiva, lo que se requiere es la desigualdad contraria -y, de hecho, la igualdad para cualquier λ positiva- para

modelar lo que podría ocurrir en un caso en que no se dé diversificación de riesgos.

4. **Invariante bajo translaciones:** Para cualquier v.a. X y $a \in \mathbb{R}$ se tiene que

$$\rho(X + a) = \rho(X) + a. \quad (2.3)$$

Significa que sumar (restar) la cantidad inicial asegurada a a la posición inicial simplemente aumenta (disminuye) la medida de riesgo en a .

Definición 2.1.1 (Medidas de Riesgo Coherentes) *Una medida de riesgo coherente, de acuerdo con la definición de Artzner et al. [4], es una función ρ de valores reales sobre el espacio de variables aleatorias de valores reales (donde los valores positivos de esta v.a. representan pérdidas) que cumple con las cuatro propiedades anteriores.*

Hemos mencionado ya algunas medidas de riesgo muy conocidas, a saber la varianza, el VaR y la ES. A continuación se presentan sus definiciones:

Definición 2.1.2 (Varianza) *La varianza σ_X^2 de una variable aleatoria X es una medida de la dispersión de la distribución de X alrededor de la media μ_X , donde $Var(X) = \sigma_X^2 = E[(X - \mu_X)^2]$.*

En el caso de dos variables aleatorias X y Y , es necesario definir la covarianza de X y Y , dada por

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X]E[Y]. \quad (2.4)$$

La varianza se utiliza como medida de riesgo en el enfoque de Markowitz. Esto se verá con más detalle en la Sección 2.2.

Definición 2.1.3 (VaR) *El Valor en Riesgo (VaR) es un alto cuantil de la distribución de pérdidas F . Matemáticamente, para $q \in (0.5, 1)$,*

$$\text{VaR}_q = F^{-1}(q) = \inf \{x \in \mathbb{R} : F(x) \geq q\}$$

Es una especie de cota superior para una pérdida que se excede sólo en pocas ocasiones. Así, si tenemos que el $\text{VaR}_{0.95} = 3$ millones, la información que se obtiene es que el 95% de las veces, la pérdida en que se incurra será menor que 3 millones.

Definición 2.1.4 (ES) *La Esperanza Condicional de la Cola (ES) o Déficit Esperado es el tamaño esperado de una pérdida, dado que sabemos que el VaR ha sido excedido. Matemáticamente, para $q \in (0.5, 1)$,*

$$ES_q = E[X \mid X > \text{VaR}_q]$$

Analizaremos ahora algunas de las características principales de estas tres medidas de riesgo.

2.2 La Varianza

Sea \mathcal{P} un conjunto de portafolios lineales

$$\mathcal{P} = \left\{ Z = \sum_{i=1}^n \lambda_i X_i \mid \lambda_i \in \mathbb{R}, \sum_{i=1}^n \lambda_i = 1 \right\}.$$

Supongamos que tenemos un cierto rendimiento esperado, es decir, pedimos además que $E[Z] = r$. Queremos entonces determinar los pesos óptimos de cada uno de los

X_i de manera que se minimice el riesgo de un portafolio bajo la condición $E[Z] = r$. El enfoque de Markowitz hace precisamente esto, midiendo el riesgo a través de la varianza; así, Markowitz selecciona el portafolio de mínima de varianza de entre los $Z \in \mathcal{P}$ que satisfacen $E[Z] = r$. Markowitz utiliza el hecho de que

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i,j=1}^n \text{Cov}(X_i, X_j)$$

y supone que los rendimientos X_i siguen una distribución Normal multivariada.

Así, la noción de varianza es un concepto central en la teoría financiera. Tanto el Modelo de Valuación del Activo Fijo (CAPM; *Capital Asset Pricing Model*) como la Teoría de Precios por Arbitraje (APT; *Arbitrage Pricing Theory*) utilizan la varianza como medida de riesgo para los diferentes instrumentos financieros y emplean una bonita teoría, que se basa esencialmente en la suposición de que los rendimientos siguen una distribución normal multivariada, para así llegar a la selección del portafolio óptimo.

La popularidad de la varianza puede explicarse de varias maneras. La varianza a menudo puede calcularse directamente: para muchas distribuciones bivariadas es muy fácil calcular los segundos momentos (varianzas y covarianzas). Además, la covarianza pueden manipularse fácilmente bajo transformaciones lineales: dadas $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $x \mapsto Ax + a$ y $B : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $x \mapsto Bx + b$, para $A, B \in \mathbb{R}^{m \times n}$, $a, b \in \mathbb{R}^m$ se tiene que $\text{Cov}[AX + a, BY + b] = ACov[X, Y]B^t$, siendo un caso especial de este hecho la relación $\text{Var}[\alpha^t X] = \alpha^t \text{Cov}[X] \alpha$, para $\alpha \in \mathbb{R}^n$. Así, la varianza de cualquier combinación lineal está completamente determinada por las covarianzas dos-a-dos entre los componentes. Este hecho se utiliza mucho en teoría

de portafolios, sobre todo en el enfoque de Markowitz (media-varianza).

Otra razón para la popularidad de la varianza es su facilidad de uso y la cantidad de información que nos brinda cuando se trata de las distribuciones normales multivariadas, en las que tener covarianza igual a cero es equivalente a tener independencia.

Sin embargo, la varianza y covarianza tienen también sus desventajas, por lo que su uso puede acarrear serios problemas, si se manejan sin los supuestos adecuados. Los siguientes ejemplos muestran las dificultades a las que podemos enfrentarnos.

2.2.1 Ejemplos de Problemas al Usar la Varianza

Recordemos que dentro del mundo gaussiano, la especificación de las distribuciones marginales X y Y y de su covarianza $Cov(X, Y)$ nos proporciona información suficiente para conocer la distribución conjunta. Sin embargo, cuando no nos encontramos en este mundo (o no sabemos si podemos suponer que la distribución conjunta es gaussiana), las marginales y la covarianza *no* determinan de manera única la distribución conjunta. Los siguientes ejemplos ilustran este hecho.

Observación 2.2.1 *Si bien estos ejemplos también son útiles para notar que la correlación lineal tiene algunos inconvenientes como medida de dependencia de variables aleatorias, en este trabajo no trataremos este tema.*

Ejemplo 2.2.2 *Sean X y Y v.a. con distribución normal estándar y supongamos que la covarianza $Cov(X, Y) = \rho$. Esta información no determina de manera única la distribución conjunta de $(X, Y)^t$, ya que podemos construir distintas distribuciones conjuntas para $(X, Y)^t$ que satisfagan $Cov(X, Y) = \rho$ y tengan marginales $N(0, 1)$.*

Así, si $(X, Y)^t$ se distribuye normal bivariado, entonces la función de distribución F de $(X, Y)^t$ está dada por

$$F_\rho(x, y) = \int_{-\infty}^x \int_{-\infty}^y \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\left\{\frac{-(s^2 - 2\rho st + t^2)}{2(1-\rho^2)}\right\} ds dt \quad (2.5)$$

Si sabemos que $Cov(X, Y) = 0$, entonces $\rho(X, Y) = 0$, de modo que su distribución conjunta queda dada por

$$F_0(x, y) = \int_{-\infty}^x \int_{-\infty}^y \frac{1}{2\pi} \exp\left\{\frac{-(s^2 + t^2)}{2}\right\} ds dt$$

En este caso tenemos que X y Y son independientes.

Consideremos ahora la función

$$C(x, y) = \int_0^x \int_0^y h(u, v) du dv = xy + \left(\int_0^x f(u) du\right) \left(\int_0^y g(v) dv\right)$$

donde

$$f(u) = \mathbf{1}_{\{(0.3, 0.7)\}}(u) - \frac{2}{3}\mathbf{1}_{\{(0.3, 0.7)^c\}}(u)$$

$$g(v) = -\mathbf{1}_{\{(0.3, 0.7)\}}(v) + \frac{2}{3}\mathbf{1}_{\{(0.3, 0.7)^c\}}(v)$$

y

$$\mathbf{1}_A(u) = \begin{cases} 1 & \text{si } u \in A \\ 0 & \text{si } u \notin A \end{cases}$$

es la función indicadora del conjunto A .

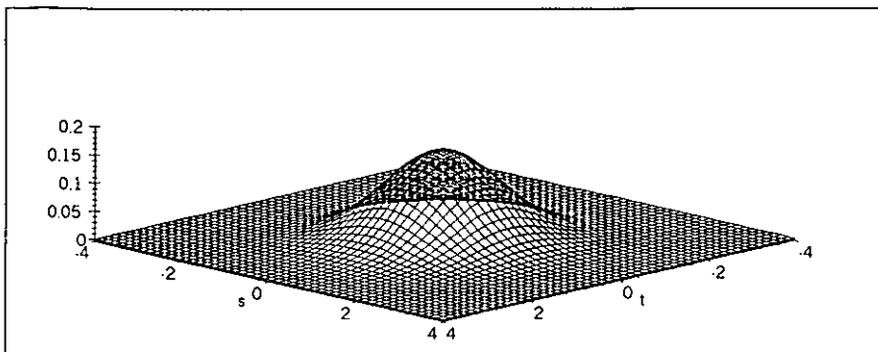
Es fácil observar que $h(x, y) = 1 + f(x)g(y)$ se anula en el cuadrado $[0.3, 0.7]^2$:

si $x \in [0.3, 0.7]$, $y \in [0.3, 0.7]$, se tiene que $h(x, y) = 1 + (1)(-1) = 0$. Esto

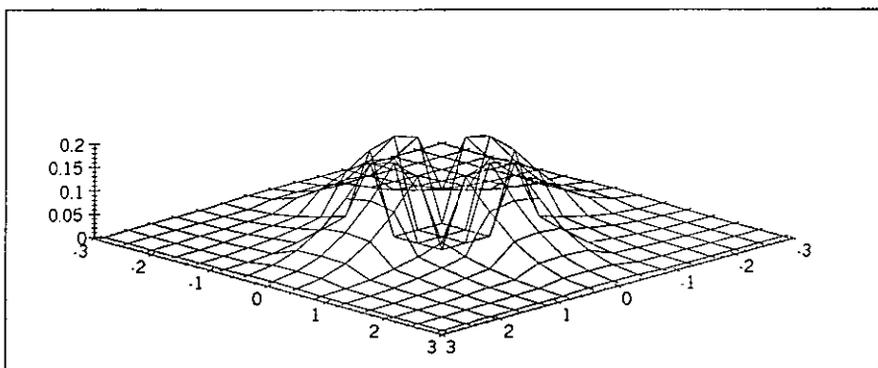
no garantiza que al tomar $F(x, y) = C(\Phi(x), \Phi(y))$ ¹ ésta no será nunca normal

¹ $C(x, y)$ se llama una cópula y es la función de distribución de un vector aleatorio en \mathbb{R}^n con marginales $U(0, 1)$. Para obtener más información sobre cópulas, puede consultarse [20].

bivariada. Además, puesto que si $1 - u \in [0.3, 0.7] \Rightarrow -u \in [-0.7, -0.3] \Rightarrow u \in [0.3, 0.7]$, tenemos que C es simétrica, es decir, $C(u, v) = C(1 - u, v)$, $0 \leq u, v \leq 1$. Esta simetría nos dice que $Cov(u, v) = Cov(1 - u, v)$, $0 \leq u, v \leq 1$, lo que implica que la covarianza es cero.



Densidad de Distrib. Normal Bivariada, con marginales $N(0, 1)$ y covarianza 0.



Densidad de $C(x, y)$, con marginales $N(0, 1)$ pero *no* Normal bivariada.

Si queremos comparar estas dos distribuciones en base a sus marginales y sus varianzas y covarianza, tenemos que, como $X \sim N(0, 1)$ y $Y \sim N(0, 1)$, entonces $Var[X] = Var[Y] = 1$ y, como se mencionó anteriormente, $Cov[X, Y] = 0$ para ambos modelos. Así, si usáramos el enfoque de Markowitz para comparar dos portafolios con

**ESTA TESIS NO SALE
DE LA BIBLIOTECA**

estos tipos de distribuciones, obtendríamos que, puesto que la varianza es la medida de riesgo que deseamos minimizar, en este caso particular los dos modelos tendrían el mismo riesgo. Pero esto es falso: basta observar la gráfica de sus densidades para notar que hay grandes diferencias entre ellos.

Tenemos entonces dos modelos con comportamientos muy diferentes, que no pueden ser diferenciados el uno del otro en base a las distribuciones marginales y la covarianza.

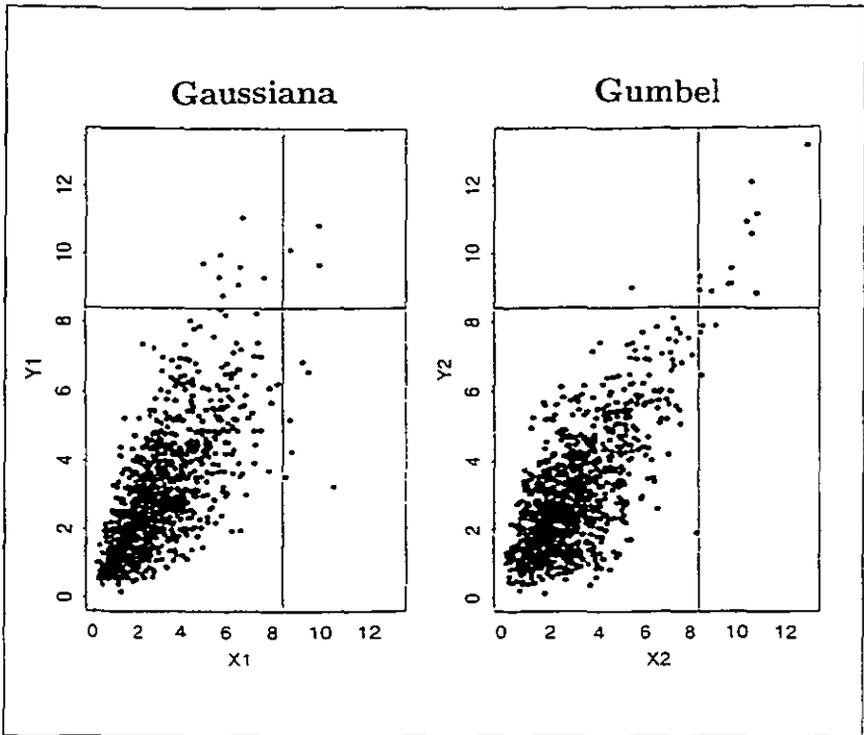
Ejemplo 2.2.3 Consideremos dos distribuciones bivariadas con marginales Gamma(3,1), que se denotarán $\Gamma_{3,1}$ y con la misma covarianza $Cov(X, Y) = 2.1$, pero con estructuras distintas, a saber

$$F_{Ga}(x, y) = \int_{-\infty}^{\phi^{-1}(\Gamma(x))} \int_{-\infty}^{\phi^{-1}(\Gamma(y))} \frac{1}{2\pi (1 - (0.71)^2)^{\frac{1}{2}}} \exp \left\{ \frac{-(s^2 - 2(0.71)st + t^2)}{2(1 - (0.71)^2)} \right\} ds dt \quad (2.6)$$

$$F_{Gu}(x, y) = \exp \left[- \left\{ (-\log \Gamma_{3,1}(x))^{1/0.54} + (-\log \Gamma_{3,1}(y))^{1/0.54} \right\}^{0.54} \right] \quad (2.7)$$

Las expresiones (2.6) y (2.7) se conocen como Cópula Gaussiana y Cópula Gumbel, respectivamente.

La siguiente figura, tomada de [9], ilustra 1000 observaciones bivariadas de estos dos modelos probabilísticos distintos para (X, Y) .



1000 observaciones bivariadas de dos distribuciones con marginales $\Gamma_{3,1}$ y con $Cov(X, Y) = 2.1$, pero con distribución conjunta distinta.

A pesar de que en ambos modelos X y Y tienen distribuciones marginales Gamma idénticas y de que la covarianza entre ellas es de 2.1 en ambos casos, es evidente que la relación entre X y Y en los modelos (2.6) y (2.7) es muy diferente desde el punto de vista cualitativo y, si pensamos que las variables aleatorias representan pérdidas, el segundo modelo es mucho más peligroso, puesto que las pérdidas extremas tienen tendencia a ocurrir juntas. Si fijamos $u = \Gamma_{3,1}^{-1}(0.99) = 8.405947$ y consideramos la probabilidad condicional de exceso $P[Y > u \mid X > u]$ bajo los dos modelos podemos,

mediante una estimación empírica basada en la Figura 3, notar que

$$\widehat{P}_{FGa} [Y > u | X > u] = 3/9,$$

$$\widehat{P}_{FGu} [Y > u | X > u] = 12/16$$

En el modelo (2.7) los excesos por encima del umbral u en un margen tienden a acompañarse de excesos en el otro margen, mientras que en el modelo (2.6) los excesos conjuntos en ambos márgenes son raros. Así, hay menos "diversificación" de riesgos grandes en el modelo (2.7). Sin embargo, en ambos casos se tiene que $Cov[X, Y] = 2.1$ y, por tratarse de marginales Gamma,

$$\begin{aligned} E[X^k] &= \int_0^{\infty} x^k \left(\frac{x^2 \exp(-x)}{\Gamma(3)} \right) dx = \frac{1}{\Gamma(3)} \int_0^{\infty} x^{k+2} \exp(-x) dx \\ &= \frac{\Gamma(k+3)}{\Gamma(3)} = (3)(4) \dots (3+k-1), \end{aligned}$$

de donde $Var(X) = Var(Y) = 12 - 9 = 3$; así, se tiene que estas dos distribuciones bivariadas no podrían diferenciarse en términos de las distribuciones marginales y de la covarianza.

Ejemplo 2.2.4 Supongamos que $(X, Y)^t$ se distribuye normal bivariado con marginales normal estándar y covarianza ρ ; denotemos la función de distribución bivariada F_ρ . Cualquier combinación convexa $F = \lambda F_{\rho_1} + (1 - \lambda) F_{\rho_2}$, $0 \leq \lambda \leq 1$, de distribuciones normales bivariadas F_{ρ_1} y F_{ρ_2} tiene también marginales normal estándar y covarianza $\rho = \lambda \rho_1 + (1 - \lambda) \rho_2$. Pero la suma $X + Y$ tiene cola más larga bajo F que bajo F_ρ :

$$P_F [X + Y > z] = \lambda (1 - \Phi) \left(\frac{z}{2(1 + \rho_1)} \right) + (1 - \lambda) (1 - \Phi) \left(\frac{z}{2(1 + \rho_2)} \right)$$

mientras que

$$P_{F_\rho} [X + Y > z] = (1 - \Phi) \left(\frac{z}{2(1 + \rho)} \right)$$

Al aplicar la regla de l'Hôpital a

$$\frac{1 - \Phi(x)}{x^{-1}\phi(x)}$$

se obtiene

$$\lim_{x \rightarrow \infty} \frac{1 - \Phi(x)}{x^{-1}\phi(x)} = \lim_{x \rightarrow \infty} \frac{-\phi(x)}{-\phi(x) \left(1 + \frac{1}{x^2}\right)},$$

que nos da la razón de Mill

$$1 - \Phi(x) = \phi(x) \left(\frac{1}{x} + O\left(\frac{1}{x^2}\right) \right).$$

Usando esta razón, podemos ver que

$$\lim_{z \rightarrow \infty} \frac{P_F [X + Y > z]}{P_{F_\rho} [X + Y > z]} = \infty \quad (2.8)$$

lo que nos muestra que los altos cuantiles de la distribución F son mucho mayores que los de la distribución F_ρ . En el Ejemplo 2.3.1 veremos que las colas de estas dos distribuciones sí pueden diferenciarse utilizando el VaR como medida de riesgo.

Una alternativa a la varianza que podría utilizarse en este caso es la kurtosis. La kurtosis de una distribución es su cuarto momento centrado μ_X^4 , y nos da una indicación de qué tan puntiaguda es la distribución. Una medida de kurtosis sin dimensión está dada por

$$\gamma_2 = \frac{\mu_X^4}{\sigma_X^4} = \frac{E[(X - \mu_X)^4]}{E[(X - \mu_X)^2]^2}.$$

Para determinar la kurtosis de una distribución usando γ_2 , tomamos como referencia a la distribución gaussiana, puesto que para cualquier distribución de este tipo, se tiene

$\gamma_2 = 3$. Para cualquier otro tipo de distribución de X , decimos que esta distribución es “menos puntiaguda” (o tiene mayor kurtosis) que la distribución gaussiana si $\gamma_2 > 3$, y que la distribución es “más puntiaguda” (tiene menor kurtosis) que la distribución gaussiana si $\gamma_2 < 3$.

Ahora bien, al calcular la kurtosis de las distribuciones F y F_ρ tratadas en este ejemplo, se obtiene lo siguiente: para F_ρ , puesto que $(X, Y)^t$ se distribuye normal bivariado con marginales normal estándar y covarianza ρ , sabemos que $Z = X + Y$ tiene distribución normal con media $E[Z] = 0$ y varianza $Var(Z) = 2(1 + \rho)$. Así, su función generadora de momentos está dada por $\Psi(t) = \exp\{(1 + \rho)t^2\}$. De aquí que $E[(Z - \mu_Z)^4] = E[Z^4] = \Psi^{(IV)}(0) = 3[2(1 + \rho)]^2$, de donde

$$\gamma_2(F_\rho) = \frac{3[2(1 + \rho)]^2}{[2(1 + \rho)]^2} = 3,$$

lo que era de esperarse, por tenerse que Z tiene distribución gaussiana. Ahora bien, en el caso de F se tiene que:

$$\begin{aligned} \gamma_2(F) &= \frac{3\lambda[2(1 + \rho_1)]^2 + 3(1 - \lambda)[2(1 + \rho_2)]^2}{[2(1 + \rho)]^2} \\ &= \frac{3 \cdot 2^2 [\lambda(1 + \rho_1)^2 + (1 - \lambda)(1 + \rho_2)^2]}{2^2 [1 + \lambda\rho_1 + (1 - \lambda)\rho_2]^2} \\ &= \frac{3[\lambda(1 + \rho_1)^2 + (1 - \lambda)(1 + \rho_2)^2]}{[\lambda(1 + \rho_1) + (1 - \lambda)(1 + \rho_2)]^2} > 3, \end{aligned}$$

puesto que

$$[\lambda a + (1 - \lambda)b]^2 - [\lambda a^2 + (1 - \lambda)b^2] = [\lambda(\lambda - 1)](a - b)^2 < 0.$$

La comparación entre las kurtosis nos indica que, como era de esperarse de acuerdo con la relación (2.8), F es menos “puntiaguda” que F_ρ , de esta manera, tenemos que F está más concentrada en las colas, es decir, es de cola más pesada que F_ρ .

Así, resulta que la kurtosis puede ser una herramienta útil para comparar las colas de las distribuciones. El problema de la kurtosis es que se requiere que el cuarto momento de la distribución sea finito, y las distribuciones de cola muy pesada no cumplen, en general, con este requisito.

En esta Sección hemos proporcionado varios ejemplos que nos hacen ver que la varianza tiene serios defectos como medida de riesgo. A través de ellos podemos ver que, en general, la varianza no es una medida de riesgo coherente. A continuación se da un resumen de los problemas que pueden surgir al utilizar la varianza.

Desventajas del uso de la Varianza

Los ejemplos anteriores nos han mostrado que no podemos confiar plenamente en la variación como medida de riesgo, pues podemos obtener el mismo resultado (la misma varianza y covarianzas) para dos distribuciones cuyas colas se comportan de manera distinta, siendo una de ellas mucho más peligrosa que la otra en términos de administración de riesgos. Para resumir los problemas que se presentan al utilizar la varianza como medida de riesgo, consideremos el caso de dos v.a. de valores reales X y Y .

- Las varianzas de X y Y pueden ser infinitas. Esto no es ideal para una medida de riesgo y causa problemas al trabajar con distribuciones de cola pesada. Por ejemplo, la varianza de los dos componentes de un vector aleatorio bivariado con distribución t_ν (es decir, para $f(x) = c(1 + x^2/\nu)^{-(\nu+2)/2}$) no existe para $\nu \leq 2$.

- La especificación de las distribuciones marginales y de la matriz de covarianzas no siempre determina una distribución conjunta única; de hecho, esto sólo sucede si se trata de una distribución normal multivariada o, más generalmente, de distribuciones elípticas (ver Sección 2.5): si sabemos que $(X, Y)^t$ siguen una distribución normal bivariada y conocemos las medias, las varianzas y la covarianza, entonces sí queda determinada una única distribución conjunta. Sin embargo, si sólo conocemos las distribuciones marginales X y Y y su covarianza, entonces existe una infinidad de posibles distribuciones bivariadas para $(X, Y)^t$ (ver Ejemplo 2.2.3). Por ello, para efectos de medición de riesgos, resulta muy osado el utilizar un modelo para la distribución conjunta que se basa solamente en las distribuciones marginales y la covarianza.
- Suponiendo que se conocen las distribuciones marginales X y Y , ni siquiera en los casos en que se tiene que la covarianza entre estas dos variables aleatorias es cero ($Cov(X, Y) = 0$) se puede determinar de manera única la distribución conjunta (ver Ejemplo 2.2.2).

Esta Sección ha servido para resumir los problemas que podemos encontrar al usar la varianza como medida de riesgo; tenemos que un problema grave de la varianza es que —ya sea por sí sola o incluso conociendo las funciones de distribución marginales— en ocasiones no nos permite distinguir entre dos distribuciones conjuntas que pueden presentar un comportamiento muy distinto en los extremos (ver Ejemplo 2.2.3); esto es un serio inconveniente para efectos de Administración de Riesgos, donde los valores extremos pueden significar la diferencia entre la supervivencia y la quiebra

de una institución financiera. Esto hace que busquemos considerar otras medidas de riesgo, con el objeto de encontrar alguna(s) que tenga(n) mejores cualidades que la varianza y cuyo uso sea, por tanto, más confiable. Para ello, en la siguiente Sección examinaremos las propiedades de una medida de riesgo cuyo uso está muy extendido en el mundo financiero de nuestros días: el Valor en Riesgo (VaR).

2.3 Valor en Riesgo

El Valor en Riesgo es una de las medidas de riesgo más usadas en la práctica. De hecho, el BIS (*Bank for International Settlement*) requiere, como parte de los acuerdos de Basilea, que las instituciones financieras a las cuales regula calculen el VaR.

Recordemos que el *Valor en Riesgo (VaR)* es un alto cuantil de la distribución de pérdidas F , la cual, como se mencionó anteriormente, es desconocida.

Matemáticamente, para $q \in (0.5, 1)$, $VaR_q = F^{-1}(q)$ (esto es, es el q -ésimo cuantil de F), donde definimos la función cuantil (también conocida como el inverso generalizado de F) mediante

$$F^{-1}(t) = \inf \{x \in \mathbb{R} : F(x) \geq t\} \quad (2.9)$$

Generalmente, se emplea el VaR para $q \geq 0.9$, si bien teóricamente éste puede calcularse para cualquier $q \in (0.5, 1)$.

Pese a la gran popularidad del VaR, esta medida de riesgo tampoco está libre de problemas. Uno de los más importantes de ellos es que no es necesariamente subaditivo. Con esto queremos decir que hay casos en que un portafolio puede dividirse en

varios subportafolios de manera que la suma del VaR correspondiente a los subportafolios es menor que el VaR del portafolio total. Esto puede ocasionar problemas si el sistema de administración de riesgos de la institución se basa en el establecimiento de límites al VaR por secciones individuales. Otro serio inconveniente del VaR es que no habla del tamaño potencial de la pérdida en caso de excederse el VaR. Así, podemos tener que $\text{VaR}_{0.99} = 10$ millones, y sin embargo puede ocurrir una pérdida de 100 millones, porque la única información que se obtiene de $\text{VaR}_{0.99}$ es que el 99% de las ocasiones, la pérdida será menor que 10 millones; pero en el 1% de casos en que se rebase esta pérdida, el $\text{VaR}_{0.99}$ no nos dice qué tanto más grande que 10 millones puede ser la pérdida.

En el siguiente ejemplo podemos constatar la utilidad del VaR como medida de riesgo.

Ejemplo 2.3.1 *Retomando el Ejemplo 2.2.4, podemos notar que, conforme se avanza más en las respectivas colas de las dos distribuciones, el Valor en Riesgo para la distribución mezclada F es mayor que el de la distribución original F_ρ .*

En efecto, usando de nuevo la razón de Mill, se tiene que

$$\bar{\Phi}(x) = 1 - \Phi(x) \sim \frac{\phi(x)}{x} = \frac{1}{\sqrt{2\pi}x} \exp\left\{\frac{-x^2}{2}\right\}, \quad x \rightarrow \infty.$$

El lado derecho de esta igualdad puede tomarse como la cola \bar{F}_ρ de la f.d. F_ρ , así que por la cerradura del DAM(Λ) (ver Sección 1.3.3) bajo equivalencia de colas, la f.d. Gaussiana Φ y F_ρ tienen las mismas constantes normalizantes c_n y d_n , donde $d_n = F_\rho^-(1 - n^{-1})$. Buscamos entonces una solución a $-\ln \bar{F}_\rho(d_n) = \ln(n)$, es

decir, a $-\ln \left[\frac{1}{\sqrt{2\pi d_n}} \exp \left\{ \frac{-d_n^2}{2} \right\} \right] = \ln n$, que se reduce a

$$\frac{1}{2} d_n^2 + \ln d_n + \frac{1}{2} \ln 2\pi = \ln n. \quad (2.10)$$

Por expansión de Taylor en (2.10), se obtiene

$$d_n = (2 \ln n)^{1/2} - \frac{\ln \ln n + \ln 4\pi}{2(2 \ln n)^{1/2}} + o\left((\ln n)^{-1/2}\right)$$

Puesto que puede tomarse $a(x) = \bar{\Phi}(x) / \phi(x)$, se tiene que $a(x) \sim x^{-1}$, de modo que

$$a\left(\frac{z}{2(1+\rho)}\right) \sim \frac{2(1+\rho)}{z}.$$

Así,

$$c_n = a(d_n) \sim \left[(2 \ln n)^{1/2} \right]^{-1} [2(1+\rho)] = 2(1+\rho) \left[-2 \ln \frac{1}{n} \right]^{1/2}.$$

Basta ahora recordar que $VaR_{\alpha, F_\rho} = F_\rho^{-1}(\alpha)$, así que al hacer $\alpha = 1 - n^{-1}$, se obtiene

$n = \frac{1}{1-\alpha}$, y de aquí

$$c_n = 2(1+\rho) [-2 \ln(1-\alpha)]^{1/2},$$

de modo que, cuando $\alpha \uparrow 1$,

$$VaR_{\alpha, F_\rho}(X+Y) \sim 2(1+\rho) [-2 \ln(1-\alpha)]^{1/2}.$$

Consideremos ahora lo que ocurre con F . Para $0 < \sigma_1 < \sigma_2$ se tiene

$$\begin{aligned} \frac{1}{\sqrt{2\pi}\sigma_1} \int_{-\infty}^z e^{-x^2/2\sigma_1^2} dx &= \frac{1}{\sqrt{2\pi}\sigma_1} \int_{-\infty}^z e^{-(x\sigma_2/\sigma_1)^2/2\sigma_2^2} dx = \frac{1\sigma_1}{\sqrt{2\pi}\sigma_1\sigma_2} \int_{-\infty}^{z\sigma_2/\sigma_1} e^{-y^2/2\sigma_2^2} dy \\ &= \frac{1}{\sqrt{2\pi}\sigma_2} \int_{-\infty}^z e^{-y^2/2\sigma_2^2} dy + \frac{1}{\sqrt{2\pi}\sigma_2} \int_z^{z\sigma_2/\sigma_1} e^{-y^2/2\sigma_2^2} dy \\ &\sim \frac{1}{\sqrt{2\pi}\sigma_2} \int_{-\infty}^z e^{-y^2/2\sigma_2^2} dy \end{aligned}$$

para valores “grandes” de z . En el caso de F , tenemos $\sigma_1^2 = 2(1 + \rho_1)$, $\sigma_2^2 = 2(1 + \rho_2)$, $\rho_2 > \rho_1$, de manera que la f.d. de $X + Y$ cuya f.d. conjunta es F queda dada por:

$$\begin{aligned} \lambda F_{\sigma_1^2} + (1 - \lambda) F_{\sigma_2^2} &= \lambda \frac{1}{\sqrt{2\pi\sigma_1}} \int_{-\infty}^z e^{-x^2/2\sigma_1^2} dx + (1 - \lambda) \frac{1}{\sqrt{2\pi\sigma_2}} \int_{-\infty}^z e^{-x^2/2\sigma_2^2} dx \\ &= \frac{1}{\sqrt{2\pi\sigma_2}} \int_{-\infty}^z e^{-x^2/2\sigma_2^2} dx + \lambda \frac{1}{\sqrt{2\pi\sigma_2}} \int_z^{z\sigma_2/\sigma_1} e^{-x^2/2\sigma_2^2} dx \\ &\sim \frac{1}{\sqrt{2\pi\sigma_2}} \int_{-\infty}^z e^{-x^2/2\sigma_2^2} dx, \end{aligned}$$

ya que el segundo término es despreciable para valores grandes de z .

Así, procediendo de la misma forma que en el caso de F_ρ , se obtiene para F

$$VaR_{\alpha,F}(X + Y) \sim 2(1 + \rho_2) [-2 \ln(1 - \alpha)]^{1/2}$$

y, puesto que $\rho_2 > \rho$,

$$\lim_{\alpha \uparrow 1} \frac{VaR_{\alpha,F}(X + Y)}{VaR_{\alpha,F_\rho}(X + Y)} = \frac{1 + \rho_2}{1 + \rho} > 1.$$

Lema 2.3.2 *El VaR es una medida de riesgo monótona, homogénea positiva e invariante bajo translaciones.*

Demostración:

1) *Monotonía.* P.D.: si $X \geq Y \Rightarrow VaR_q^X \geq VaR_q^Y$.

Para cualquier $z \in \mathbb{R}$, $F_X(z) = P[X \leq z] \leq P[Y \leq z] = F_Y(z)$. De manera que

$$VaR_q^X = \inf \{x \in \mathbb{R} : F_X(x) \geq q\} \geq \inf \{y \in \mathbb{R} : F_Y(y) \geq q\} = VaR_q^Y$$

2) *Homogeneidad Positiva.* P.D.: Para $\lambda \geq 0$, $VaR_q^{\lambda X} = \lambda VaR_q^X$.

Sea $\lambda > 0$ y sea $Y = \lambda X$. Entonces $P[Y \leq y] = P[\lambda X \leq y] = P[X \leq y/\lambda] = F_X(y/\lambda)$.

Supongamos que $VaR_q^X = \inf \{x \in \mathbb{R} : F_X(x) \geq q\} = x^*$.

Entonces es claro que $VaR_q^Y = \inf \{y \in \mathbb{R} : F_X(y/\lambda) \geq q\} = \lambda x^*$, de donde $\lambda VaR_q^X = VaR_q^{\lambda X}$.

3) *Invariante bajo Translaciones*. P.D.: Para $a \in \mathbb{R}$, $VaR_q^{X+a} = VaR_q^X + a$.

Sea $a \in \mathbb{R}$ y sea $Y = X+a$. Entonces $P[Y \leq y] = P[X + a \leq y] = P[X \leq y - a] = F_X(y - a)$. Supongamos que $VaR_q^X = x^*$. Entonces, si $y = x^* + a$, se tiene que $F_X(y - a) = F_X((x^* + a) - a) = F_X(x^*)$, de donde es claro que $VaR_q^{X+a} = VaR_q^Y = \inf \{y \in \mathbb{R} : F_X(y - a) \geq q\} = x^* + a = VaR_q^X + a$. ■

2.3.1 Ejemplos de Problemas al Utilizar el VaR

Acabamos de probar que el VaR es una medida de riesgo monótona, homogénea positiva e invariante bajo translaciones. Sin embargo, el VaR no es una medida de riesgo coherente, porque en general no satisface la propiedad de subaditividad (ver (2.1)). Los siguientes ejemplos nos muestran algunos problemas ocasionados por la no-subaditividad del VaR.

Ejemplo 2.3.3 Sean X y Y variables aleatorias con distribución idéntica $F_X(x) = 1 - x^{-1/2}$, $x \geq 1$. Esta distribución es de cola muy pesada y su esperanza $E[X]$ no es finita. Considérense los riesgos $X + Y$ y $2X$.

Se tiene que $f_X(x) = \frac{x^{-3/2}}{2}$. Haciendo $Z = X + Y$, podemos calcular

$$F_Z(z) = 1 - \frac{2\sqrt{z-1}}{z}, \quad z > 2$$

Si hacemos $W = 2X$, tenemos que

$$F_W(w) = P[W \leq w] = P[2X \leq w] = P\left[X \leq \frac{w}{2}\right] = F_X\left(\frac{w}{2}\right) = 1 - \left(\frac{w}{2}\right)^{-1/2}, \quad w \geq 2.$$

Entonces

$$P[X + Y \leq z] = 1 - \frac{2\sqrt{z-1}}{z} < 1 - \frac{\sqrt{2z}}{z} = P[2X \leq z]$$

para $z > 2$. Se sigue que

$$VaR_\alpha(X + Y) > VaR_\alpha(2X) = VaR_\alpha(X) + VaR_\alpha(Y)$$

para $\alpha \in (0.5, 1)$. El VaR no es subaditivo para esta distribución, con lo que los argumentos de diversificación no funcionan: resulta mejor tomar un solo riesgo y duplicarlo que tomar dos riesgos independientes.

Ejemplo 2.3.4 Sea Ω un conjunto infinito y consideremos dos v.a.i.i.d. X_1 y X_2 con función de densidad

$$f_{X_1}(x) = f_{X_2}(x) = \begin{cases} 0.90 & \text{si } x \in [-1, 0] \\ 0.05 & \text{si } x \in [0, 2] \\ 0 & \text{en otro caso.} \end{cases}$$

Supongamos que cada una representa, tomando las pérdidas como positivas, un valor futuro neto aleatorio, es decir, un riesgo que puede resultar interesante.

Tenemos entonces que

$$F_{X_1}(x) = F_{X_2}(x) = \begin{cases} 0 & \text{si } x < -1 \\ 0.9x + 0.9 & \text{si } x \in [-1, 0] \\ 0.05x + 0.9 & \text{si } x \in [0, 2] \\ 1 & \text{si } x > 2 \end{cases}$$

De modo que, si nos interesa el valor en riesgo al 90% de X_1 o de X_2 , obtendremos

$$VaR_{0.9}(X_1) = VaR_{0.9}(X_2) = \inf \{x \in \mathbb{R} : F(x) \geq 0.9\} = 0$$

Sin embargo, usando la independencia de X_1 y X_2 es fácil notar que el valor en riesgo al 90% para $Z = X_1 + X_2$ es definitivamente mayor que 0; en efecto

$$F_{X_1+X_2}(z) = \begin{cases} 0 & \text{si } z < -2 \\ 0.405(z+2)^2 & \text{si } z \in [-2, -1] \\ 0.81 - 0.405z^2 + 0.045(1+z)^2 & \text{si } z \in [-1, 0] \\ 0.855 + 0.09z & \text{si } z \in [0, 1] \\ 0.985 + 0.005z - 0.045(2-z)^2 & \text{si } z \in [1, 2] \\ 1 - 0.00125(4-z)^2 & \text{si } z \in [2, 4] \\ 1 & \text{si } z \geq 4 \end{cases}$$

de donde

$$VaR_{0.9}(X_1 + X_2) = \inf \{x \in \mathbb{R} : F_{X_1+X_2}(x) \geq 0.9\} = 0.5$$

De aquí que el control por separado de estos riesgos no nos permitiría controlar su suma, si nos interesara el valor en riesgo al 90%.

Ejemplo 2.3.5 *Se tienen dos opciones sobre una acción con la misma fecha de ejercicio T . La primera opción, que se denotará A , con precio inicial u , paga 1000 si el valor de la acción en el tiempo T es mayor que la cantidad fija U y nada en caso contrario; la segunda opción, que se denotará B , con precio inicial l , paga 1000 si el valor de la acción al tiempo T es menor que L (con $L < U$) y nada en caso contrario.*

Así,

$$A = \begin{cases} -u & \text{si } S_T \leq U \\ -1000 + u & \text{si } S_T > U \end{cases}$$

$$B = \begin{cases} l & \text{si } S_T \geq L \\ -1000 + l & \text{si } S_T < L \end{cases}$$

Se escogen L y U de manera que $P\{S_T < L\} = P\{S_T > U\} = 0.008$ y se calculan los valores en riesgo al 99% del valor futuro neto de las posiciones de dos inversionistas que emiten, respectivamente, dos opciones A y dos opciones B . Estos son $-2 \cdot u$ y $-2 \cdot l$, respectivamente. En contraste, el número positivo $1000 - l - u$ es el valor en riesgo al 99% del valor futuro neto de la posición de un inversionista que emite $A + B$, pues

$$A + B = \begin{cases} 1000 - u - l & \text{si } S_T \leq L \\ -u - l & \text{si } L \leq S_T \leq U \\ 1000 - l - u & \text{si } U < S_T \end{cases}$$

En los ejemplos que hemos visto surgen problemas relacionados con la subaditividad al utilizar el Valor en Riesgo porque la distribución *conjunta* con que se trabaja *no* es una distribución normal o, más generalmente, elíptica (ver Sección 2.5); más adelante veremos que el VaR sí se comporta "bien" respecto a la suma (es decir, sí es subaditivo) cuando la distribución *conjunta* con que se trabaja es una distribución Gaussiana. Sin embargo, al igual que para la varianza y covarianza, no basta con saber que las distribuciones marginales son Gaussianas, ~~sino que es necesario estar~~ seguro de que la distribución conjunta es Gaussiana.

2.4 Esperanza Condicional de la Cola

Desde hace mucho tiempo, los actuarios han calculado la prima para pólizas con deducible usando el tamaño condicional de la reclamación, dado que la reclamación excede el deducible. De la misma manera, los tratados de reaseguro han involucrado la distribución condicional de una reclamación para una póliza (o de la reclamación total para un portafolio de pólizas), dado que ésta está por encima del nivel de retención del asegurador cesionista. Para responder a la pregunta de “¿qué tan malo es lo malo?”, la cual el Valor en Riesgo deja sin responder, algunas personas han identificado el deducible (o el nivel de retención) con el cuantil utilizado en la medición de riesgos financieros. Así es como surge la Esperanza Condicional de la Cola.

Recordemos que la *Esperanza Condicional de la Cola* (*ES*) es el tamaño esperado de una pérdida, dado que sabemos que el VaR ha sido excedido. Tenemos entonces que, para $q \in (0.5, 1)$,

$$ES_q(X) = E[X \mid X > VaR_q^X]$$

Además de considerar el tamaño potencial de la pérdida dado que se ha excedido el VaR, la esperanza condicional de la cola (o el déficit esperado) es una medida de riesgo subaditiva, por lo que no tiene las limitaciones y problemas que se mencionaron para la varianza y el VaR.

Utilizando la igualdad

$$E[X \mid X > VaR_q^X] = E[X - VaR_q^X + VaR_q^X \mid X > VaR_q^X]$$

y el hecho de que VaR_q^X es una constante, se obtiene la siguiente relación entre

$ES_q(X)$ y Var_q^X :

$$ES_q(X) = Var_q^X + E[X - Var_q^X \mid X > Var_q^X]. \quad (2.11)$$

Recordemos la definición de la función media de exceso $e(u)$:

Definición 2.4.1 (Func. Media de Exceso) *La función media de exceso de una v.a. X con f.d. F está dada por*

$$e(u) = E[X - u \mid X > u]. \quad (2.12)$$

Vemos entonces que el segundo término del lado derecho de la igualdad (2.11) es precisamente $e(Var_q)$. Así, tenemos la siguiente relación:

$$ES_q(X) = Var_q^X + e(Var_q^X). \quad (2.13)$$

Además, sabemos que la función media de exceso es precisamente la media de la distribución de exceso expresada como una función del umbral u , donde, como se mencionó en el capítulo anterior, la función de distribución de exceso está dada por

$$\begin{aligned} F_u(x) &= P[X - u \leq x \mid X > u] \\ &= \frac{F(x+u) - F(u)}{1 - F(u)}, \end{aligned} \quad (2.14)$$

para $0 \leq x < x_F - u$, donde x_F es el extremo derecho de F .

De manera que, de acuerdo con (1.42) y (2.14), $e(u)$ está dada por:

$$e(u) = \int_{x=0}^{x_F-u} x dF_u(x) = \int_{x=u}^{x_F} \frac{(x-u) dF(x)}{1-F(u)} = \int_{x=u}^{x_F} \frac{\bar{F}(x) dx}{\bar{F}(u)} = \frac{1}{\bar{F}(u)} \int_{x=u}^{x_F} \bar{F}(x) dx.$$

Esto puede calcularse para varias distribuciones continuas, obteniéndose la siguiente tabla:

Nombre	Cola \bar{F} o densidad f	Parámetros	$e(u)$
Pareto	$\bar{F}(x) = \left(\frac{\kappa}{\kappa+x}\right)^\alpha$	$\alpha, \kappa > 0$	$\frac{\kappa+u}{\alpha-1}, \alpha > 1$
Lognormal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$	$\mu \in \mathbb{R}, \sigma > 0$	$\frac{\sigma^2 u}{\ln u - \mu} (1 + o(1))$
Exponencial	$\bar{F}(x) = e^{-\lambda x}$	$\lambda > 0$	λ^{-1}
Weibull	$\bar{F}(x) = e^{-cx^\alpha}$	$c > 0, \alpha \geq 0$	$\frac{u^{1-\alpha}}{c\alpha} (1 + o(1))$
Gamma	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	$\alpha, \beta > 0$	$\beta^{-1} \left(1 + \frac{\alpha-1}{\beta u} + o\left(\frac{1}{u}\right)\right)$
Normal truncada	$f(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2}$	-	$u^{-1} (1 + o(1))$
DPG	$\bar{F}(x) = \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}}$	$\xi \in \mathbb{R}, \beta > 0$	$\frac{\beta + \xi u}{1 - \xi}, \beta + \xi u > 0$

Función Media de Exceso para algunas distribuciones continuas.

Observación 2.4.2 Claramente, esta distribución Weibull $F_{c,\alpha}$ con $c, \alpha > 0, x \geq 0$ es la f.d. de una v.a. positiva. Sin embargo, en el contexto de la Teoría de Valores Extremos, la distribución Weibull Ψ_α está concentrada en $(-\infty, 0)$: $\Psi_\alpha(x) = 1 - F_{1,\alpha}(-x), x < 0$.

Haciendo uso de esta tabla, es fácil calcular la ES para las distribuciones allí mencionadas.

Ejemplo 2.4.3 Para calcular la $ES_{0.95}$ de una v.a. X que tiene distribución exponencial con $\lambda = 3$, procedemos de la siguiente manera:

Sabemos que la f.d. de X es $F(x) = 1 - e^{-3x}$. Ahora bien, por (2.13), tenemos que

$$\begin{aligned} ES_{0.95}(X) &= VaR_{0.95}^X + e(VaR_{0.95}^X) \\ &= VaR_{0.95}^X + (3)^{-1}, \end{aligned}$$

de acuerdo con los resultados de la Tabla 1.

Por otra parte

$$\begin{aligned} VaR_{0.95}^X &= \inf \{x \in \mathbb{R} \mid F(x) \geq 0.95\} \\ &= \inf \{x \in \mathbb{R} \mid 1 - e^{-3x} \geq 0.95\} \\ &= \inf \{x \in \mathbb{R} \mid \ln(0.05) = -3x\} \\ &= 0.99857, \end{aligned}$$

de donde

$$ES_{0.95}(X) = 0.99857 + 1/3 = 1.33191,$$

lo que quiere decir que, si sabemos que se ha excedido el $VaR_{0.95}^X$, el tamaño potencial de la pérdida es 1.3319.

Teorema 2.4.4 *ES es una medida de riesgo coherente en el sentido de la Definición*

2.1.1.

Demostración:

1) *Monotonía*. P.D.: si $X \geq Y \Rightarrow ES_q(X) \geq ES_q(Y)$.

Supongamos que $X \geq Y$. Entonces,

$$ES_q(Y) = E[Y \mid Y > VaR_q^Y] \leq E[X \mid Y > VaR_q^Y]$$

$$\begin{aligned}
&= VaR_q^X + E \left[X - VaR_q^X \mid Y > VaR_q^Y \right] \\
&= VaR_q^X + \frac{E \left[(X - VaR_q^X) \mathbf{1}_{\{Y > VaR_q^Y\}} \right]}{P \left[Y > VaR_q^Y \right]} \\
&= VaR_q^X + \frac{E \left[(X - VaR_q^X) \mathbf{1}_{\{Y > VaR_q^Y\}} \right]}{1 - q} \\
&= VaR_q^X + \frac{E \left[(X - VaR_q^X) \mathbf{1}_{\{Y > VaR_q^Y\}} \mathbf{1}_{\{X > VaR_q^X\}} \right]}{1 - q} \\
&\quad + \frac{E \left[(X - VaR_q^X) \mathbf{1}_{\{Y > VaR_q^Y\}} \mathbf{1}_{\{X \leq VaR_q^X\}} \right]}{1 - q}.
\end{aligned}$$

El tercer término del último renglón es menor o igual que cero, de modo que

$$\begin{aligned}
ES_q(Y) &\leq VaR_q^X + \frac{E \left[(X - VaR_q^X) \mathbf{1}_{\{Y > VaR_q^Y\}} \mathbf{1}_{\{X > VaR_q^X\}} \right]}{1 - q} \\
&\leq VaR_q^X + \frac{E \left[(X - VaR_q^X) \mathbf{1}_{\{X > VaR_q^X\}} \right]}{1 - q} \\
&= VaR_q^X + E \left[(X - VaR_q^X) \mid X > VaR_q^X \right] = ES_q(X).
\end{aligned}$$

2) *Subaditividad*. P.D.: $ES_q(X + Y) \leq ES_q(X) + ES_q(Y)$.

Sabemos que

$$\begin{aligned}
ES_q(X + Y) &= E \left[X + Y \mid X + Y > VaR_q^{X+Y} \right] \\
&= E \left[X \mid X + Y > VaR_q^{X+Y} \right] + E \left[Y \mid X + Y > VaR_q^{X+Y} \right].
\end{aligned}$$

Ahora bien, por (2.11) se tiene

$$\begin{aligned}
E \left[X \mid X + Y > VaR_q^{X+Y} \right] &= VaR_q^X + E \left[(X - VaR_q^X) \mid X + Y > VaR_q^{X+Y} \right] \\
&= VaR_q^X + \frac{E \left[(X - VaR_q^X) \mathbf{1}_{\{X + Y > VaR_q^{X+Y}\}} \right]}{P \left[X + Y > VaR_q^{X+Y} \right]} \\
&= VaR_q^X + \frac{E \left[(X - VaR_q^X) \mathbf{1}_{\{X + Y > VaR_q^{X+Y}\}} \mathbf{1}_{\{X > VaR_q^X\}} \right]}{1 - q} \\
&\quad + \frac{E \left[(X - VaR_q^X) \mathbf{1}_{\{X + Y > VaR_q^{X+Y}\}} \mathbf{1}_{\{X \leq VaR_q^X\}} \right]}{1 - q}.
\end{aligned}$$

Pero si $X \leq VaR_q^X$, entonces $X - VaR_q^X \leq 0$, por lo que el tercer término del lado derecho es negativo o cero, de donde

$$\begin{aligned} E[X | X + Y > VaR_q^{X+Y}] &\leq VaR_q^X + \frac{E[(X - VaR_q^X) \mathbf{1}_{\{X+Y > VaR_q^{X+Y}\}} \mathbf{1}_{\{X > VaR_q^X\}}]}{1 - q} \\ &\leq VaR_q^X + \frac{E[(X - VaR_q^X) \mathbf{1}_{\{X > VaR_q^X\}}]}{1 - q} \\ &= VaR_q^X + E[(X - VaR_q^X) | X > VaR_q^X] = ES_q(X). \end{aligned}$$

De la misma manera, $E[Y | X + Y > VaR_q^{X+Y}] \leq ES_q(Y)$. Así,

$$ES_q(X + Y) \leq ES_q(X) + ES_q(Y).$$

3) *Homogeneidad Positiva*. P.D.: Si $\lambda \geq 0$, entonces $ES_q(\lambda X) = \lambda ES_q(X)$.

Haciendo uso del Lema 2.3.2, se tiene

$$\begin{aligned} ES_q(\lambda X) &= E[\lambda X | \lambda X > VaR_q^{\lambda X}] = E[\lambda X | \lambda X > \lambda VaR_q^X] \\ &= E[\lambda X | X > VaR_q^X] = \lambda E[X | X > VaR_q^X] = \lambda ES_q(X). \end{aligned}$$

4) *Invariante bajo Translaciones*. P.D.: Para $a \in \mathbb{R}$, $ES_q(X + a) = ES_q(X) + a$.

$$ES_q(X + a) = E[X + a | X + a > VaR_q^{X+a}].$$

Por el Lema 2.3.2, esto se reduce a

$$\begin{aligned} ES_q(X + a) &= E[X + a | X + a > VaR_q^X + a] \\ &= E[X | X > VaR_q^X] + E[a | X > VaR_q^X] \\ &= ES_q(X) + a. \quad \blacksquare \end{aligned}$$

Este Teorema resulta muy importante, pues afirma que la Esperanza Condicional de la Cola (ES) es una medida de riesgo coherente. Es, por tanto, una medida de riesgo mucho más confiable que la varianza y el VaR, ya que éstas no son subaditivas en general y pueden ocasionar serios problemas si son usadas como único recurso en la medición de riesgos. En cambio, de acuerdo con este Teorema, la ES no causará ningún problema, inclusive si las distribuciones conjuntas con que se trabaja no son Gaussianas.

Puesto que el VaR es una medida de riesgo muy usada en la práctica e incluso a veces es requerida por los reguladores, en este trabajo seguiremos considerándolo como una medida de riesgo importante. Sin embargo, hay que enfatizar que, en estos casos, el VaR debería ser usado en conjunto con la ES, de modo que puedan entregársele al regulador los resultados que requiere pero que, al mismo tiempo, la empresa esté utilizando, internamente, una medida de riesgo coherente que le permita controlar su exposición al riesgo de manera efectiva.

2.5 El Mundo Elíptico y la Medición de Riesgos

En las secciones anteriores hemos visto que pueden surgir problemas al utilizar el VaR y la varianza como medidas de riesgo. Esto se debe a que, originalmente, el VaR y la varianza fueron concebidos para utilizarse con la distribución Gaussiana o, más generalmente, con distribuciones elípticas; así, cuando estas medidas de riesgo se usan con distribuciones que *no* pertenecen al mundo elíptico, no puede garantizarse que cumplan con la propiedad de subaditividad. En esta sección describiremos brevemente

las distribuciones esféricas y elípticas, así como el comportamiento de la varianza, el VaR y la ES dentro del mundo elíptico. Para obtener más información sobre las distribuciones elípticas puede consultarse [11].

Las distribuciones esféricas son una extensión de la distribución normal estándar multivariada $N_n(0, 1)$, es decir, de la distribución de variables normal estándar independientes. Son una familia de distribuciones simétricas para vectores no correlacionados y con media cero.

Definición 2.5.1 *Un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)^t$ tiene una distribución esférica si para cualquier mapeo ortogonal $T \in \mathbb{R}^{n \times n}$ (es decir, los mapeos que satisfacen $TT^t = T^tT = I_{n \times n}$) se tiene que*

$$T\mathbf{X} \stackrel{d}{=} \mathbf{X}. \quad (2.15)$$

La función característica $\psi(\mathbf{t}) = E[\exp(it^t\mathbf{X})]$ de las distribuciones esféricas tiene una forma particularmente sencilla. Existe una función $\phi: \mathbb{R}_+ \rightarrow \mathbb{R}$ tal que $\psi(\mathbf{t}) = \phi(\mathbf{t}^t\mathbf{t}) = \phi(t_1^2 + \dots + t_n^2)$. A esta función ϕ se le llama el generador característico de la distribución esférica \mathbf{X} , lo cual se denota

$$\mathbf{X} \sim S_n(\phi).$$

Si \mathbf{X} tiene densidad $f(\mathbf{x}) = f(x_1, \dots, x_n)$, entonces esto es equivalente a $f(\mathbf{x}) = g(\mathbf{x}^t\mathbf{x}) = g(x_1^2 + \dots + x_n^2)$ para alguna función $g: \mathbb{R}_+ \rightarrow \mathbb{R}_+$, por lo que la interpretación más sencilla de las distribuciones esféricas nos dice que hablamos de aquellas distribuciones cuya densidad es constante en esferas.

Ejemplo 2.5.2 *Algunos ejemplos de densidades en la clase esférica son la distribución t multivariada con ν grados de libertad $f(\mathbf{x}) = c(1 + \mathbf{x}^t\mathbf{x}/\nu)^{-(n+\nu)/2}$ y la dis-*

tribución logística $f(\mathbf{x}) = c \exp(-\mathbf{x}'\mathbf{x}) / [1 + \exp(-\mathbf{x}'\mathbf{x})]^2$, donde c es una constante normalizadora. Hay que notar que éstas son las distribuciones de variables aleatorias no-correlacionadas pero, contrariamente al caso de la normal, no son las distribuciones de variables aleatorias independientes. De hecho, dentro de la clase de las distribuciones esféricas la normal multivariada es la única distribución de variables aleatorias independientes.

Las distribuciones esféricas tienen una representación estocástica alternativa: $\mathbf{X} \sim S_n(\phi)$ si y sólo si

$$\mathbf{X} \stackrel{d}{=} R \cdot \mathbf{U},$$

donde el vector aleatorio \mathbf{U} se distribuye uniformemente sobre la hiperesfera unitaria $S_{n-1} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}'\mathbf{x} = 1\}$ en \mathbb{R}^n y $R \geq 0$ (llamada la *variable generadora*) es una variable aleatoria positiva, independiente de \mathbf{U} . Así, las distribuciones esféricas pueden interpretarse como mezclas de distribuciones uniformes en esferas de distintos radios en \mathbb{R}^n .

Ejemplo 2.5.3 En el caso de la distribución normal estándar multivariada, la variable generadora satisface $R \sim \sqrt{\chi_n^2}$. En el caso de la distribución t multivariada con ν grados de libertad, se da $R^2/n \sim F(n, \nu)$, donde $F(n, \nu)$ denota una distribución F con n y ν grados de libertad.

Un paso más allá de las distribuciones esféricas se encuentran las distribuciones elípticas.

Las distribuciones elípticas extienden a la distribución normal multivariada $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

es decir, la distribución normal con media μ y matriz de covarianzas Σ . Matemáticamente, son los mapeos afines de las distribuciones esféricas en \mathbb{R}^n .

Definición 2.5.4 Sea $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ un mapeo afín tal que $T(\mathbf{x}) = A\mathbf{x} + \mu$ para algunas $A \in \mathbb{R}^{n \times n}$ y $\mu \in \mathbb{R}^n$. \mathbf{X} tiene una distribución elíptica si $\mathbf{X} = T(\mathbf{Y})$ y $\mathbf{Y} \sim S_n(\phi)$.

Puesto que la función característica puede escribirse de la siguiente manera

$$\begin{aligned}\psi(\mathbf{t}) &= E[\exp(it^t \mathbf{X})] = E[\exp(it^t (A\mathbf{Y} + \mu))] \\ &= \exp(it^t \mu) \exp(i(A^t \mathbf{t})^t \mathbf{Y}) = \exp(it^t \mu) \phi(\mathbf{t}^t \Sigma \mathbf{t}),\end{aligned}$$

donde $\Sigma := AA^t$.

La notación para las distribuciones elípticas es la siguiente:

$$\mathbf{X} \sim E_n(\mu, \Sigma, \phi).$$

Por ejemplo, la distribución normal multivariada $N_n(\mu, \Sigma) = E_n(\mu, \Sigma, \phi)$, con $\phi(\mathbf{t}) = \exp(-\mathbf{t}^2/2)$.

Si \mathbf{Y} tiene densidad $f(\mathbf{y}) = g(\mathbf{y}^t \mathbf{y})$ y si A es una matriz regular (es decir, $\det(A) \neq 0$ y, por tanto, Σ es definida estrictamente positiva), entonces $\mathbf{X} = A\mathbf{Y} + \mu$ tiene densidad

$$h(\mathbf{x}) = \frac{1}{\sqrt{\det(\Sigma)}} g((\mathbf{x} - \mu)^t \Sigma^{-1} (\mathbf{x} - \mu)),$$

por lo que los contornos de igual densidad ahora son elipsoides.

El conocer la distribución de \mathbf{X} no determina por completo la representación elíptica $E_n(\mu, \Sigma, \phi)$, pues determina de manera única μ , pero no Σ y ϕ , las cuales

quedan determinadas excepto por constantes positivas. En efecto, si \mathbf{X} es elíptica y no-degenerada, existen μ , A y $\mathbf{Y} \sim S_n(\phi)$ tales que $\mathbf{X} \stackrel{d}{=} A\mathbf{Y} + \mu$, pero para cualquier $\lambda \in \mathbb{R} \setminus \{0\}$ se tiene que también $\mathbf{X} \stackrel{d}{=} (A/\lambda)\lambda\mathbf{Y} + \mu$, donde $\lambda\mathbf{Y} \sim S_n(\tilde{\phi})$ y $\tilde{\phi}(u) := \phi(\lambda^2 u)$. En general, si $\mathbf{X} \sim E_n(\mu, \Sigma, \phi) = E_n(\tilde{\mu}, \tilde{\Sigma}, \tilde{\phi})$, entonces $\mu = \tilde{\mu}$ y existe $c > 0$ tal que $\tilde{\Sigma} := c\Sigma$ y $\tilde{\phi}(u) = \phi(u/c)$.

Un enfoque particular sería el elegir Σ de manera que pueda interpretarse directamente como la matriz de covarianzas de \mathbf{X} . Para ello, sea $\mathbf{X} \sim E_n(\mu, \Sigma, \phi)$, de manera que $\mathbf{X} \stackrel{d}{=} A\mathbf{Y} + \mu$ para $\Sigma = AA^t$ y \mathbf{Y} un vector aleatorio que satisfaga $\mathbf{Y} \sim S_n(\phi)$. Es equivalente tener $\mathbf{Y} \stackrel{d}{=} R \cdot \mathbf{U}$, donde \mathbf{U} se distribuye uniformemente en S^{n-1} y R es una variable aleatoria positiva independiente de \mathbf{U} . Si $E[R^2] < \infty$, se sigue que $E[\mathbf{X}] = \mu$ y $Cov[\mathbf{X}] = AA^t E[R^2]/n = \Sigma E[R^2]/n$, puesto que $Cov[\mathbf{U}] = I_{n \times n}/n$. Así, si empezamos con el generador característico $\tilde{\phi}(u) := \phi(u/c)$, donde $c = n/E[R^2]$, aseguramos que $Cov[\mathbf{X}] = \Sigma$. De manera que una distribución elíptica queda completamente determinada por su media, su matriz de covarianzas y su generador característico.

Veremos ahora algunas de las razones por las que la varianza (y covarianza) es una medida de riesgo que surge de manera natural en el mundo de las distribuciones elípticas. La primera razón y la más importante es que las distribuciones elípticas satisfacen muchas de las propiedades de la distribución normal multivariada. Así, las combinaciones lineales, las distribuciones marginales y las distribuciones condicionales de las variables aleatorias elípticas pueden determinarse en gran medida mediante el uso de álgebra lineal, utilizando la matriz de covarianzas, la media y el generador característico. En particular, se satisfacen las siguientes propiedades:

1. Cualquier combinación lineal de un vector aleatorio con distribución elíptica es también elíptica y tiene el mismo generador característico ϕ . Esto es, si $\mathbf{X} \sim E_n(\mu, \Sigma, \phi)$ y $\mathbf{B} \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, entonces

$$\mathbf{B}\mathbf{X} + b \sim E_m(\mathbf{B}\mu + b, \mathbf{B}\Sigma\mathbf{B}^t, \phi).$$

En particular las componentes X_1, \dots, X_n de \mathbf{X} son todas variables aleatorias con distribución elíptica.

2. Las distribuciones marginales de las distribuciones elípticas son también elípticas y tienen el mismo generador. Sea $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim E_n(\mu, \Sigma, \phi)$ con

$$X_1 \in \mathbb{R}^p, X_2 \in \mathbb{R}^q, p + q = n. \text{ Sea } E[\mathbf{X}] = \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \mu_1 \in \mathbb{R}^p, \mu_2 \in \mathbb{R}^q \text{ y}$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \text{ Entonces}$$

$$X_1 \sim E_p(\mu_1, \Sigma_{11}, \phi), X_2 \sim E_q(\mu_2, \Sigma_{22}, \phi).$$

3. Suponemos que Σ es definida estrictamente positiva. La distribución condicional de X_1 dado X_2 es también elíptica, aunque, en general, tiene un generador diferente:

$$X_1 | X_2 \sim E_p(\mu_{1.2}, \Sigma_{11.2}, \tilde{\phi}),$$

$$\text{donde } \mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Otra característica importante de las distribuciones elípticas es que estas distribuciones pueden someterse a los enfoques estándar de la administración de riesgos.

Respaldan el uso del Valor en Riesgo (*VaR*) como medida de riesgo, así como el enfoque media-varianza de Markowitz para administración de riesgos y optimización de portafolios.

Supongamos que $\mathbf{X} = (X_1, \dots, X_n)^t$ representa n riesgos con distribución elíptica; consideremos portafolios *lineales* de estos riesgos

$$\left\{ Z = \sum_{i=1}^n \lambda_i X_i \mid \lambda_i \in \mathbb{R} \right\}$$

con distribución F_Z .

En el mundo elíptico el uso del *VaR* como medida de riesgo tiene sentido porque el *VaR* es una medida de riesgo coherente en el sentido de Artzner *et al.* (ver Definición 2.1.1) en este mundo. De hecho, en el mundo elíptico el uso de cualquier medida de riesgo homogénea-positiva e invariante bajo translaciones para clasificar riesgos o para determinar los pesos óptimos que minimizan el riesgo de un portafolio bajo la condición de que se obtenga un cierto rendimiento es equivalente al enfoque de Markowitz, en el que la varianza se usa como medida de riesgo. Medidas de riesgo alternativas, como podrían ser el VaR_q o la esperanza condicional de la cola ES_q , nos dan distintos valores numéricos, pero no tienen ningún efecto en el manejo de los riesgos. Esto se precisa en el siguiente Teorema.

Teorema 2.5.5 *Supongamos que $\mathbf{X} \sim E_n(\mu, \Sigma, \phi)$ con $\sigma^2[X_i] < \infty$ para cada i . Sea*

$$\mathcal{P} = \left\{ Z = \sum_{i=1}^n \lambda_i X_i \mid \lambda_i \in \mathbb{R} \right\}$$

el conjunto de portafolios lineales. Entonces es cierto lo siguiente:

1. (Subaditividad del VaR). Para cualesquiera dos portafolios $Z_1, Z_2 \in \mathcal{P}$ y $0.5 \leq q < 1$, $VaR_q(Z_1 + Z_2) \leq VaR_q(Z_1) + VaR_q(Z_2)$.

2. (Equivalencia de la varianza y otras medidas de riesgo homogéneas positivas). Sea ρ una medida de riesgo de valores reales en el espacio de variables aleatorias de valores reales y que depende sólo de la distribución de la variable aleatoria X . Supongamos que esta medida satisface que, para $\lambda \geq 0$, $\rho(\lambda X) = \lambda \rho(X)$ (es decir, satisface la propiedad de homogeneidad positiva dada en (2.2)). Entonces, para $Z_1, Z_2 \in \mathcal{P}$ $\rho(Z_1 - E[Z_1]) \leq \rho(Z_2 - E[Z_2]) \Leftrightarrow \sigma^2[Z_1] \leq \sigma^2[Z_2]$.

3. (Portafolio de Markowitz que minimiza el riesgo). Sea ρ como en el punto anterior (2) y supongamos que también es invariante bajo traslaciones, es decir, satisface que para cualquier $a \in \mathbb{R}$, $\rho(X + a) = \rho(X) + a$. Sea

$$\mathcal{E} = \left\{ Z = \sum_{i=1}^n \lambda_i X_i \mid \lambda_i \in \mathbb{R}, \sum_{i=1}^n \lambda_i = 1, E[Z] = r \right\}$$

el conjunto de portafolios que dan un rendimiento esperado r . Entonces

$$\arg \min_{Z \in \mathcal{E}} \rho(Z) = \arg \min_{Z \in \mathcal{E}} \sigma^2[Z].$$

Demostración:

La observación básica para la prueba de este Teorema es que $(Z_1, Z_2)^t$ tiene una distribución elíptica, por lo que Z_1 , Z_2 y $Z_1 + Z_2$ también tienen distribuciones elípticas.

(1) Sea t_q el q -ésimo cuantil de la distribución estandarizada del tipo correspondiente. Entonces:

$$VaR_q(Z_1) = E[Z_1] + \sigma[Z_1]t_q,$$

$$\text{VaR}_q(Z_2) = E[Z_2] + \sigma[Z_2]t_q,$$

$$\text{VaR}_q(Z_1 + Z_2) = E[Z_1 + Z_2] + \sigma[Z_1 + Z_2]t_q.$$

Puesto que $\sigma[Z_1 + Z_2] \leq \sigma[Z_1] + \sigma[Z_2]$ y $t_q \geq 0$, se tiene que

$$\text{VaR}_q(Z_1 + Z_2) \leq \text{VaR}_q(Z_1) + \text{VaR}_q(Z_2).$$

(2) Puesto que Z_1 y Z_2 son v.a. del mismo tipo, debe existir $a > 0$ tal que $Z_1 - E[Z_1] \stackrel{d}{=} a(Z_2 - E[Z_2])$. Se sigue entonces que

$$\rho(Z_1 - E[Z_1]) \leq \rho(Z_2 - E[Z_2]) \Leftrightarrow a \leq 1 \Leftrightarrow \sigma^2[Z_1] \leq \sigma^2[Z_2].$$

(3) Se sigue del punto anterior y del hecho de que se optimiza sobre portafolios con la misma esperanza. ■

Este teorema muestra que en el mundo elíptico el portafolio de Markowitz que minimiza la varianza, minimiza también otras importantes medidas de riesgo, como el Valor en Riesgo (VaR) y la Esperanza Condicional de la Cola (ES).

Hasta aquí hemos presentado una introducción al concepto de medida de riesgo, así como las definiciones de tres medidas de riesgo muy utilizadas en la práctica: la Varianza, el Valor en Riesgo y la Esperanza Condicional de la Cola. También hemos visto las características de estas medidas de riesgo y hemos ilustrado con ejemplos los problemas que pueden surgir al utilizar el VaR y la Varianza fuera del mundo elíptico.

Sin embargo, nuestro interés principal es el uso de las medidas de riesgo para “medir” eventos extremos, los cuales generalmente se nos presentan a través de observaciones del mundo real. Esto ocasiona que *no* siempre conozcamos exactamente la distribución que siguen estos datos. En el siguiente Capítulo veremos cómo podemos abordar este problema, utilizando la teoría expuesta en el Capítulo 1 (Preliminares).

Capítulo 3

Medidas de Riesgo y Valores

Extremos

El riesgo por eventos extremos está presente en todas las áreas de la administración de riesgos. Ya sea que nos interese el riesgo de mercado, de crédito, operacional o de seguros, uno de los mayores retos para el administrador de riesgos es la implementación de modelos en los que se consideran los eventos raros y dañinos, haciendo posible la medición de sus consecuencias.

Cualquiera que sea el tipo de riesgo que se esté considerando, el enfoque que usaremos será el mismo: intentaremos modelarlo de tal manera que se contemple la posibilidad de un resultado extremo. Utilizando ese modelo, trataremos de medir el riesgo con una medida que proporcione información acerca del resultado extremo. Puesto que tratamos de medir riesgos extremos, usaremos como medidas de riesgo el VaR y la Esperanza Condicional de la Cola (ES), pues ambas medidas de riesgo intentan describir la cola de una distribución de pérdidas. Como hasta ahora, tomaremos

a las pérdidas como números positivos y a las ganancias como números negativos.

Ejemplo 3.0.6 *Supongamos que una compañía aseguradora tiene los datos x_1, x_2, \dots, x_n , que son reclamos de seguro por incendio en un período de tiempo específico. Dependiendo del tipo de incendio que ocasionó los reclamos, podría o no justificarse una condición del tipo " x_1, \dots, x_n provienen de una muestra i.i.d. X_1, \dots, X_n con f.d. F ". Supongamos que sí es posible imponer esta condición. Supongamos también que necesitamos poner precio a un contrato de seguros para pérdidas excesivas; es claro que muy pocos de nuestros datos entrarán en el intervalo que hemos caracterizado como pérdidas excesivas. Si ajustamos la forma paramétrica de una distribución a todo el conjunto de datos, lo más seguro es que no se ajuste particularmente bien a esta área de la cola en que los datos son escasos. La compañía aseguradora tiene entonces dos opciones: puede decidir no hacer el contrato de pérdidas excesivas por tener poca experiencia en ese tipo de pérdidas; o bien, si desea vender el contrato, debe obtener un buen estimado de la distribución en el área de la cola.*

Para resolver este problema, la compañía puede usar los métodos -basados en la Teoría de Valores Extremos (TVE)- que se introducen en este capítulo. Estos métodos nos ofrecen buenos modelos para explicar los eventos extremos que han ocurrido en el pasado y, si bien no predicen el futuro con certidumbre, si nos dan una idea bastante aproximada de la frecuencia y magnitud que podrían tener los eventos extremos en el futuro. Hay que señalar que la estimación fuera del rango de los datos, como la requerida en el ejemplo anterior, sólo es posible si se imponen restricciones adicionales al modelo. No existen técnicas mágicas que proporcionen

resultados confiables a cambio de nada.

Sean X_1, X_2, \dots variables aleatorias idénticamente distribuidas, con función de distribución F desconocida. Estos riesgos aleatorios pueden representar cualquiera de las siguientes:

- Rendimientos (negativos) diarios de un activo o portafolio financiero: pérdidas y ganancias.
- Rendimientos de mayor o menor frecuencia.
- Pérdidas operacionales.
- Reclamaciones catastróficas en seguros.
- Pérdidas de crédito.
- Niveles de agua de un río.

Además, pueden representar riesgos que podemos observar directamente o riesgos que debemos simular mediante un procedimiento Monte Carlo, debido a la dificultad de obtención de los datos. Hay situaciones en las que, a pesar de que se lleva a cabo una simulación a partir de un modelo estocástico conocido, la complejidad del sistema es tan grande que no conocemos exactamente cuál es la distribución de pérdidas F .

Confiamos en los resultados para sumas (que incluyen a la distribución Normal, las distribuciones α -estables y el movimiento Browniano) cuando queremos modelar/poner precio/determinar el tamaño de las reservas de fenómenos aleatorios basados en promedios. De la misma manera, confiamos en las técnicas estadísticas

basadas en estas herramientas cuando se aplican para estimar el comportamiento de la distribución “no muy lejos” de la media. Sin embargo, hay ocasiones en que nos interesan los extremos; en estos casos, los resultados para máximos son una herramienta que se adapta mejor que los resultados para sumas a los datos de eventos extremos. Considérese el siguiente ejemplo.

Ejemplo 3.0.7 *Se dice que dentro de un cierto portafolio las reclamaciones se comportan como una distribución exponencial con media 10 (millones de pesos). Se han observado ya 100 reclamaciones de este tipo, siendo la pérdida más grande 50. ¿Todavía podemos creer en este modelo? ¿Y si la pérdida más grande hubiera sido 100?*

Solución:

La suposición básica es que

$$X_1, X_2, \dots, X_{100} \text{ son i.i.d. con } P(X_1 \leq x) = 1 - e^{-\frac{x}{10}}, \quad x \geq 0$$

Entonces, para $M_n = \max(X_1, X_2, \dots, X_n)$,

$$P(M_{100} > x) = 1 - (P(X_1 \leq x))^{100} = 1 - (1 - e^{-\frac{x}{10}})^{100}$$

De aquí se obtiene

$$P(M_{100} \geq 50) = 0.4914$$

$$P(M_{100} \geq 100) = 0.00453$$

Sin embargo, en vez de hacer los cálculos exactos que acabamos de llevar a cabo, consideremos el siguiente argumento asintótico. Primero, para toda $n \geq 1$ y $x \in \mathbb{R}$,

$$P\left(\frac{M_n}{10} - \log n \leq x\right) = P(M_n \leq 10(x + \log n))$$

$$\begin{aligned}
&= \left(1 - e^{-\left(\frac{10(x+\log n)}{10}\right)}\right)^n \\
&= \left(1 - e^{-x - \log n}\right)^n \\
&= \left(1 - \frac{e^{-x}}{e^{\log n}}\right)^n \\
&= \left(1 - \frac{e^{-x}}{n}\right)^n,
\end{aligned}$$

de modo que

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n}{10} - \log n \leq x\right) = e^{-e^{-x}} \equiv \Lambda(x)$$

Así, podemos usar la aproximación

$$P(M_n \leq x) \approx \Lambda\left(\frac{x}{10} - \log n\right)$$

para obtener

$$P(M_{100} \geq 50) \approx 0.4902$$

$$P(M_{100} \geq 100) \approx 0.00453$$

que es muy similar a lo obtenido con los cálculos exactos que realizamos.

El punto importante aquí es que si nos hicieran esta misma pregunta en un caso en el que tuviéramos mucha menos información específica acerca de $F(x) = P(X_1 \leq x)$, podríamos aplicar TVE para poder proceder.

A continuación se proporciona una breve descripción del contenido de este capítulo. En la Sección 3.1 se mencionan dos diferentes tipos de modelos para medir riesgos extremos que pueden construirse con base en la TVE: el Modelo de Picos sobre el Umbral (POT) y el Modelo de Máximos de Bloque; asimismo, se dan algunos ejemplos para cuya solución estos modelos resultan de gran utilidad. En la Sección

3.2 se expone el Modelo de Picos sobre el Umbral en el que, utilizando el Teorema Gnedenko-Balkema-Pickands-de Haan (TGBPdH) (ver Teorema 1.4.7), se aproxima la distribución de exceso F_u por encima del umbral u mediante una Distribución Pareto Generalizada (DPG). Aquí vemos, paso a paso, cómo obtener los estimadores necesarios (el de la distribución de exceso, el de la cola de la distribución y el del VaR_q) para finalmente obtener el estimador de la esperanza condicional de la cola ES_q . Se proporciona, asimismo, el Estimador de Máxima Verosimilitud (EMV) como una opción para obtener los estimados de los parámetros de la DPG. La Sección 3.3 está dedicada al Modelo de Máximos de Bloque. Este modelo utiliza el Teorema de Fisher-Tippet (ver Teorema 1.2.13) para justificar el ajuste de una Distribución de Valores Extremos Generalizada (VEG) a los máximos por bloque de la distribución subyacente desconocida. Puede entonces calcularse una especie de cuantil, conocido como el nivel de retorno, que puede usarse para obtener estimaciones del VaR_q . Puesto que la parte estadística se encuentra asociada también al Modelo de Máximos de Bloque, en la Sección 3.4 se ofrecen tres técnicas para estimar los parámetros de la distribución VEG: el Estimador de Máxima Verosimilitud (EMV), el Estimador de Hill (EH) y el Estimador Deckers-Einmahl-de Haan (EDEdH); al mismo tiempo, se hace mención de algunas de las propiedades de cada uno de ellos.

En este Capítulo no se presentan ejemplos que ilustren detalladamente la aplicación del Modelo de Máximos de Bloque o del Modelo de Picos sobre el Umbral porque ambos modelos se utilizarán en el Capítulo 4 para el análisis de datos reales.

La teoría y ejemplos de este capítulo pueden encontrarse en [6], [17], [18], [8] y [7]. Otras referencias son [16], [19], [5] y [23].

3.1 La TVE puesta en acción

Crear un modelo para un riesgo implica seleccionar una distribución particular. Para hacerlo apropiadamente es necesario, además de utilizar datos históricos, tener una base teórica que fundamente la elección de la función de distribución. Esta base teórica es, en nuestro caso, la Teoría de Valores Extremos.

Consideraremos dos tipos de modelos para valores extremos. El grupo de modelos más antiguo es el de los Modelos de Máximos de Bloque, que son modelos para las mayores observaciones recolectadas de grandes muestras de observaciones idénticamente distribuidas. Por ejemplo, si anotamos cada hora o diariamente las pérdidas y ganancias de un instrumento o grupo de instrumentos financieros particulares, el método de máximos de bloque nos proporcionaría un modelo que pudiera ser adecuado para el máximo trimestral o anual de tales observaciones. Podemos ver una posible aplicación para este método en la definición de pérdidas acentuadas, que se presenta en el siguiente ejemplo.

Ejemplo 3.1.1 Los grandes bancos se están interesando por los métodos actuariales para determinar el tamaño de sus reservas y así poder protegerse contra pérdidas crediticias futuras. Así, Swiss Bank Corporation diseñó el ACRA (Actuarial Credit Risk Accounting) para la administración de riesgos. En su estructura de medición de riesgos, utilizan las siguientes definiciones para pérdidas:

- **Pérdidas Esperadas:** Son las pérdidas que debe suponerse que ocurrirán continuamente como consecuencia de participar en cierto tipo de negocios.

- **Pérdidas Inesperadas:** Son las pérdidas inusuales, aunque predecibles, que el Banco debe ser capaz de absorber en el curso normal de su negocio.
- **Pérdidas Acentuadas:** Son las situaciones posibles, aunque improbables, que el Banco debe poder sobrevivir.

La TVE permite cuantificar las fronteras entre estos distintos tipos de pérdidas, lo que el Banco utiliza para determinar qué tipo de recursos (ganacias, reservas o capital contable) son los que deben utilizarse para cubrir cada una de sus pérdidas.

Un grupo de modelos más moderno es el de los modelos de Picos Sobre el Umbral (POT; *Peaks Over Threshold*), que son modelos para todas las observaciones (grandes) que exceden un umbral *alto*. Los modelos POT se consideran en general más útiles para aplicaciones prácticas, debido a que usan más eficientemente los datos de valores extremos, que a menudo son limitados.

Dentro de la clase de los modelos POT pueden distinguirse dos estilos de análisis. Existen los modelos semi-paramétricos, contruidos alrededor del estimador de Hill y sus similares; y los modelos completamente paramétricos, basados en la Distribución de Pareto Generalizada (DPG). Si bien ambos enfoques están justificados teóricamente y son útiles en la práctica cuando se utilizan de manera correcta, en este trabajo utilizaremos los modelos paramétricos. Con ellos se obtienen fórmulas paramétricas sencillas de medidas de riesgo extremo para las cuales es relativamente fácil dar estimaciones del error estadístico utilizando las técnicas de inferencia máximo-verosímil.

Recordemos que nuestras medidas de riesgo extremo están definidas en términos

de la distribución de pérdidas F . Tenemos que, para $q \in [0.95, 1)$,

$$VaR_q = F^{-}(q),$$

$$ES_q = E[X | X > VaR_q],$$

donde la restricción para los valores de q conforme al capítulo anterior se debe a que ahora nos interesan los riesgos extremos, lo que nos sitúa en la cola de la distribución.

Al igual que nos ocurre con F , VaR_q y ES_q son cantidades teóricas que nunca conoceremos. Nuestra meta en la medición de riesgos es obtener estimadores \widehat{VaR}_q y \widehat{ES}_q de estas medidas.

3.2 Modelo de Picos Sobre el Umbral (POT)

La meta del modelo POT es obtener un estimador de la esperanza condicional de la cola ES_q , para $q \in (0.95, 1)$. Para ello, se usa el Teorema Gnedenko-Balkema-Pickands-de Haan (ver Teorema 1.4.7) para aproximar la distribución de exceso F_u por encima de un umbral determinado u mediante una DPG. Una vez que se tiene el estimador de la distribución de exceso, es fácil reformular las expresiones en términos de la cola \overline{F} de la distribución y, puesto que el VaR_q es un alto cuantil de la distribución, éste puede obtenerse invirtiendo la fórmula para el estimador de colas. Podemos entonces proceder a estimar el valor de ES_q , utilizando la relación que, como vimos en la Sección 2.4, existe entre la esperanza condicional de la cola y el VaR_q :

$$ES_q = VaR_q + E[X - VaR_q | X > VaR_q].$$

Estos pasos se explican detalladamente a continuación.

Supongamos que X, X_1, \dots, X_n son i.i.d. con f.d. $F \in DAM(H_\xi)$ para algún $\xi \in \mathbb{R}$.

En general, supondremos que la distribución F subyacente es una distribución con extremo derecho infinito, es decir, que permite la posibilidad de que ocurran pérdidas arbitrariamente grandes, aún cuando atribuye probabilidades insignificantes a los resultados irrazonablemente grandes, como lo hacen la distribución normal y la distribución t -student. Sin embargo, también puede darse el caso de que se utilice F con extremo derecho finito. Un ejemplo de esto es la distribución Beta en el intervalo $[0, 1]$, que asigna probabilidad cero a los resultados mayores que 1 y que podría usarse, por ejemplo, como la distribución de pérdidas de crédito expresadas en proporción a la exposición.

Empezamos por elegir un umbral u y denotamos

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

el número de excesos de u por parte de X_1, \dots, X_n . Denotemos a los correspondientes excesos Y_1, \dots, Y_{N_u} .

En el Capítulo 1 (ver 1.41) hemos definido la distribución de pérdidas excesivas por encima de un umbral alto u mediante

$$F_u(y) = P[X - u \leq y \mid X > u] = P[Y \leq y \mid X > u], \quad 0 \leq y < x_F - u \quad (3.1)$$

La distribución de exceso representa la probabilidad de que una pérdida exceda el umbral u por a lo más una cantidad y , dado que sabemos que ha excedido el umbral.

Es muy útil notar que la distribución de exceso F_u puede escribirse en términos de la

distribución F subyacente:

$$\overline{F}(u + y) = \overline{F}(u) \overline{F}_u(y). \quad (3.2)$$

Recordemos ahora la definición de la Distribución Pareto Generalizada (DPG):

una DPG está dada por

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

donde $\beta > 0$, y donde $x \geq 0$ cuando $\xi \geq 0$ y $0 \leq x \leq -\beta/\xi$ cuando $\xi < 0$. De aquí,

es claro que la cola de una DPG $G_{\xi, \beta}$ con parámetros $\xi \in \mathbb{R}$ y $\beta > 0$ es

$$\overline{G}_{\xi, \beta}(x) = \begin{cases} (1 + \xi x/\beta)^{-1/\xi} & \text{si } \xi \neq 0, \\ \exp(-x/\beta) & \text{si } \xi = 0. \end{cases} \quad x \in D(\xi, \beta),$$

donde

$$D(\xi, \beta) = \begin{cases} [0, \infty) & \text{si } \xi \geq 0, \\ [0, -\beta/\xi] & \text{si } \xi < 0. \end{cases} \quad (3.3)$$

El Teorema Gnedenko-Balkema-Pickands-de Haan nos dice que si X_1, \dots, X_n son i.i.d. con f.d. $F \in DAM(H_\xi)$, $\xi \in \mathbb{R}$, entonces existe una función $\beta: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ tal que

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0. \quad (3.4)$$

Basándonos en este resultado límite para $\overline{F}_u(y)$, podemos pensar en la siguiente aproximación para valores grandes de u :

$$\overline{F}_u(y) \approx \overline{G}_{\xi, \beta(u)}(y). \quad (3.5)$$

Es importante notar que que β es una función del umbral u . En la práctica, u debe tomarse suficientemente grande. Dado un valor (grande) de u , ξ y $\beta = \beta(u)$ se estiman a partir de los datos de excesos, por lo que los estimados resultantes dependen de u .

La relación (3.2) nos deja ver un método para estimar la cola de F mediante la estimación, por separado, de $\bar{F}_u(y)$ y $\bar{F}(u)$. Un estimador natural para $\bar{F}(u)$ está dado por la f.d. empírica

$$\left(\widehat{\bar{F}(u)}\right) = \bar{F}_n(u) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i > u\}} = \frac{N_u}{n}.$$

Esto está justificado por el Teorema de Glivenko-Cantelli (ver Apéndice C), que nos dice que la función de distribución empírica es una aproximación uniforme a la verdadera función de distribución.

Por otra parte, la aproximación de la distribución de exceso mediante una DPG dada en la relación (3.5) -la cual podemos utilizar porque u es grande- motiva un estimado de la forma

$$\left(\widehat{\bar{F}_u}(y)\right) \approx \bar{G}_{\hat{\xi}, \hat{\beta}}(y) \quad (3.6)$$

para $\hat{\xi} = \hat{\xi}_{N_u}$ y $\hat{\beta} = \hat{\beta}_{N_u}$ apropiados. En el Apartado 3.2.1 se mencionará un método para obtener estos estimadores de los parámetros de la DPG.

Una pregunta que surge en este momento es ¿por qué no utilizamos el estimador empírico (o, dicho de otra manera, el método de Simulación Histórica) para estimar toda la cola de F , es decir, $F(x)$ para toda $x \geq u$? La respuesta es que la Simulación Histórica (SH) es un método bastante pobre en la cola de la distribución, pues los datos se vuelven escasos. Al fijar el umbral en u estamos considerando que se tienen suficientes observaciones que exceden a u como para permitir que obtengamos un estimado razonable de $\bar{F}(u)$ a través de SH, pero para niveles más altos el método sería demasiado poco confiable.

Al sustituir nuestro estimado de $\bar{F}(u)$ obtenido por SH, junto con el estimado

sugerido por (3.6), se obtiene el estimado de la cola $\overline{F}(u+y)$:

$$\left(\widehat{\overline{F}}(u+y)\right) = \frac{N_u}{n} \left(1 + \widehat{\xi} \frac{y}{\widehat{\beta}}\right)^{-1/\widehat{\xi}}, \quad y > 0. \quad (3.7)$$

Este estimado puede verse como un tipo de estimado por SH aumentado con TVE y puede construirse siempre que creamos que los datos vienen de una distribución común, aunque sus propiedades estadísticas se entienden mejor en las situaciones en que los datos se suponen independientes o sólo "levemente dependientes".

Es fácil notar que $\widehat{\overline{F}}(u+y)$ es también una Distribución Pareto Generalizada (DPG) con el mismo parámetro de forma ξ , pero con parámetro de escala $\tilde{\beta} = \beta(1 - F(u))^\xi$ y parámetro de localización $\tilde{\nu} = \nu - \tilde{\beta} \left[(1 - F(u))^{-\xi} - 1 \right] / \xi$.

Lo que estamos haciendo es extrapolar al área en que los datos se vuelven escasos y son, por tanto, una guía poco confiable para su distribución.

En los casos Fréchet y Gumbel ($\xi \geq 0$), la restricción para el dominio en (3.7) es $y \geq 0$, enfatizando el hecho de que estimamos \overline{F} en la cola superior (o cola derecha).

Ahora, dada una probabilidad $q > F(u)$, lo único que necesitamos para obtener el estimador \widehat{VaR}_q de VaR_q es invertir la fórmula (3.7) del estimador de colas. Así,

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

de donde

$$\frac{\widehat{\beta}}{\widehat{\xi}} \left(\left(\frac{n}{N_u} \widehat{\overline{F}}(u+y) \right)^{-\widehat{\xi}} - 1 \right) = y \quad (3.8)$$

Sea $x = u + y$. Entonces, al hacer $\widehat{\overline{F}}(u+y) = \widehat{\overline{F}}(x) = 1 - q$ se tiene que $x = y + u = \widehat{VaR}_q$, y al sustituirlo en (3.8) se obtiene la fórmula de estimación del VaR_q :

$$\widehat{VaR}_q = u + \frac{\widehat{\beta}}{\widehat{\xi}} \left(\left(\frac{n}{N_u} (1 - q) \right)^{-\widehat{\xi}} - 1 \right). \quad (3.9)$$

En lenguaje estadístico estándar, éste es el estimador de un cuantil, donde el cuantil es un parámetro desconocido de una distribución subyacente desconocida.

Más aún, si $\hat{\xi} < 0$, un estimado del extremo derecho x_F de F está dado por

$$x_F = u - \frac{\hat{\beta}}{\hat{\xi}}$$

Esta relación se obtiene haciendo $\hat{x}_F = \hat{x}_1$ (es decir, $q = 1$) en la relación (3.9).

Como dijimos antes, la meta del método POT es la obtención de un estimado de la Esperanza Condicional de la Cola (ES). Tenemos, además, que ES y VaR están relacionados mediante la expresión

$$ES_q = VaR_q + E[X - VaR_q | X > VaR_q].$$

Es fácil notar que el término $E[X - VaR_q | X > VaR_q]$ es simplemente la esperanza de la distribución de exceso $F_{VaR_q}(y)$ por encima del umbral VaR_q . Además, nuestro modelo (3.5) para la distribución de exceso por encima del umbral u tiene la propiedad de que, para umbrales más altos VaR_q con $q > F(u)$ la distribución de exceso por encima de este umbral más alto es también una DPG con el mismo parámetro de forma, aunque con escala distinta. Esto se demostró en el Teorema 1.4.7, y un argumento alternativo es el siguiente: si suponemos que los excesos sobre el umbral u tienen distribución DPG, es decir, $(X - u | X > u) \sim G_{\xi, \beta}$, basta notar que para $VaR_q > u$ podemos escribir

$$X - VaR_q | X > VaR_q = (X - u) - (VaR_q - u) | (X - u) > (VaR_q - u),$$

de donde es fácil ver que

$$F_{VaR_q}(y) \approx G_{\xi, \beta + \xi(VaR_q - u)}(y). \quad (3.10)$$

La ecuación (3.10) nos proporciona un modelo explícito para las pérdidas de exceso por encima del VaR_q . Con este modelo podemos calcular muchas de las características de este tipo de pérdidas; por ejemplo, si notamos que (para $\xi < 1$) la función media de exceso de la distribución dada en (3.10) es $(\beta + \xi(VaR_q - u)) / (1 - \xi)$ (ver Tabla en la Sección 2.4 para $e(u)$ de una DPG), podemos calcular el estimado de la esperanza condicional de la cola. Se obtiene que

$$\begin{aligned}\widehat{ES}_q &= \widehat{VaR}_q + e(\widehat{VaR}_q) \\ &= \widehat{VaR}_q + \frac{\widehat{\beta} + \widehat{\xi}(\widehat{VaR}_q - u)}{1 - \widehat{\xi}} \\ &= \frac{\widehat{VaR}_q}{1 - \widehat{\xi}} + \frac{\widehat{\beta} - \widehat{\xi}u}{1 - \widehat{\xi}}.\end{aligned}$$

Reacomodando, tenemos

$$\frac{\widehat{ES}_q}{\widehat{VaR}_q} = \frac{1}{1 - \widehat{\xi}} + \frac{\widehat{\beta} - \widehat{\xi}u}{(1 - \widehat{\xi})\widehat{VaR}_q}. \quad (3.11)$$

Vale la pena examinar esta razón más detenidamente en los casos en que la distribución subyacente tiene un extremo derecho x_F infinito. En este caso la razón queda mayormente determinada por el factor $1 / (1 - \widehat{\xi})$. El segundo término del lado derecho de (3.11) se vuelve insignificamente pequeño conforme la probabilidad q se va acercando más y más a 1. Esta observación asintótica subraya la importancia del parámetro de forma ξ en la estimación de la cola, pues es este parámetro el que determina cómo difieren nuestras dos medidas de riesgo en las regiones extremas de la distribución de pérdidas.

Ahora que obtuvimos un estimado de ES, debemos subrayar que nunca debe hacerse el análisis para un único valor de u . De hecho, es recomendable hacer gráficas

para comparar las estimaciones resultantes a través de un amplio rango de valores de u . La mayoría de las gráficas exhibirán una conducta de este tipo: gran variabilidad para valores grandes de u (pocas observaciones) contra sesgo para valores pequeños de u (muchas observaciones pero, al mismo tiempo, puede no ser válida la aproximación (3.5)). En el Capítulo 4 veremos algunas técnicas gráficas que pueden ayudarnos a elegir un valor adecuado para u .

Tenemos entonces que, para la utilización del método POT, necesitamos:

- Un umbral u suficientemente alto.
- Estimados $\hat{\xi}$ y $\hat{\beta}$ para la DPG.

Ya hemos comentado sobre las dificultades para elegir un umbral “óptimo”. A continuación presentamos uno de los métodos más conocidos para obtener los estimados de los parámetros de la DPG.

3.2.1 Estimador de Máxima Verosimilitud para los parámetros de la DPG

En la relación (3.6) se hace necesaria la obtención de estimados de los parámetros de la DPG con la que se aproximará la función de distribución de exceso F_u por encima del umbral u . Uno de los métodos para obtener estos estimados es usando el Estimador de Máxima Verosimilitud (EMV), el cual se presenta enseguida.

Partimos de datos $\mathbf{X} = (X_1, \dots, X_n)$ i.i.d. con f.d. F . Suponemos que F es una

DPG con parámetros ξ y β , de manera que su densidad f es

$$f(x) = \frac{1}{\beta} \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}-1}, \quad x \in D(\xi, \beta),$$

donde $D(\xi, \beta)$ está dado en (3.3).

La función logarítmica de verosimilitud es igual a

$$l((\xi, \beta); \mathbf{X}) = -n \ln \beta - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \ln \left(1 + \frac{\xi}{\beta} X_i\right).$$

Nótese que los argumentos de esta función deben satisfacer la restricción $X_i \in D(\xi, \beta)$.

Ahora, las ecuaciones de verosimilitud pueden derivarse y resolverse numéricamente,

dando como resultado los EMV $\hat{\xi}_n$ y $\hat{\beta}_n$. El método funciona bien si $\xi > -1/2$ y, en

este caso, puede probarse que

$$n^{1/2} \left(\hat{\xi}_n - \xi, \quad \frac{\hat{\beta}_n}{\beta} - 1 \right) \xrightarrow{d} N(0, M^{-1}), \quad n \rightarrow \infty,$$

donde

$$M^{-1} = (1 + \xi) \begin{pmatrix} 1 + \xi & -1 \\ -1 & 2 \end{pmatrix}$$

y $N(\mu, \Sigma)$ denota una distribución normal bivariada cuya media es el vector μ y con matriz de covarianzas Σ . Se dan las propiedades usuales del EMV, es decir, la eficiencia y la consistencia.

Debido a la relación (3.5), resulta más realista el suponer un modelo DPG para los excesos Y_1, \dots, Y_N , donde $N = N_u$ es independiente de las Y_i . Las ecuaciones condicionales de verosimilitud que resultan se resuelven más fácilmente mediante la reparametrización $(\xi, \beta) \rightarrow (\xi, \tau)$, donde $\tau = -\xi/\beta$. Esto nos da la solución

$$\hat{\xi} = \hat{\xi}(\tau) = N^{-1} \sum_{i=1}^N \ln(1 - \tau Y_i),$$

donde τ satisface

$$h(\tau) = \frac{1}{\tau} + \frac{1}{N} \left(\frac{1}{\xi(\tau)} + 1 \right) \sum_{i=1}^N \frac{Y_i}{1 - \tau Y_i} = 0.$$

La función $h(\tau)$, definida para $\tau \in (-\infty, \max(Y_1, \dots, Y_N))$ es continua en 0.

Tenemos pues ya un método, llamado Picos Sobre el Umbral (POU), que nos permite calcular los estimados de dos importantes medidas de riesgo (la ES y el VaR), utilizando una aproximación de la función de excesos F_u mediante una DPG. En la siguiente sección se presenta el modelo de Máximos de Bloque, el cual nos permitirá obtener estimados del VaR.

3.3 Modelo de Máximos de Bloque

Consideremos una muestra X_1, \dots, X_n de v.a.i.i.d. con f.d. $F \in DAM(H_\xi)$ para algún $\xi \in \mathbb{R}$. Sea $0 < p < 1$ y sea x_p el p -cuantil correspondiente. Todo el sentido de requerir la condición $F \in DAM(H_\xi)$ es el poder estimar cuantiles fuera del rango de los datos, es decir, $p > 1 - 1/n$. Por supuesto, esto último es equivalente a encontrar estimadores para la cola $\bar{F}(x)$ para x grande.

Ejemplo 3.3.1 *Un conjunto de datos muy conocido en TVE es el del caudal del río Nidd en Yorkshire, Inglaterra. El conjunto básico está integrado por 154 observaciones de caudales por encima de los 65 CUMECS durante el período de tiempo comprendido entre 1934 y 1970. Los hidrólogos prepararon los datos utilizando una dura técnica para eliminar la posible presencia de estacionalidad y tendencia en los datos. Aunque el conjunto completo contiene una serie de valores para cada año, para*

el primer análisis se consideran solamente los máximos anuales. De esta manera, se evitan la dependencia intra-anual y resulta válido suponer que los datos x_1, \dots, x_{35} son observaciones de una sucesión X_1, \dots, X_{35} de v.a.i.i.d. con distribución de valores extremos H . Queremos contestar preguntas como: ¿Cuál es la probabilidad de que el nivel máximo de agua del año próximo exceda un nivel x ? ¿Cuál es la probabilidad de que el nivel máximo de agua del año próximo exceda todos los niveles anteriores? ¿Cuánto tiempo (en años, digamos) esperamos que pase antes de que se dé un cierto nivel (elevado) de agua?

Es claro que el poder responder estas preguntas depende de lo que sepamos acerca de la f.d. H_ξ . Para ello será de gran utilidad el Teorema de Fisher-Tippet, que recordamos a continuación.

Sean X_1, X_2, \dots, X_n v.a.i.i.d. con función de distribución F , y sean $(a_n), (b_n)$ sucesiones de constantes tales que para alguna distribución límite H no degenerada se tiene

$$F^n(a_n x + b_n) = P\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{d} H(x), \quad x \in \mathbb{R}. \quad (3.12)$$

Entonces H es del tipo de la Distribución de Valores Extremos Generalizada (VEG), es decir, $H(x) = H_\xi((x - \mu)/\sigma)$, donde

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

Este Teorema es la base teórica que necesitamos para el modelo de máximos de bloque, pues sugiere ajustar una distribución VEG a los datos de máximos muestrales, siempre que podamos conseguir ese tipo de datos.

Para implementar un análisis de pérdidas acentuadas basado en el modelo de máximos de bloque se requiere muchísima información, ya que debemos definir bloques y reducir los datos a máximos de bloque solamente. Supongamos que se tienen datos de rendimientos (negativos) diarios, los cuales se dividen en k grandes bloques de tamaño aproximadamente igual; por ejemplo, podemos tomar bloques anuales o semestrales. Sea $M_n^{(j)} = \max(X_1^{(j)}, X_2^{(j)}, \dots, X_n^{(j)})$ el máximo de las n observaciones que integran el bloque j -ésimo. No conocemos la distribución subyacente de nuestros rendimientos, pero hacemos uso del Teorema de Fisher-Tippet: empleando el método de máxima verosimilitud (o algún otro método para estimar los parámetros de la f.d. VEG), ajustamos la distribución de VEG a los datos de los máximos de bloque $M_n^{(1)}, M_n^{(2)}, \dots, M_n^{(k)}$. Esto es, suponemos que el tamaño de nuestros bloques es suficientemente grande como para que el resultado límite del Teorema de Fisher-Tippet pueda tomarse como aproximadamente exacto. Así, tendríamos que existen $a_n > 0$, $b_n \in \mathbb{R}$ tales que, para algún ξ

$$P \left\{ \frac{M_n - b_n}{a_n} \leq x \right\} \approx H_\xi(x).$$

Ahora, haciendo $y = a_n x + b_n$, se obtiene

$$\begin{aligned} P \{ M_n \leq y \} &\approx H_\xi \left(\frac{y - b_n}{a_n} \right) \\ &= H_{\xi, b_n, a_n}(y). \end{aligned}$$

Esto nos muestra que las constantes normantes a_n y b_n se encargan de los parámetros de localización y escala (μ y σ) de la distribución VEG. El parámetro crucial es el parámetro de forma ξ . En la Sección 3.4 se presentan algunos métodos para la estimación de ξ .

Para un análisis completo debemos intentar con diferentes tamaños de bloque y comparar los resultados. Al definir apropiadamente los períodos de tiempo (es decir, los tamaños de los bloques) creamos independencia en el modelo. Así, podemos suponer que nuestros datos constan de n observaciones de máximos i.i.d. que tienen exactamente una distribución VEG $H_\theta = H_{\xi,\mu,\sigma}$.

La definición usual del q -ésimo cuantil de una distribución con f.d. F es

$$x_q = F^{-1}(q). \quad (3.13)$$

Consideraremos una especie alternativa de cuantil que es particularmente fácil de estimar a partir de nuestro modelo ajustado: el nivel de retorno. Este es un concepto tomado de la hidrología que resulta útil en la administración de riesgos.

Supongamos que consideramos un modelo de máximos anuales. Definimos el nivel de retorno de k -años $R_{365,k}$ mediante

$$P\{M_{365} > R_{365,k}\} = 1/k, \quad k > 1.$$

Este es un nivel que esperaríamos que fuera excedido solamente en un año de cada k años, en promedio. De manera similar, $R_{n,k}$ es un nivel que esperamos que sea excedido en un n -bloque (es decir, un bloque de tamaño n) de entre cada k n -bloques, en promedio. En general, el nivel $R_{n,k}$ de retorno k para bloques de tamaño n se define como

$$P\{M_n > R_{n,k}\} = \frac{1}{k}, \quad (3.14)$$

Llamamos períodos difíciles o acentuados a los n -bloques en que se excede el nivel de retorno.

De modo que si pensamos que los máximos de bloques de tamaño n siguen la distribución VEG con f.d. $H_{\xi, \mu, \sigma}$, entonces $R_{n,k}$ es una función de los parámetros de esta f.d.:

$$R_{n,k} = H_{\xi, \mu, \sigma}^{-1} \left(1 - \frac{1}{k} \right) = \mu - \frac{\sigma}{\xi} \left(1 - \left(-\log \left(1 - \frac{1}{k} \right) \right)^{-\xi} \right).$$

Por tanto, podemos estimar $R_{n,k}$ usando los estimadores de los parámetros (ξ, μ, σ) , los cuales, como ya se mencionó, pueden obtenerse por el método de Máxima Verosimilitud o por algún otro método.

El nivel de retorno es un cuantil particular de la distribución marginal asociado con una probabilidad particular. En los casos en que los datos son i.i.d. es fácil calcular esta probabilidad, puesto que de (3.13) y (3.14) se deduce que

$$\left(1 - \frac{1}{k} \right)^{1/n} = P^{1/n} \{M_n \leq R_{n,k}\} = F(R_{n,k}),$$

lo que quiere decir que $R_{n,k} = x_{(1-1/k)^{1/n}}$. El nivel de retorno de 20 años sería el cuantil 0.999859 y el nivel de retorno de 10 semestres sería el cuantil 0.9994244. De manera similar, para calcular el cuantil 0.999 podríamos calcular el nivel de retorno de 6 semestres o de 11.5 trimestres.

Si nuestros datos no son i.i.d., lo que podemos decir acerca de los niveles de retorno se vuelve vago. Sabemos que el nivel de retorno $R_{n,k}$ se excede en promedio en un n -bloque de entre k n -bloques. Sin embargo, no sabemos cuántos excesos ocurrirán en ese n -bloque extremo o período acentuado en particular, puesto que esto dependerá de la propensidad de la serie a apiñonarse. En este sentido, $R_{n,k}$ nos daría una idea de la frecuencia de los períodos acentuados, pero no de la verdadera frecuencia de los resultados extremos promediados a largo plazo. Si queremos información más precisa

necesitamos conocer el índice extremo θ de la serie, que nos indica qué tan propensa es la serie a formar "grupos".

Alternativamente, en el lenguaje utilizado por los ingenieros, un evento de t -años relativo a una v.a. X con f.d. F corresponde al valor

$$x_t = F^{-1}(1 - t^{-1}).$$

Su interpretación obvia es que, con probabilidad $1 - t^{-1}$, el nivel x_t no será excedido. Con una formulación diferente, x_t es el cuantil superior $(1 - t^{-1})$ de F . Puesto que a través de los niveles de retorno podemos calcular cuantiles, entonces también podemos determinar eventos de t -años.

Ejemplo 3.3.2 *El 1 de febrero de 1953, varios diques se colapsaron en Holanda debido a una fuerte tormenta, la cual ocasionó severas inundaciones y la muerte de más de 1800 personas. El gobierno holandés nombró un Comité que debía determinar cuál debería de ser la altura de los diques (en sus diferentes localizaciones) para evitar inundaciones futuras con una probabilidad suficientemente alta. Es decir, el gobierno holandés pidió al Comité estimar las alturas de los diques que implicaran un evento de t -años, para t suficientemente grande. El Comité utilizó 166 observaciones del máximo nivel de agua registrado anualmente, tomando en cuenta una inundación comparable ocurrida en 1570, en la que el nivel del agua fue de $NAP+4m$, donde NAP es el Nivel Normal de Amsterdam. La inundación de 1953 correspondió a un nivel de $NAP+3.85m$. En base a estos datos, el Comité estimó que el cuantil $(1 - 10^{-4})$ era $NAP+5.14m$. Es evidente que la estimación de este tipo de evento de 10000-años requiere estimar mucho más allá del rango de los datos. La única manera de resolver*

un problema de este tipo es haciendo suposiciones adicionales sobre el modelo, es decir, sobre la f.d. F de los máximos anuales.

Ejemplo 3.3.3 Consideremos un tratado de Exceso de Pérdidas (XL-Treaty). Para poner precio a estos tratados, la prioridad u típicamente se determina como un evento de t -años correspondiente a un reclamo específico con f.d. F , es decir,

$$u = u_t = F^{-1} \left(1 - \frac{1}{t} \right).$$

Si lleváramos a cabo un análisis por el método de Máximos de Bloque, podríamos determinar la prioridad del tratado mediante el nivel de retorno.

3.4 Estimación de los Parámetros para la distribución VEG

En las secciones anteriores se expusieron dos métodos, basados en la TVE, que nos permiten obtener medidas de riesgo extremo. Estos modelos requieren la estimación de parámetros, en un caso de los correspondientes a la Distribución de Valores Extremos Generalizada (VEG); y en el otro, los de una Distribución de Pareto Generalizada (DPG). Vemos así que el problema estadístico está implícito en ambos modelos. Por ello, en esta sección se presentan tres estimadores: el de Máxima Verosimilitud (EMV), el de Hill (EH) y el estimador Deckers-Einmahl-de Haan (EDEDH).

El EMV es extensamente utilizado en la práctica y sus propiedades han sido ampliamente investigadas. El EDEDH funciona para $\xi \in \mathbb{R}$, mientras que el EH está diseñado sólo para $\xi > 0$. Por otra parte, el estimador de Hill es una herramienta

muy popular para detectar colas pesadas y para estimar el índice $1/\xi$ de una DPG para $\xi > 0$.

La pregunta sobre cuál estimador debería usarse en la práctica no tiene una respuesta definitiva. Todo depende de los posibles valores de ξ y de las propiedades precisas de la f.d. subyacente F .

3.4.1 El Estimador de Máxima Verosimilitud (EMV)

Supongamos que H_θ tiene densidad h_θ . Entonces la *función de verosimilitud* basada en los datos $\mathbf{X} = (X_1, \dots, X_n)$ está dada por

$$L(\theta; \mathbf{X}) = \prod_{i=1}^n h_\theta(X_i) \mathbf{1}_{\{1+\xi(X_i-\mu)/\sigma > 0\}}.$$

Denotemos mediante $l(\theta; \mathbf{X}) = \ln L(\theta; \mathbf{X})$ la *función logarítmica de verosimilitud*. El *estimador de máxima verosimilitud (EMV)* para θ es igual a

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} l(\theta; \mathbf{X}),$$

es decir, $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ maximiza $l(\theta; \mathbf{X})$ sobre un espacio apropiado de parámetros Θ . En el caso de $H_{0,\mu,\sigma}$ se obtiene

$$l((0, \mu, \sigma); \mathbf{X}) = -n \ln \sigma - \sum_{i=1}^n \exp \left\{ -\frac{X_i - \mu}{\sigma} \right\} - \sum_{i=1}^n \frac{X_i - \mu}{\sigma}.$$

Al diferenciar esta función con respecto a μ y a σ se obtienen las ecuaciones de verosimilitud en el caso Gumbel:

$$\begin{aligned} 0 &= n - \sum_{i=1}^n \exp \left\{ -\frac{X_i - \mu}{\sigma} \right\}, \\ 0 &= n + \sum_{i=1}^n \frac{X_i - \mu}{\sigma} \left(\exp \left\{ -\frac{X_i - \mu}{\sigma} \right\} - 1 \right), \end{aligned}$$

para las cuales no existe solución explícita. En los casos en que $\xi \neq 0$ es necesario utilizar métodos numéricos.

Las propiedades clásicas de regularidad del EMV son válidas siempre que $\xi > -1/2$; en estos casos, los estimadores $(\widehat{\xi}_{N_u}, \widehat{\sigma}_{N_u})$, basados en una muestra de N_u excesos sobre un umbral u , tienen distribución normal asintóticamente.

Específicamente, para un umbral u fijo se tiene que

$$N_u^{1/2} \begin{pmatrix} \widehat{\xi}_{N_u} \\ \widehat{\sigma}_{N_u} \end{pmatrix} \xrightarrow{d} N \left[\begin{pmatrix} \xi \\ \sigma \end{pmatrix}, \begin{pmatrix} (1+\xi)^2 \sigma(1+\xi) & \\ \sigma(1+\xi) & 2\sigma^2(1+\xi) \end{pmatrix} \right],$$

cuando $N_u \rightarrow \infty$. Este resultado nos permite calcular errores estándar aproximados para nuestros estimadores de máxima verosimilitud.

Dado que la mayoría de las distribuciones que se encuentran dentro de los seguros y las finanzas tienen soporte no-acotado por la derecha (lo cual sólo es posible para $\xi \geq 0$), la técnica de EMV es un procedimiento útil y confiable en estos campos.

3.4.2 El Estimador de Hill (EH) para $\xi = \alpha^{-1} > 0$.

Supongamos que X_1, \dots, X_n son i.i.d. con f.d. $F \in DAM(\Phi_\alpha)$, $\alpha > 0$, de manera que (ver Teorema 1.3.5) $F(x) = x^{-\alpha}L(x)$, $x > 0$ para una función de variación lenta L . Las distribuciones con ese tipo de colas constituyen los ejemplos principales para modelar fenómenos de cola pesada.

Un estimador natural para α es el Estimador de Hill (EH)

$$\widehat{\alpha}_{k,n}^{(H)} = \left(\frac{1}{k} \sum_{j=1}^k \ln X_{j,n} - \ln X_{k,n} \right)^{-1}, \quad (3.15)$$

donde $k = k(n) \rightarrow \infty$ (es decir, usamos un número suficientemente grande de estadísticos de orden) y $\frac{n}{k(n)} \rightarrow \infty$ (puesto que nos interesan las propiedades de la cola, debemos concentrarnos en los estadísticos de orden superior).

Dijimos que EH es un estimador natural para α porque puede derivarse a partir de varios métodos, entre ellos el método del EMV:

Supongamos por el momento que X es una v.a. con f.d. F tal que para $\alpha > 0$, $P[X > x] = \bar{F}(x) = x^{-\alpha}$, $x \geq 1$. Entonces es inmediato que $Y = \ln X$ tiene f.d. $P[Y > y] = e^{-\alpha y}$, $y \geq 0$, es decir, $Y \sim \text{Exp}(\alpha)$. De manera que el EMV de α está dado por

$$\hat{\alpha}_n = \bar{Y}_n^{-1} = \left(\frac{1}{n} \sum_{j=1}^n \ln X_j \right)^{-1} = \left(\frac{1}{n} \sum_{j=1}^n \ln X_{j,n} \right)^{-1}.$$

Una generalización trivial es $\bar{F}(x) = Cx^{-\alpha}$, $x \geq u > 0$, donde u es conocida. Si interpretamos esta ecuación como completamente especificada, es decir, $C = u^\alpha$, entonces inmediatamente obtenemos como EMV de α

$$\hat{\alpha}_n = \left(\frac{1}{n} \sum_{j=1}^n \ln \left(\frac{X_{j,n}}{u} \right) \right)^{-1} = \left(\frac{1}{n} \sum_{j=1}^n \ln X_{j,n} - \ln u \right)^{-1}. \quad (3.16)$$

Si bien a menudo no tenemos la información paramétrica precisa que consideramos en estos ejemplos, si suponemos que nos encontramos en el DAM(Φ_α), podemos a su vez suponer que \bar{F} se comporta como una f.d. Pareto por encima de cierto umbral conocido u . Sea

$$K = \text{card} \{i : X_{i,n} > u, \quad i = 1, \dots, n\}. \quad (3.17)$$

Condicionando con el evento $\{K = k\}$, la estimación de máxima verosimilitud de α y C en $\bar{F}(x) = Cx^{-\alpha}$ se reduce maximizar la densidad conjunta de $\{X_{k,n}, \dots, X_{1,n}\}$,

donde

$$f_{X_{1,n}, \dots, X_{k,n}}(x_1, \dots, x_k) = \frac{n!}{(n-k)!} (1 - Cx_k^{-\alpha})^{n-k} (C^k \alpha^k) \prod_{i=1}^k x_i^{-(\alpha+1)}, \quad u < x_k < \dots < x_1.$$

Así, los EMV condicionales que se obtienen son

$$\begin{aligned} \hat{\alpha}_{k,n}^{(H)} &= \left(\frac{1}{k} \sum_{j=1}^k \ln \left(\frac{X_{j,n}}{X_{k,n}} \right) \right)^{-1} = \left(\frac{1}{k} \sum_{j=1}^k \ln X_{j,n} - \ln X_{k,n} \right)^{-1} \\ \hat{C}_{k,n} &= \frac{k}{n} X_{k,n}^{\hat{\alpha}_{k,n}^{(H)}}. \end{aligned}$$

De esta manera, vemos que el Estimador de Hill tiene la misma forma que el EMV en el modelo exacto de (3.16), pero reemplazando la cantidad determinista u por el umbral aleatorio $X_{k,n}$, para K definida como en (3.17).

Teorema 3.4.1 (Propiedades del Estimador de Hill) *Supongamos que (X_n) es estrictamente estacionario con distribución marginal F que satisface, para algún $\alpha > 0$ y $L \in \mathcal{R}_0$,*

$$\overline{F}(x) = P[X > x] = x^{-\alpha} L(x), \quad x > 0.$$

Sea $\hat{\alpha}^{(H)} = \hat{\alpha}_{k,n}^{(H)}$ el estimador de Hill.

(a) (Consistencia débil) Si $k \rightarrow \infty$, $k/n \rightarrow 0$ cuando $n \rightarrow \infty$, entonces

$$\hat{\alpha}^{(H)} \xrightarrow{P} \alpha.$$

(b) (Consistencia fuerte) Si $k/n \rightarrow 0$, $k/\ln \ln n \rightarrow \infty$ cuando $n \rightarrow \infty$ y (X_n)

es una sucesión i.i.d., entonces

$$\hat{\alpha}^{(H)} \xrightarrow{c.s.} \alpha.$$

Observación 3.4.2 *Para obtener normalidad asintótica se necesita hacer suposiciones de variación regular de segundo orden para \bar{F} . Estas condiciones no pueden ser verificadas en la práctica. Generalmente, para la elección de k se hace una gráfica de Hill*

$$\left\{ \left(k, \hat{\alpha}_{k,n}^{(H)} \right) : k = 2, \dots, n \right\},$$

buscando señales de estabilidad.

Observación 3.4.3 *El estimador de Hill es muy sensible con respecto a la dependencia entre los datos. Si se sabe que los datos son dependientes, deben aplicarse técnicas especiales para poder obtener residuales y aplicar entonces el estimador de Hill a los residuales.*

3.4.3 El Estimador Deckers-Einmahl-de Haan (EDEDH) para

$$\xi \in \mathbb{R}$$

Una desventaja del Estimador de Hill es que está diseñado para $F \in DAM(\Phi_\alpha)$ (es decir, para $F \in DAM(H_\xi), \xi > 0$). Ya se mencionó que esta clase de modelos sirve para una gran variedad de aplicaciones en finanzas y seguros. Deckers, Einmahl y de Haan hicieron una extensión del Estimador de Hill, de manera que pueda cubrirse toda la clase $H_\xi, \xi \in \mathbb{R}$. En el Teorema 1.3.15 vimos que, para $F \in DAM(H_\xi), \xi < 0$, el extremo derecho x_F de F es finito; por simplicidad, supondremos que es positivo. Puesto que todas las condiciones para el dominio de atracción del máximo de H_ξ tienen alguna relación con la variación regular, éstas pueden reformularse para obtener

estimadores para todo $\xi \in \mathbb{R}$. El Estimador Deckers-Einmahl-de Haan está dado por:

$$\widehat{\xi} = 1 + H_n^{(1)} + \frac{1}{2} \left(\frac{(H_n^{(1)})^2}{H_n^{(2)}} - 1 \right)^{-1},$$

donde

$$H_n^{(1)} = \frac{1}{k} \sum_{j=1}^k (\ln X_{j,n} - \ln X_{k+1,n})$$

es el recíproco del estimador de Hill (modulo un cambio poco importante de k a $k+1$)

y

$$H_n^{(2)} = \frac{1}{k} \sum_{j=1}^k (\ln X_{j,n} - \ln X_{k+1,n})^2.$$

Puesto que $H_n^{(1)}$ y $H_n^{(2)}$ pueden interpretarse como momentos empíricos, $\widehat{\xi}$ también se conoce como el estimador de momentos de ξ .

En esta sección hemos expuesto varios métodos para estimar el parámetro de forma ξ de las f.d. en el dominio de atracción del máximo de una distribución VEG. Si bien el parámetro de forma ξ es el más importante, también es necesario estimar los parámetros de localización y escala. Las propiedades de los estimadores dados en esta sección dependen fuertemente de la conducta de orden superior de la cola \overline{F} de la distribución subyacente. Desgraciadamente, esta conducta no puede verificarse en la práctica.

Se ha visto que la determinación del número k de estadísticas de orden superior que se usarán es un punto delicado del planteamiento. En general es conveniente hacer una gráfica de Hill (o relacionada), buscando una tendencia hacia la estabilidad (es decir, una línea medianamente horizontal) a través de varios valores de k . Es preferible elegir a k dentro de esa región.

Es muy importante recalcar que nunca debe hacerse una sola estimación. Siempre deben calcularse y graficarse los estimadores de las cantidades relevantes (de los cuantiles, digamos) en un amplio rango de valores de k .

Para los datos de (rea)seguros normalmente es fácil notar que se trata de datos de cola pesada, cuya cola tiene una conducta muy similar a la Pareto desde valores de x relativamente bajos. Esto no es tan evidente en el mundo financiero, debido a que los datos financieros presentan una estructura de dependencia más complicada. Por esta razón, cualquier análisis que se realice debe apoyarse en técnicas exploratorias para el análisis de los datos. En el siguiente Capítulo veremos algunas de estas técnicas.

Hemos visto dos modelos matemáticos para la obtención de medidas de riesgo para valores extremos: el modelo de Picos sobre el Umbral y el modelo de Máximos de Bloque. Al utilizar un modelo matemático intentamos utilizar la herramienta correcta para deducir conclusiones con base científica de los datos. Sin embargo, también es muy importante que los datos se reporten de manera correcta y que las respuestas que intentamos obtener del modelo siempre tomen en cuenta las suposiciones en que este modelo está basado. Contamos ya con la TVE, que es una base teórica confiable. Sin embargo, el mundo real nos informa acerca de los eventos extremos a través de datos estadísticos, tales como grandes reclamos de seguros, niveles de agua de los ríos, grandes caídas (o incrementos) de los valores del mercado accionario durante un período determinado de tiempo, niveles extremos de indicadores ambientales como el ozono o el monóxido de carbono, la velocidad del viento en algún lugar, la altura de las olas durante una tormenta o los valores máximo y mínimo de rendimiento de un portafolio.

Siempre existirán varios niveles de incertidumbre al realizar análisis de datos. El primer nivel es la incertidumbre en los parámetros. Incluso cuando se cuenta con datos abundantes y de buena calidad, e incluso con un buen modelo, los estimadores de los parámetros están sujetos a un error estándar. Obtenemos un rango de estimaciones de parámetros que son compatibles con las suposiciones del modelo. Pero la inferencia es sensible a pequeños cambios de los parámetros, particularmente del parámetro de forma ξ .

También está presente la incertidumbre del modelo: podemos tener buenos datos pero un mal modelo. Lo que sabemos es que al usar los métodos de valores extremos al menos estamos trabajando con una buena clase de modelos, pero sólo son aplicables por encima de umbrales altos y nosotros debemos elegir dónde fijar el umbral. Si fijamos un umbral demasiado alto tendremos pocos datos, lo que implicará una mayor incertidumbre de parámetros. Si fijamos un umbral demasiado bajo perderemos la justificación teórica para el modelo. En general, la inferencia es muy sensible a la elección del umbral, aunque éste no es siempre el caso.

Tan grave como la incertidumbre de parámetros y de modelo es la incertidumbre en los datos. En cierto sentido nunca tendremos suficientes datos para el análisis de valores extremos. Mientras que una muestra de 1000 puntos puede resultar bastante amplia para hacer inferencias acerca de la media de una distribución mediante el uso del Teorema del Límite Central, nuestra inferencia sobre la cola de la distribución es menos segura, puesto que son pocos los puntos que entran en la región de la cola. Además, la inferencia es muy sensible al valor de las mayores pérdidas observadas, por lo que la introducción de nuevas pérdidas extremas al conjunto de datos puede tener

un impacto sustancial. Por ello, cabe aún la utilización de escenarios “extremos”, puesto que los datos históricos pueden enriquecerse con pérdidas hipotéticas para investigar las consecuencias de eventos extremos que aún no han sido observados.

Capítulo 4

Análisis de Datos

En este Capítulo se pondrán en práctica los dos Modelos de análisis descritos en el Capítulo 3: el Modelo de Máximos de Bloque (MDB) y el Modelo de Picos Sobre el Umbral (POT). Para ello, se llevará a cabo el análisis de tres casos: el primero (Sección 4.2), correspondiente a datos de incendios industriales en Dinamarca; el segundo (Sección 4.3), donde se examina la variación (negativa) en el precio diario de la acción de BMW; y el tercero (4.4), correspondiente a las pérdidas en el tipo de cambio *FIX* de pesos mexicanos a dólares americanos (MXN/USD). Para cada uno de estos casos se realizan cuatro actividades básicas: primero, un análisis preliminar —como se explica en la Sección 4.1— basado en gráficas, entre las que se encuentran la gráfica de la serie de tiempo, el histograma, el Boxplot, el QQ-Plot y la gráfica de la función media de exceso; estas gráficas son una herramienta muy útil, que nos proporciona información acerca de nuestros datos y en particular de su comportamiento en el área de la cola. Enseguida se realiza el análisis de Máximos de Bloque para distintos tamaños de bloque, obteniendo los correspondientes estimadores del VaR_q para $q =$

0.999; después se lleva a cabo el análisis de Picos Sobre el Umbral para los mismos datos, utilizando varios umbrales, donde se toma como umbral inicial el indicado por la gráfica de la función muestral media de exceso. También en este caso se calculan los estimadores del VaR_q y se obtienen además los correspondientes a la Esperanza Condicional de la Cola (ES_q). Por último, se hace una breve comparación de los resultados obtenidos para el estimador de VaR_q por cada uno de los dos métodos y se examina cuál de ellos se apega más a los datos históricos de la serie. Cada uno de estos cuatro pasos está ampliamente ilustrado con gráficas y tablas para los tres conjuntos de datos. El primer caso es particularmente instructivo. En la Sección 4.3 se da además un breve resumen de los resultados que justifican la aplicación del método de Máximos de Bloque a series estacionarias.

Todas las gráficas de este Capítulo están realizadas en S-Plus y, a excepción de las gráficas correspondientes a la Esperanza Condicional de la Cola (ES), fueron hechas utilizando las funciones del *software* EVIS (*Extreme Values in S-Plus*), creado en ETH-Zurich por el Dr. Alexander McNeil. Este *software* está disponible en

<http://www.math.ethz.ch/~mcneil/>

La gráfica en la que se obtiene el estimador de ES y algunas de las funciones necesarias para la realización de las Tablas (específicamente las que evalúan niveles de retorno y los transforman a cuantiles, o viceversa, y la que calcula el índice extremo dado el número de excesos del umbral) son de diseño propio. Los datos de los primeros dos casos (incendios industriales en Dinamarca y *log-returns* de la acción BMW) forman parte de los conjuntos de datos incluidos en EVIS, mientras que los

datos del tipo de cambio FIX se obtuvieron de la página *web* del Banco de México: www.banxico.org.mx .

4.1 Análisis Exploratorio de Datos

Antes de enfrascarse en un análisis estadístico de los datos, es muy importante examinar los datos mediante su observación. En nuestra era, con los continuos avances de la computación, la exploración gráfica de los datos se está volviendo cada vez más importante. Y es que cada conjunto de datos es único, por lo que debe haber una persona que analice los datos y considere su significado a cada paso del análisis. El proceso no puede y no debería automatizarse por completo.

4.1.1 Para Empezar

Dado un conjunto de datos, el análisis normalmente comienza con la gráfica de la serie de tiempo, un Boxplot, un histograma y una gráfica de la distribución empírica F . A continuación se da una breve descripción de cada una de estas herramientas.

Gráfica de la Serie de Tiempo

Nos permite identificar las pérdidas más extremas y sus fechas aproximadas de ocurrencia. También podemos ver si hay evidencia de que las pérdidas grandes se encuentren agrupadas, lo que podría poner en duda la suposición de que los datos son independientes e idénticamente distribuidos.

Boxplot

Los Boxplots son gráficas de las localizaciones de las siguientes características básicas de la distribución de los datos: la mediana, los cuartiles superior e inferior, así como la extensión de los datos más allá de los cuartiles.

En esta gráfica es muy fácil detectar cuántas observaciones extremas se tienen y cuáles son sus valores. Por ello, los Boxplots nos permiten captar rápidamente la localización, la escala (ancho) y la forma aproximada de la distribución de los datos. Por ejemplo, si los cuartiles superior e inferior están aproximadamente a la misma distancia de la mediana, entonces la distribución de los datos es relativamente simétrica con respecto al centro.

Histograma

El Boxplot es una herramienta sencilla y poderosa para la visualización de los datos, pero conlleva una reducción importante de los datos. A menudo se quiere que los datos no se reduzcan tanto y a la vez obtener más detalles acerca de la forma de la distribución; es entonces cuando se hace un histograma. El histograma muestra el amplio rango de los datos, al tiempo que nos permite ver si hay evidencia de simetría en los datos.

Por otra parte, si el histograma se realiza en escala logarítmica, esta herramienta nos permite verificar si los datos tuvieran una cola derecha con distribución lognormal, lo cual se evidenciaría en la gráfica a través de una conocida forma de campana.

4.1.2 Gráficas de Probabilidad y de Cuantiles (PP-Plot y QQ-Plot)

Si se tienen los datos i.i.d. X, X_1, \dots, X_n , comenzamos por definir la muestra ordenada (los estadísticos de orden) $X_{n,n} \leq \dots \leq X_{1,n}$.

Lema 4.1.1 Sean X_1, \dots, X_n v.a.i.i.d. con f.d. F . Sean U_1, \dots, U_n v.a.i.i.d. con distribución uniforme en $(0, 1)$ y denotemos $U_{n,n} < \dots < U_{1,n}$ los correspondientes estadísticos de orden. Entonces es cierto lo siguiente:

1. $F^{-1}(U_1) \stackrel{d}{=} X_1$.

2. Para todo $n \in \mathbb{N}$

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

3. La v.a. $F(X_1)$ tiene distribución uniforme en $(0, 1)$ si y sólo si F es una función continua.

PP-Plot

La transformación de cuantiles del lema anterior es la base teórica sobre la cual yace la gráfica de probabilidad. Esta transformación implica que, para F continua, las v.a.s $U_i = F(X_i)$ para $i = 1, \dots, n$ son i.i.d. con distribución uniforme en $(0, 1)$. Además,

$$(F(X_{k,n}))_{k=1, \dots, n} \stackrel{d}{=} (U_{k,n})_{k=1, \dots, n}.$$

De aquí, se sigue que

$$E[F(X_{k,n})] = \frac{n-k+1}{n+1}, \quad k = 1, \dots, n.$$

Nótese también que, si denotamos mediante F_n a la distribución empírica, se tiene que $F_n(X_{k,n}) = (n - k + 1) / n$.

Llamamos gráfica de probabilidad o PP-Plot a la gráfica

$$\left\{ \left(F(X_{k,n}), \frac{n - k + 1}{n + 1} \right) : k = 1, \dots, n \right\}. \quad (4.1)$$

El Teorema Glivenko-Cantelli (ver Apéndice C) justifica la linealidad aproximada de esta gráfica.

QQ-Plot

Una gráfica más común que el PP-Plot es la gráfica de cuantiles (o cuantil contra cuantil), mejor conocida como QQ-Plot:

$$\left\{ \left(X_{k,n}, F^{-1} \left(\frac{n - k + 1}{n + 1} \right) \right) : k = 1, \dots, n \right\} \quad (4.2)$$

Esto es, los cuantiles de la función de distribución empírica se grafican en el eje- x contra los cuantiles de la distribución que suponemos que siguen los datos.

De la misma manera que para el PP-Plot, la linealidad aproximada de esta gráfica está justificada por el Teorema Glivenko-Cantelli.

Existen algunas variantes del QQ-Plot dado en (4.2), las cuales son de la siguiente forma:

$$\{(X_{k,n}, F^{-1}(p_{k,n})) : k = 1, \dots, n\},$$

donde $p_{k,n}$ es cierta posición de graficación; casi siempre tomaremos $p_{k,n} = \frac{n - k + 0.5}{n}$, lo cual permite cierta corrección de continuidad.

Ejemplo 4.1.2 Para una distribución Gumbel

$$\Lambda(x) = \exp \{-e^{-x}\}, \quad x \in \mathbb{R},$$

este método se aplica fácilmente. Supongamos que queremos determinar si la muestra X_1, X_2, \dots, X_n proviene de Λ . Para ello, tomamos la muestra ordenada y graficamos $X_{k,n}$ (más precisamente la k -ésima observación más grande $x_{k,n}$) contra $\Lambda^{-}(p_{k,n}) = -\ln(-\ln p_{k,n})$, donde $p_{k,n}$ es una posición de graficación como la mencionada anteriormente. Si la distribución Gumbel se ajusta bien a nuestros datos, entonces este QQ-Plot debería ser aproximadamente lineal.

Propiedades del QQ-Plot. El QQ-Plot es una herramienta muy útil para:

1. **Comparación de Distribuciones:** Si los datos se generaron mediante una muestra aleatoria de la distribución de referencia, la gráfica debería verse prácticamente lineal. Esto también es válido si los datos provienen de una transformación lineal de la distribución.
2. **“Outliers”:** Si uno o varios de los datos están contaminados por grandes errores o si son por alguna razón muy diferentes en cuanto a sus valores del resto de los puntos -siendo que estos últimos tienen una distribución que se aproxima a la distribución de referencia- podemos detectar fácilmente a los puntos “extraños” mediante la gráfica.
3. **Localización y escala:** Debido a que el aplicar una transformación lineal a una distribución simplemente transforma su gráfica con esta misma transformación lineal, podemos estimar gráficamente (a través de las intersecciones y la pendiente) los parámetros de escala y localización para una muestra de datos, si suponemos que los datos provienen de la distribución de referencia.

4. **Forma:** Podemos utilizar la gráfica para saber si los datos provienen de una distribución con cola más pesada o con cola más ligera que la de la distribución de referencia. Si la distribución empírica tiene cola derecha más pesada (tendencia a presentar mayores valores) que la distribución de referencia, entonces la gráfica mostrará tendencia hacia la concavidad con respecto a la forma ideal; mientras que si la distribución empírica tiene cola derecha menos pesada que la distribución de referencia, habrá tendencia hacia la convexidad. Para la cola izquierda, el comportamiento es el opuesto.

Advertencia sobre el QQ-Plot

Debemos mencionar, sin embargo, que, en ocasiones, incluso los datos generados mediante determinada distribución pueden mostrar desviaciones de la conducta típica para ese tipo de distribución (digamos, por ejemplo, datos generados con la distribución exponencial). En general, mientras más datos se tengan, más podemos confiar en el mensaje que nos da el QQ-Plot.

4.1.3 La Función Media de Exceso

La gráfica de la función media de exceso es una herramienta útil para distinguir entre distribuciones de cola pesada y distribuciones de cola corta. Recordemos la definición de la función media de exceso:

Definición 4.1.3 Sea X una v.a. con extremo derecho x_F ; entonces

$$e(u) = E[X - u \mid X > u], \quad 0 \leq u < x_F \quad (4.3)$$

es la función media de exceso de X . También decimos, para cada u fijo, que la cantidad $e(u)$ es el exceso medio sobre el umbral u .

La función media de exceso describe el exceso esperado de un umbral u dado que sabemos que éste ha sido excedido.

Podemos hacer un análisis gráfico de la conducta de la cola utilizando la función muestral media de exceso $e_n(u)$, que es un estimador empírico de la función media de exceso dada en (4.3). La función muestral media de exceso se define

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)_+}{\sum_{i=1}^n \mathbf{1}_{\{X_i > u\}}}, \quad (4.4)$$

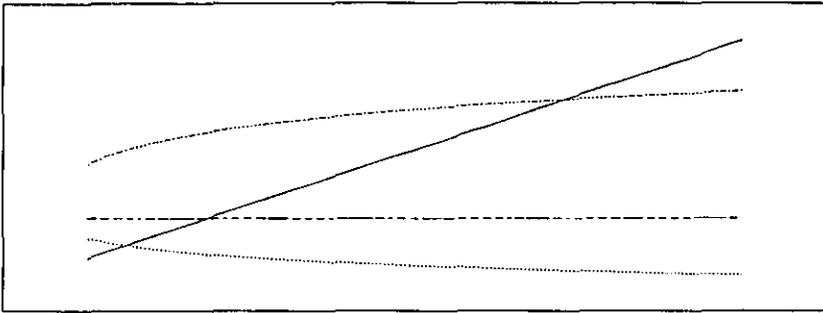
es decir, es la suma de los excesos por encima del umbral u dividida entre el número de puntos que exceden este umbral.

Supongamos que tenemos la muestra X_1, \dots, X_n . Como de costumbre, ordenamos esta muestra para obtener los estadísticos de orden $X_{1,n} > \dots > X_{n,n}$. Se procede entonces a hacer la gráfica de la función muestral media de exceso, dada por

$$\{(u, e_n(u)) \mid X_{n,n} < u < X_{1,n}\}. \quad (4.5)$$

La interpretación de $e(u)$ es la siguiente: si los puntos muestran tendencia hacia arriba, entonces hay señales de conducta de cola pesada. Los datos con distribución exponencial darían una recta aproximadamente horizontal; mientras que los datos de una distribución de cola corta mostrarían una tendencia hacia abajo. En particular, si la gráfica empírica (4.5) parece seguir una línea razonablemente recta con pendiente positiva por encima de un cierto valor de u , entonces podemos suponer que los datos siguen, en el área de la cola y por encima del valor del umbral u , una Distribución Pareto Generalizada (DPG) con parámetro de forma ξ positivo.

Nótese que, puesto que cerca de la mayor observación $X_{1,n}$ son muy pocos los datos que forman parte del cálculo de $e_n(u)$, existe un aumento en la variabilidad (sensibilidad a cambios en los datos) cerca del extremo superior de la gráfica. En ocasiones conviene detener el proceso de graficación en $X_{4,n}$ o $X_{3,n}$ para evitar esta conducta errática.



Función media de exceso teórica para distribuciones Pareto (sólida), Normal (puntos), Exponencial (guiones) y Lognormal (mezcla).

4.1.4 Observaciones sobre el Análisis Exploratorio de Datos

Las técnicas gráficas de exploración de datos que se expusieron en esta sección parten de la suposición de que los datos son independientes e idénticamente distribuidos. Si bien a menudo es posible suponer que los datos en el área de seguros representan observaciones i.i.d., éste no es el caso en general cuando se trata de datos financieros: es bien sabido que, aunque a primera vista pudieran parecer independientes con baja correlación, si se toman las raíces cuadradas o los valores absolutos de los datos, se podrá evidenciar la presencia de una alta correlación, por lo que hay una tendencia a que los valores grandes (aunque muy posiblemente de signos contrarios, es decir,

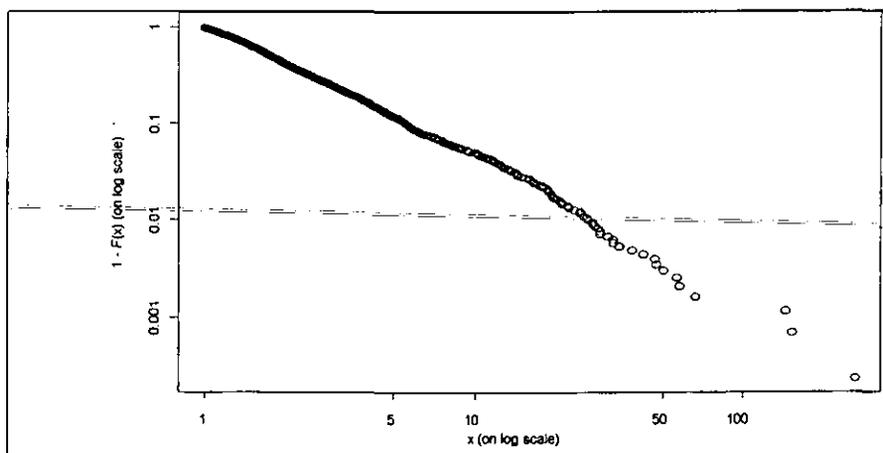
pérdidas y ganancias) ocurran juntos. Sin embargo, podemos suponer que la serie de tiempo subyacente es estacionaria, con lo que muchos resultados para límites (como la Ley Fuerte de los Grandes Números) siguen siendo válidos en general. Esto nos permite interpretar las gráficas exploratorias empíricas que hemos descrito (el histograma, la función media de exceso muestral, el QQ-Plot, etc.) como cercanas a sus contrapartes teóricas. También existen varios métodos estadísticos para eliminar las tendencias que pudieran mostrar los datos, como sería agregar una variable de tendencia a los datos, sacar promedios o diferenciar. Mediante estos métodos se intenta obtener residuales i.i.d. a los que puedan aplicarse los métodos gráficos mencionados en esta sección.

4.2 Caso 1: Pérdidas Industriales por Incendios en Dinamarca

La base de datos comprende 2492 pérdidas y puede considerarse esencialmente como todas las pérdidas danesas por incendios que exceden 1 millón de coronas danesas (DKK), en el período comprendido entre enero de 1980 y diciembre de 1990, más algunas pérdidas más pequeñas (por debajo de 1 millón de DKK). La cantidad reflejada corresponde a la pérdida total del evento mencionado e incluye daños a edificios, daños a los contenidos de los edificios y pérdida de utilidades. Los datos se han ajustado por inflación para reflejar valores de 1985.

Restringiremos nuestra atención a las 2156 pérdidas que exceden un millón de

DKK, puesto que para nuestros propósitos es menos importante que el modelo explique las pérdidas más pequeñas; si nos interesaran también las pérdidas menores podríamos, en todo caso, utilizar una distribución mezclada, de manera que un modelo pudiera aplicarse a la cola y otro al cuerpo principal de los datos. Hacemos esta acalaración porque, además, típicamente en seguros tendremos datos históricos de pérdidas que exceden cierta cantidad (la cual se conoce como desplazamiento). Es prácticamente imposible obtener datos de todas las pérdidas y, de cualquier manera, las pérdidas pequeñas son menos importantes. Las aseguradoras en general deben enfrentarse a pérdidas mayores, pues los asegurados se encargan ellos mismos de las pérdidas pequeñas y pueden incluso no reportarlas. En el caso de nuestros datos, tenemos que el desplazamiento es 1. Esto puede observarse en la siguiente gráfica, que corresponde a la cola de la función de distribución empírica de los datos daneses, usando escala logarítmica en ambos ejes.



Estimador empírico de la cola \bar{F} de los datos daneses, en doble escala logarítmica.

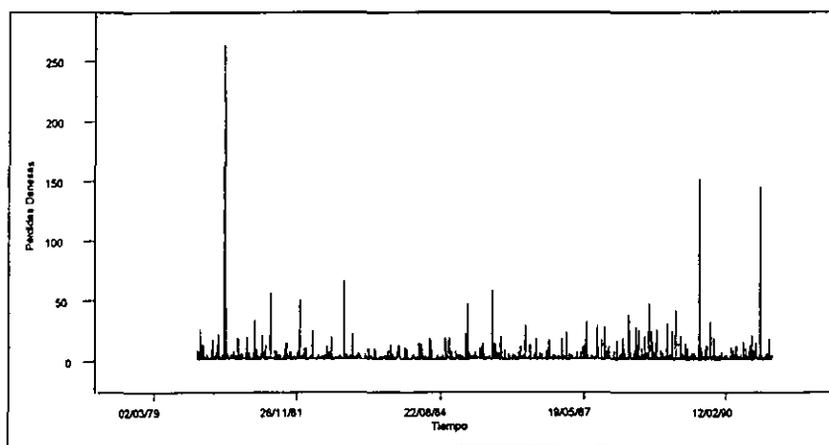
Supongamos que debemos poner precio a la capa que va de 50 a 200 (millones de

DKK). En este intervalo sólo tenemos seis pérdidas observadas, por lo que si ajustamos una distribución a todo el conjunto de datos, ésta puede no ajustarse particularmente bien a esta área de la cola donde los datos son escasos. Debemos entonces obtener un buen estimador del área de la cola para esta distribución. Para ello, usaremos los métodos de TVE explicados en el Capítulo anterior.

4.2.1 Análisis Preliminar de Datos de Pérdidas por Incendios en Dinamarca

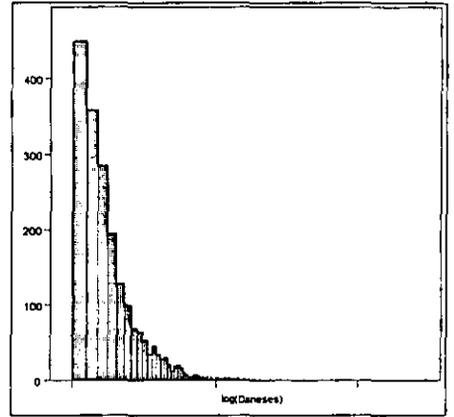
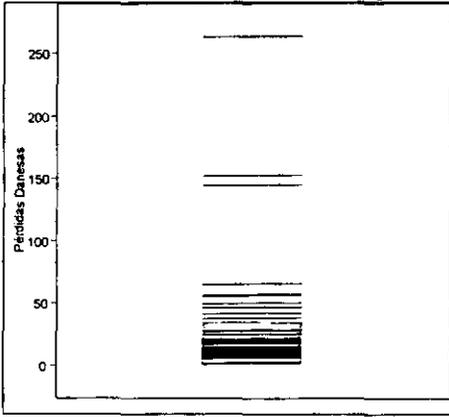
Al inicio de este Capítulo mencionamos que conviene realizar un análisis gráfico de los datos antes de llevar a cabo el análisis estadístico completo. El análisis gráfico nos puede dar una buena idea acerca del comportamiento de los datos, por lo cual es un buen punto de partida para decidir si la aplicación de cierto modelo es o no plausible.

Comenzaremos pues por la gráfica de la serie de tiempo de los datos daneses.



Serie de tiempo de 2157 datos de incendios daneses entre 1980 y 1990.

Con la gráfica de la serie de tiempo podemos notar que las pérdidas más grandes ocurrieron en los años 1980, 1989 y 1990, respectivamente. La gráfica no muestra agrupamiento de las pérdidas grandes, por lo que es creíble la suposición de independencia de los datos. Veremos enseguida un Boxplot y un histograma, en escala logarítmica, de los datos. Se realiza además una Tabla con los principales datos estadísticos correspondientes a las pérdidas danesas.



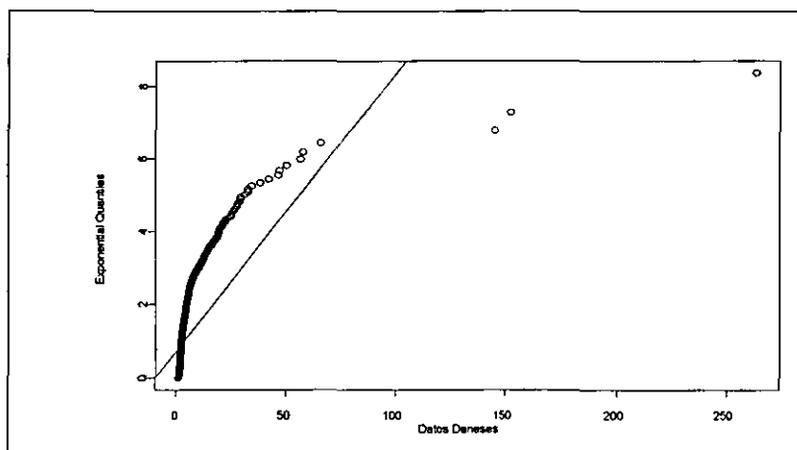
Boxplot de las pérdidas danesas.

Histograma en escala logarítmica.

Mínimo	1.000
Primer cuartil	1.321
Media	3.385
Mediana	1.778
Tercer Cuartil	2.967
Máximo	263.250
Desv. Estándar	8.507
Total datos	2167

Resumen Estadístico Daneses.

El Boxplot nos muestra que casi todas las pérdidas de nuestra base de datos están por debajo de los 40 millones de DKK, con sólo 6 pérdidas entre los 50 y los 250 millones de DKK. El resumen estadístico de los datos daneses nos confirma que la mayor parte de nuestros datos corresponden a pérdidas entre 1 y 4 millones de DKK, al mostrar que el tercer cuartil está cerca de los 3 millones de DKK y que la media empírica está apenas por encima de esta cantidad. Por su parte, el histograma nos permite ver el amplio rango de los datos; su forma parece indicar que los datos podrían provenir de una distribución cuya cola derecha es pesada. Para tratar de corroborar esta información, investigamos ahora la posibilidad de que nuestros datos provengan de una distribución exponencial, es decir, de una distribución de cola mediana. Para ello, procedemos a hacer un QQ-Plot:



QQ-Plot de Datos Daneses

Recordemos que si los datos provinieran de una distribución exponencial, el QQ-Plot debería ser prácticamente lineal. En este caso, tenemos que los datos muestran una desviación cóncava con respecto a la forma ideal, lo que indica que nuestra dis-

tribución tiene la cola más pesada que una distribución exponencial. Cuando se expusieron las propiedades del QQ-Plot mencionamos que, en ocasiones, incluso si los datos se generan de una distribución exponencial, el QQ-Plot puede indicar que los datos se alejan de la conducta exponencial típica. En general, mientras más datos se tengan, más claro será el mensaje que nos envíe el QQ-Plot. En este análisis contamos con más de 2000 datos, de manera que parece razonable decir que la cola de nuestros datos es más pesada que la de una distribución exponencial.

4.2.2 Análisis de Máximos de Bloque para Datos Daneses

Procedemos ahora a realizar el análisis de máximos de bloque para los datos de pérdidas por incendios en Dinamarca. Lo primero que debemos decidir es, pues, el tamaño de los bloques. Queremos que los bloques sean bastante grandes como para que las observaciones puedan tomarse como efectivamente independientes, pero también queremos tener suficientes bloques como para poder realizar el análisis. El análisis exploratorio que realizamos de los datos nos ha indicado que la distribución subyacente tiene la cola más pesada que una distribución exponencial; esto nos hace pensar que esta distribución podría estar en el Dominio de Atracción del Máximo de la Fréchet.

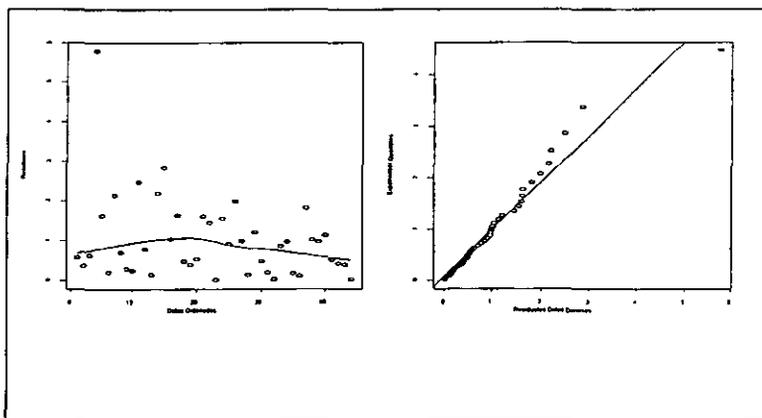
Nuestro primer modelo considerará bloques trimestrales: los 2167 datos se reducen así a 44 máximos trimestrales. Los estimadores de los parámetros, obtenidos por máxima verosimilitud, son en este caso $\hat{\xi} = 0.512$, $\hat{\sigma} = 11.069$ y $\hat{\mu} = 19.047$. De acuerdo con estos parámetros, estamos ajustando los máximos a una distribución

Fréchet con la cola tan pesada que no tiene segundo momento (varianza) finito.

El ajuste del modelo Fréchet a los máximos de bloque se verifica usando residuales ordinarios: para cada máximo de bloque M_n^i , $i = 1, \dots, m$, el residual correspondiente se define mediante

$$W_i = \left(1 + \xi \frac{M_n^i - \mu}{\sigma} \right)^{-1/\xi}. \quad (4.6)$$

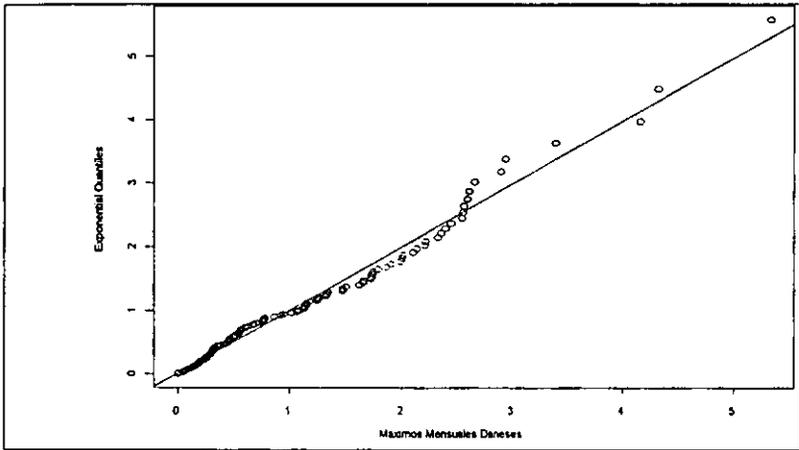
Estos residuales deberían ser i.i.d. con distribución exponencial con parámetro $\lambda = 1$, y esta hipótesis puede verificarse gráficamente mediante un QQ-Plot. De la misma manera, una gráfica de dispersión de los residuales nos permite constatar si el tamaño de los bloques es suficientemente grande como para poder suponer independencia entre ellos.



Dispersión y QQ-Plot de Residuales para Máximos Trimestrales Daneses.

Vemos que en el caso de bloques trimestrales para los datos daneses, tanto la suposición de independencia como la de distribución exponencial de los residuales parecen aceptables, lo que sugiere que los máximos de bloque están adecuadamente modelados por la distribución Fréchet correspondiente.

Modifiquemos ahora el tamaño de los bloques: consideraremos ahora los 132 máximos mensuales. Los estimadores de los parámetros son, en este caso, $\hat{\xi} = 0.623$, $\hat{\sigma} = 5.971$ y $\hat{\mu} = 8.376$. El QQ-Plot de los residuales de estos máximos muestra que también en este caso es aceptable suponer que su comportamiento es exponencial, por lo que el modelo Fréchet es aceptable.



QQ-Plot de los residuales de máximos mensuales para datos daneses.

Si hacemos los bloques más grandes y tomamos máximos semestrales, tenemos únicamente 22 datos de máximos. Los estimadores de los parámetros que se obtienen son $\hat{\xi} = 0.618$, $\hat{\sigma} = 17.333$ y $\hat{\mu} = 26.362$.

Utilizando nuestro modelo de datos mensuales, el estimador del nivel de retorno de 20-meses $R_{16,20}$ (el nivel que será excedido, en promedio, en un mes dentro de cada período de 20 meses) es 59.763. Los estimadores de los niveles de retorno de 20-trimestres y de 20-semestres $R_{49,20}$ y $R_{99,20}$ son 96.348 y 174.135, respectivamente. Es claro que si nuestros modelos para máximos para los diferentes tamaños de bloque son consistentes, entonces el nivel de retorno de 20-semestres debería parecerse al

nivel de retorno de 40-trimestres y al de 120-meses. En la Tabla siguiente se confirma que éste es el caso calculando $R_{16,120}$, $R_{49,40}$ y $R_{99,20}$; estos valores se muestran en el renglón R_{20} . La Tabla también resume los estimadores de los parámetros, con el Error Estándar (E.E.) correspondiente, obtenidos para nuestros modelos de máximos mensuales, trimestrales y semestrales.

	Mes	E.E.	Trimestre	E.E.	Semestre	E.E.
$\approx n$	16		49		99	
m	132		44		22	
ξ	0.623	0.103	0.512	0.133	0.618	0.249
σ	5.971	0.633	11.069	1.807	17.333	4.457
μ	8.376	0.612	19.047	1.855	26.362	4.326
$R_{n,20}$	59.763		96.348		174.135	
R_{20}	187.454		139.426		174.135	
R_{10}	120.97		96.348		111.00	

Estimadores VEG y niveles de retorno para datos daneses.

Puesto que, como hemos mencionado antes,

$$R_{n,k} = x_{(1-1/k)^{1/n}}$$

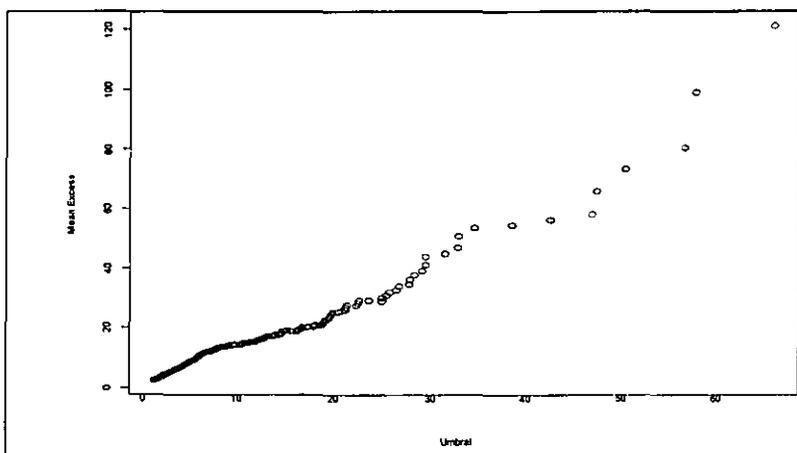
tenemos que el nivel de retorno de 20-semestres corresponde al cuantil 0.99948, el nivel de retorno de 20-trimestres al 0.99895 y el de 20-meses al cuantil 0.99679.

Recíprocamente, el cuantil 0.995 corresponde, aproximadamente, a los niveles de retorno de 13-meses, 5-trimestres o 3-semestres. Los estimadores puntuales son,

respectivamente, 45.03, 44.03 y 47.31.

4.2.3 Análisis POT para Datos Daneses

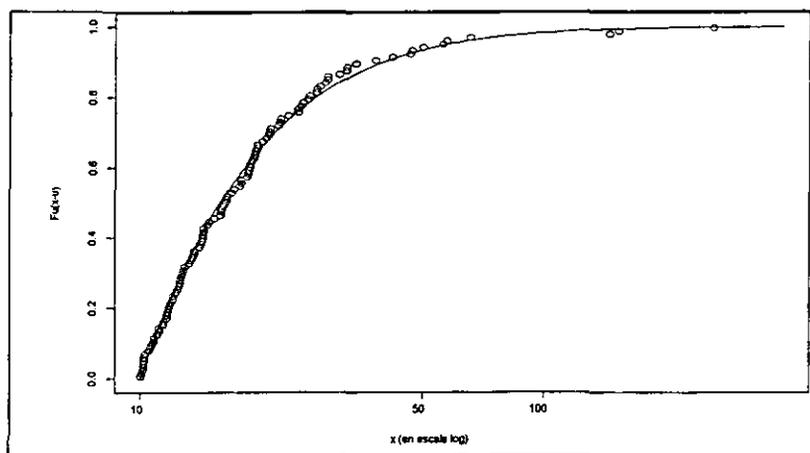
En el Capítulo 3 se dijo que el primer paso para realizar un análisis de datos utilizando el método de Picos Sobre el Umbral (POT) es la elección de un umbral. Como se mencionó en la Sección 4.1, una herramienta que puede darnos una indicación sobre qué umbral podríamos escoger para un primer análisis es la gráfica de la función muestral media de exceso.



Gráfica de la función muestral media de exceso para datos daneses.

La gráfica de la función media de exceso empírica para los datos daneses tiene pendiente positiva, lo que nos indica que los datos provienen de una distribución de cola pesada. Por otro lado, puede evidenciarse que la gráfica se vuelve bastante recta al llegar al umbral $u = 10$, y tal vez de nuevo a partir de $u = 20$. De hecho, la gráfica en general es bastante recta como para sugerir que el ajuste de la Distribución Pareto Generalizada (DPG) podría ser razonable inclusive para todo el conjunto de datos.

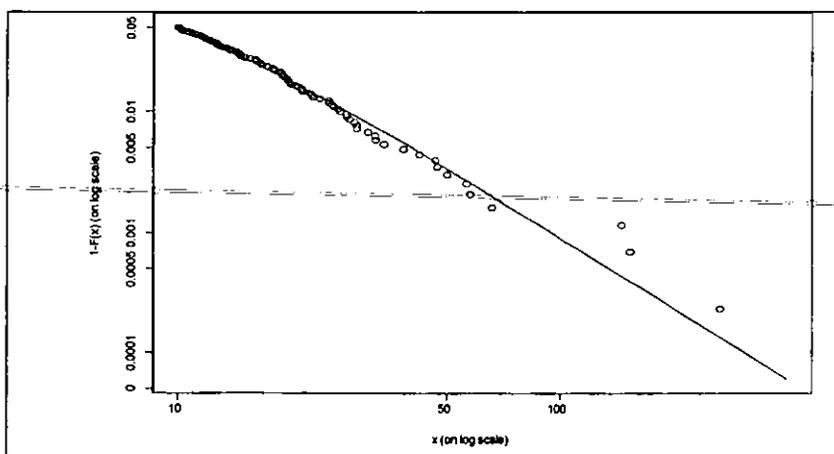
Intentaremos primero realizar el ajuste de la DPG a los datos que exceden el umbral $u = 10$. Para ello, utilizamos la forma de tres parámetros de la DPG, asignándole el valor del umbral al parámetro de localización ν . De esta manera, nuestros $n = 2156$ datos de pérdidas se transforman en $N_u = 109$ excesos del umbral. Con base a estos datos, obtenemos que los estimadores de máxima verosimilitud de los parámetros de forma y de escala son, respectivamente, $\hat{\xi} = 0.497$ y $\hat{\beta} = 6.98$. El valor de $\hat{\xi}$ muestra que los datos son de cola pesada y sugiere que un buen modelo explicativo pudiera tener varianza infinita. En la siguiente Figura se grafica, sobreimpuesta sobre la función de distribución empírica (punteada) de los 109 excesos, la curva correspondiente al modelo de DPG estimado.



La DPG se ajusta a los 109 excesos del umbral 10. Los estimadores de los parámetros son $\hat{\xi} = 0.497$, $\hat{\nu} = 10$ y $\hat{\beta} = 6.98$.

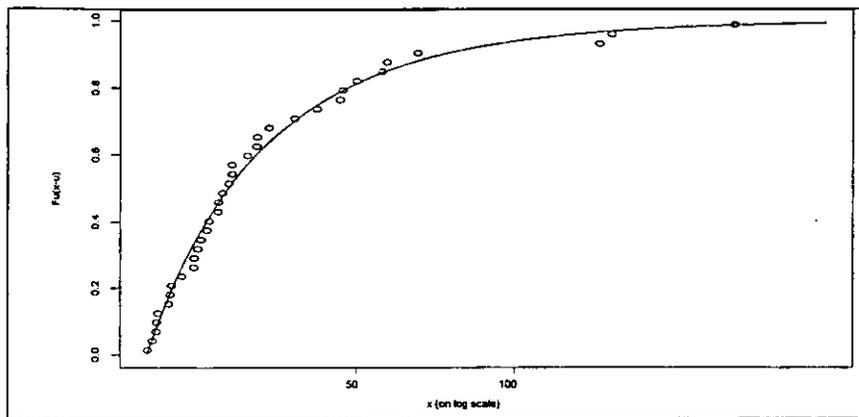
El ajuste obtenido parece bastante bueno a simple vista. Recordemos que éste es un estimador de la distribución de exceso por encima del umbral 10. Como se describió en la Sección 3.2, podemos transformar los parámetros de escala y localización para

obtener un modelo DPG que se ajuste a la zona de la cola (por encima del umbral $u = 10$ dado) de la distribución subyacente. En este caso, el estimador de $F(u)$ que se obtiene por Simulación Histórica (SH) es 0.95 (lo cual corresponde a $\frac{2156-109}{2156}$), de manera que el umbral se encuentra (aproximadamente) en el 95o. percentil muestral. Al combinar esto con nuestro modelo paramétrico para la distribución de exceso se obtiene el estimador de colas que se muestra en la siguiente Figura. En ella, el eje- y indica las probabilidades de la cola $\bar{F}(x) = 1 - F(x)$. La esquina superior izquierda de la gráfica muestra que el umbral de 10 corresponde a una probabilidad de la cola de 0.05, lo cual está de acuerdo con lo estimado por SH. También en este caso los puntos representan las 109 mayores pérdidas, mientras que la curva sólida muestra que la fórmula del estimador de colas permite extrapolar hacia el área en que los datos se vuelven escasos y son, por tanto, una guía poco confiable para la distribución subyacente.

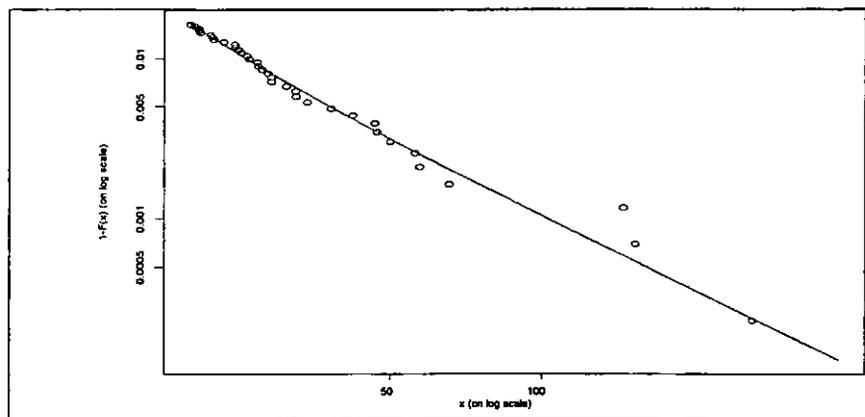


Ajuste de la DPG a la cola de la distribución por encima del umbral 10. Los estimadores de los parámetros son $\hat{\xi} = 0.497$, $\hat{\nu} = -0.859$ y $\hat{\beta} = 1.58$.

Hemos dicho que el análisis nunca debe realizarse para un único umbral. En nuestro caso, la gráfica de la función media de exceso daba cierta indicación de que el modelo DPG podría ser plausible también para el umbral $u = 20$. Escogemos entonces $u = 20$ y obtenemos lo siguiente:

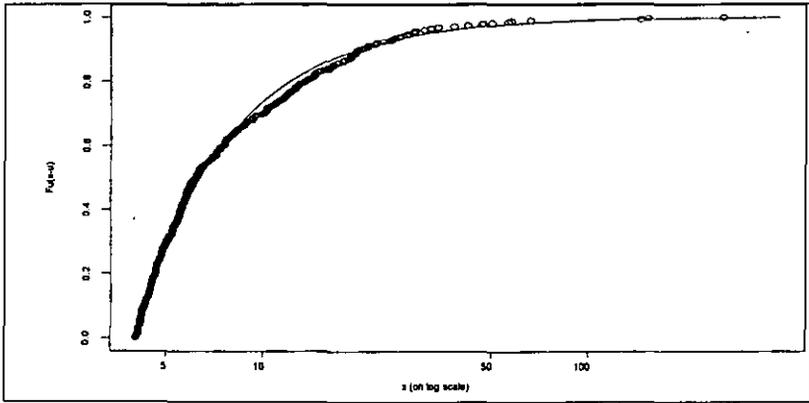


La DPG se ajusta a los 36 excesos del umbral 20. Los estimadores de los parámetros son $\hat{\xi} = 0.684$, $\hat{\nu} = 20$ y $\hat{\beta} = 9.64$.



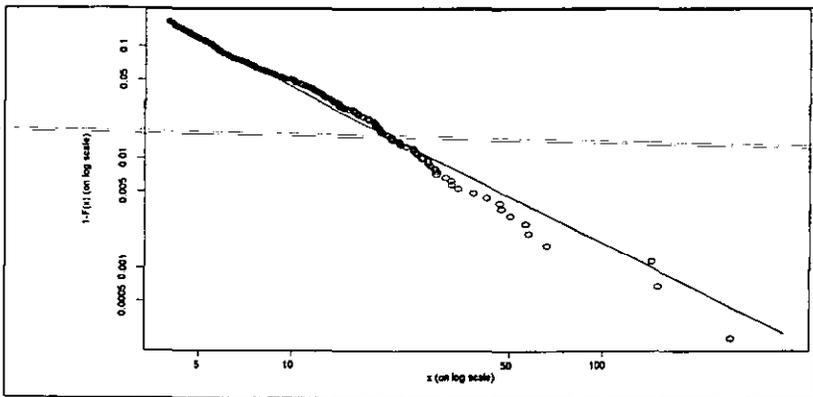
Ajuste de la cola de la DPG a los datos daneses con umbral 20. Los estimadores de los parámetros son $\hat{\xi} = 0.684$, $\hat{\nu} = 6.77$ y $\hat{\beta} = 0.584$.

Ambas Figuras evidencian que el ajuste de la DPG a los datos daneses por encima del umbral $u = 20$ es bastante bueno, si bien la cantidad de datos que se consideran (36) es mucho menor que los 109 datos que tomaban parte en el análisis con $u = 10$. Finalmente, en vista de que había evidencia de que el modelo DPG era razonable para el conjunto completo de los datos, realizamos el ajuste para el umbral $u = 4$:



Ajuste de la DPG a los 362 excesos del umbral 4. Los estimadores de los parámetros

$$\text{son } \hat{\xi} = 0.72, \hat{\nu} = 4 \text{ y } \hat{\beta} = 2.63.$$

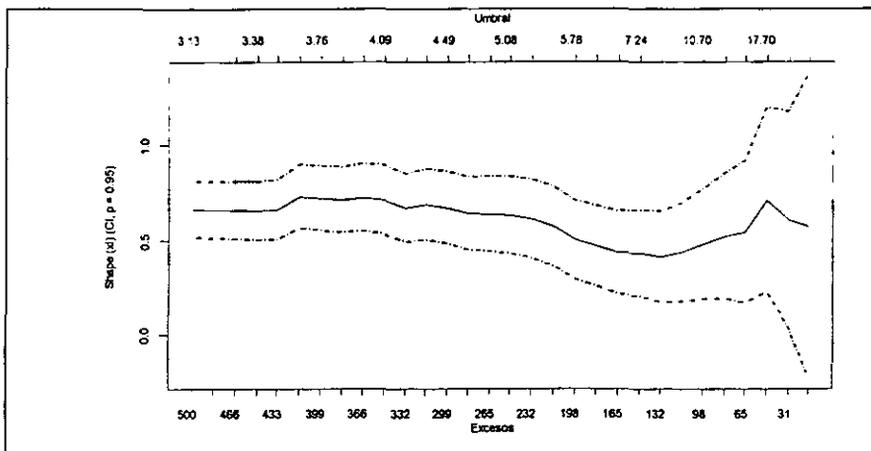


Ajuste de la cola de la DPG a los datos daneses, con umbral 4. Se tiene $\hat{\xi} = 0.72$,

$$\hat{\nu} = 1.35 \text{ y } \hat{\beta} = 0.725.$$

Las dos Figuras anteriores muestran que la DPG se ajusta de manera aceptable a los datos por encima del umbral $u = 4$, si bien el ajuste en el área de la cola no es tan bueno como el obtenido para $u = 10$ o para $u = 20$.

El parámetro crucial para la estimación de altos cuantiles o la determinación de precios de capas altas utilizando el modelo DPG es el parámetro de forma ξ , llamado también índice de la cola. Básicamente, entre mayor sea el valor de ξ , mayores serán los estimadores que obtendremos para los cuantiles. En la siguiente Figura se ajustan los modelos DPG con distintos umbrales, para obtener estimadores de ξ , así como intervalos de confianza asintóticos para los estimadores del parámetro. En el eje- x inferior se marca el número de puntos que exceden el umbral; en el eje- x superior se marca el umbral. El parámetro de forma ξ se grafica en el eje- y . En total se ajustan 30 modelos.



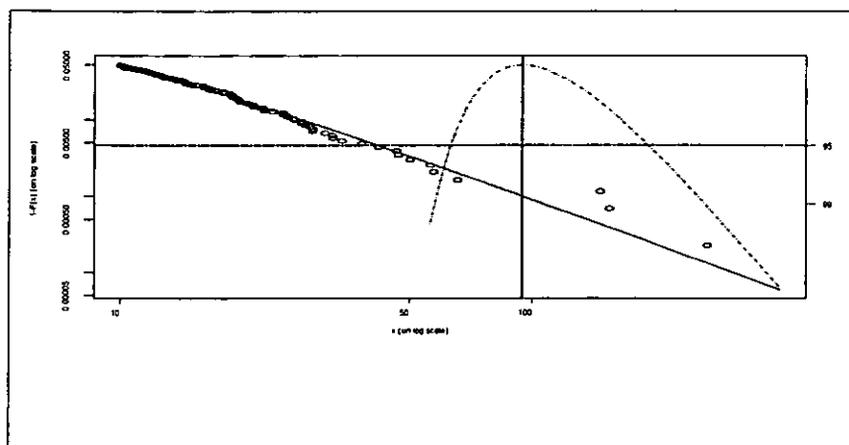
Estabilidad del parámetro de forma ξ conforme se varía el umbral.

Al usar esta Figura para elegir un umbral “óptimo” nos enfrentamos a un intercambio entre sesgo y varianza: puesto que nuestro modelo se basa en un resultado

límite para umbrales altos, si escogemos un umbral demasiado bajo nuestros estimadores podrían estar sesgados porque el resultado límite no puede aplicarse. Por otra parte, si escogemos un umbral demasiado alto, tendremos muy pocos datos y nuestros estimadores estarán sujetos a altos errores estándar. Así, una elección sensible estaría aproximadamente en el centro de la gráfica, tal vez en un umbral entre 4 y 10 en este caso. Sin embargo, la decisión sobre qué umbral es más conveniente elegir depende del uso que se dará a los resultados. Si estamos intentando poner precio al seguro de una alta capa, es posible que queramos ser más bien conservadores y obtener respuestas que sean quizá demasiado altas y no demasiado bajas. En el caso de los datos daneses, es posible que decidiéramos utilizar un umbral menor a 10, tal vez tomar $u = 4$. El modelo de la DPG puede no ajustarse tan bien a los datos por encima de este umbral más bajo como lo hace a los datos por encima de $u = 10$, pero puede resultar más seguro utilizar el umbral más bajo para hacer los cálculos. Por otra parte, puede haber razones de negocios (competencia) para intentar mantener la prima baja; esto pudiera significar que se prefiera basar los cálculos sólo en las pérdidas más grandes, de manera que escogeríamos un umbral mayor.

Si realizamos el análisis fijando el umbral en $u = 10$, se obtiene que el estimador del $\text{VaR}_{0.999}$ para los datos daneses es 94.34. Esto se muestra en la siguiente figura: la línea punteada vertical se intersecta con el estimador de la cola en el punto (94.34, 0.001), lo que nos permite leer el estimador del VaR en el eje- x . La curva punteada es una herramienta que nos permite calcular intervalos de confianza (mediante el método de verosimilitud de perfiles) para el VaR. El eje- y del lado derecho de la gráfica es una escala de confianza (no una escala de cuantiles). La línea punteada horizontal

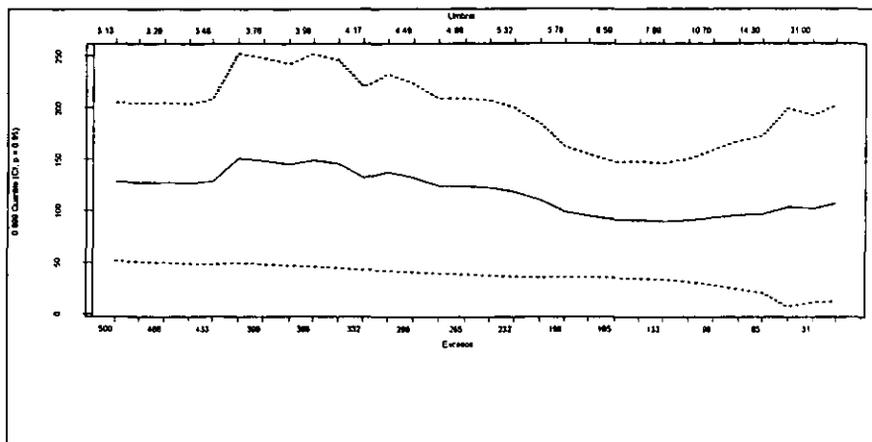
corresponde a una confianza del 95%; las coordenadas en x de los dos puntos donde la curva punteada intersecta a la línea horizontal son los extremos del intervalo de confianza del 95%: (65.24, 188.30). Es importante notar que el intervalo obtenido es asimétrico: esto refleja el problema de estimar altos cuantiles para datos de cola pesada: es más fácil acotar el intervalo por abajo que hacerlo por arriba. Obsérvese que se obtiene un intervalo de confianza del 99% más amplio si se baja la línea horizontal al valor 99 del eje de confianza.



El estimador del $Var_{0.999}$ es 94.34.

En la siguiente Figura se grafican los estimadores del cuantil 0.999 para distintos umbrales; de esta manera, se muestra que los estimadores de los cuantiles dependen de la elección del umbral. Si el modelo es bueno, se esperaría que este cuantil sea excedido por aproximadamente una de cada mil pérdidas que excedan un millón de DKK; estas pérdidas son raras, pero peligrosas para el asegurador. En un conjunto de 2156 pérdidas, es posible que hayamos visto solamente dos o tres pérdidas de esta magnitud, de manera que éste es un problema difícil de estimación de cuantiles que

involucra interpolación —basada en el modelo— en el área de la cola.



Estimadores del cuantil 0.999 conforme se incrementa el umbral (eje x superior) y se disminuyen el número de excesos (eje x inferior).

Tenemos pues que, de la misma manera en que los estimadores de los parámetros (particularmente el estimador del importante parámetro de forma ξ) varían conforme se mueve el umbral, los estimadores de los cuantiles también varían si se aumenta o se disminuye el umbral. Hemos visto que, utilizando el modelo con un umbral $u = 10$, el estimador de $\text{VaR}_{0.999}$ es 94.34. Pero si aumentamos el umbral hasta $u = 20$, el estimador de $\text{VaR}_{0.999}$ se convierte en 102.23, donde el amplísimo intervalo de confianza del 95% está dado por (63.583, 309.536). De la misma manera, si disminuimos el umbral a $u = 4$, el estimador de $\text{VaR}_{0.999}$ aumenta hasta 146.26. Claramente, hay una diferencia considerable entre los estimadores de $\text{VaR}_{0.999}$ cuando $u = 4$ y $u = 10$; si intentamos estimar cuantiles más altos, como $\text{VaR}_{0.9999}$ la diferencia se vuelve aún más marcada. Estimar el cuantil 0.9999 es equivalente a estimar el tamaño de una pérdida entre 10000. En nuestro conjunto de datos es posible que aún no hayamos

visto una pérdida de esta magnitud, de manera que éste es un problema extremadamente difícil, que conlleva la extrapolación del modelo más allá del rango de los datos.

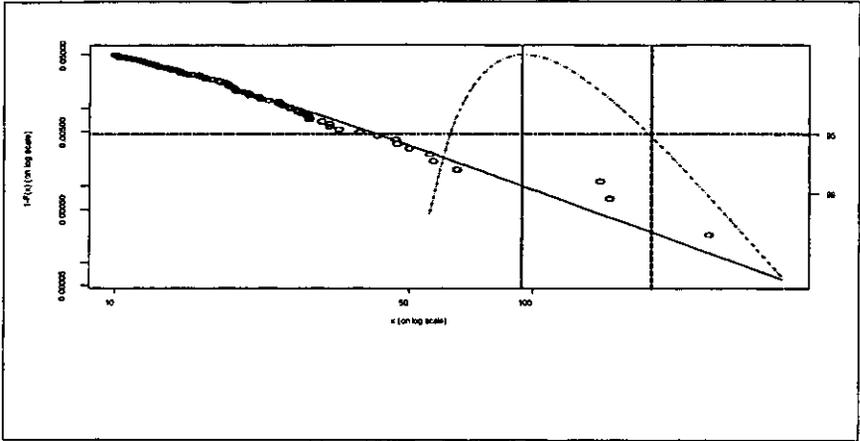
La estimación del $\text{VaR}_{0.995}$ es un problema ligeramente más sencillo: es muy posible que ya hayamos visto alrededor de diez u once pérdidas de esta magnitud en nuestra base de datos. Para los umbrales de 10 y 4 los estimadores son 40.17 y 46.11, respectivamente, de manera que la discrepancia no es tan grande.

Así, la sensibilidad de los estimadores de los cuantiles puede no ser muy severa cuando se trata de estimar cuantiles moderadamente altos, dentro del rango de los datos, pero se incrementa para los cuantiles más lejanos. Esto no es sorprendente, pues la estimación de cuantiles en las orillas de los datos o más allá del rango de los datos es un problema inherentemente difícil que representa un reto para cualquier método.

Calculemos ahora los estimadores de la Esperanza Condicional de la Cola. Recordemos que

$$\widehat{ES}_q = \frac{\widehat{\text{VaR}}_q}{1 - \widehat{\xi}} + \frac{\beta - \widehat{\xi}u}{1 - \widehat{\xi}}.$$

Así, para $u = 10$, tenemos que $1 / (1 - \widehat{\xi}) \approx 2.0$ y $(\beta - \widehat{\xi}u) / (1 - \widehat{\xi}) \approx 4.0$. Esto nos dice que, esencialmente, \widehat{ES}_q se obtiene duplicando el valor de $\widehat{\text{VaR}}_q$. Así, el estimador de $ES_{0.999}$ es 191.54, el cual se ha marcado en la siguiente Figura con una segunda recta vertical.



El estimador de la Esperanza Condicional de la Cola para $u = 10$ es

$$\widehat{ES}_{0.999} = 191.54.$$

De la misma manera, se obtiene que $\widehat{ES}_{0.995} = 83.85$ para $u = 10$; así, de acuerdo con este ajuste del modelo, esperamos que cuando $\text{VaR}_{0.995}$ sea excedido (es decir, cuando se exceda 40.17 millones de DKK) el tamaño aproximado de estas pérdidas sea 83.85 millones de DKK. Puesto que $\text{VaR}_{0.995}$ es el tamaño de cinco de entre cada mil pérdidas, aproximadamente, nuestros 2167 datos podrían incluir 10 u 11 pérdidas que hayan excedido esta cantidad; al revisar los datos, notamos que hay 10 pérdidas por encima de $\text{VaR}_{0.995}$, como era de esperarse. Estas pérdidas tienen las siguientes magnitudes (en millones de DKK): 263.250, 56.225, 50.066, 65.708, 46.500, 57.411, 47.020, 42.092, 152.413 y 144.658. El valor promedio de estas pérdidas es 92.53, si bien hay que notar la presencia de tres pérdidas que están muy por encima de este valor. Por su parte, para $\text{VaR}_{0.999}$, tenemos 3 datos en nuestra serie de datos daneses que han excedido esta cantidad: el promedio de estas tres pérdidas es 186.77; esta cantidad está muy de acuerdo con el estimador $\widehat{ES}_{0.999} = 191.53$.

La siguiente Tabla muestra estimadores del VaR_q para $q = 0.995, 0.999$ y 0.9999 y de ES_q para $q = 0.995$ y 0.999 , para algunos umbrales, junto con los correspondientes estimadores de forma ξ y de escala β de la DPG. Se indica también el error estándar (E.E.) en la estimación de ξ .

u	3	4	5	10	20
Excesos	532	362	254	109	36
$\hat{\xi}$	0.67	0.72	0.63	0.50	0.68
E. E.	0.07	0.09	0.11	0.14	0.28
$\hat{\beta}$	2.19	2.63	3.81	6.98	9.64
$\widehat{\text{VaR}}_{0.995}$	43.85	46.11	43.20	40.17	37.94
$\widehat{\text{VaR}}_{0.999}$	128.96	146.23	121.17	94.34	102.23
$\widehat{\text{VaR}}_{0.9999}$	600.87	766.95	522.10	304.90	471.32
$\widehat{\text{ES}}_{0.995}$	132.49	164.06	118.99	83.85	107.31
$\widehat{\text{ES}}_{0.999}$	388.52	522.35	330.62	191.53	310.84

Estimadores de ξ , β , VaR y ES para varios umbrales.

4.2.4 Comparación de Cuantiles Obtenidos con el Método Máximos de Bloque y con el Método POT para Pérdidas por Incendios en Dinamarca

A continuación se presenta una comparación entre los estimadores obtenidos mediante el Modelo de Máximos de Bloque (MDB) y los obtenidos mediante el Modelo de Picos

sobre el Umbral (POT). Los umbrales del método POT están tomados de manera que se considere el mismo número de datos que los que se utilizan en el modelo de Máximos de Bloque (MDB); así, por ejemplo, tenemos 132 excesos del umbral 7.99, 44 excesos del umbral 18.55 y 22 excesos del umbral 25.95. El promedio del valor del estimador de ξ en el Modelo MDB es 0.584, mientras que en el POT es 0.671. Los estimadores del VaR son mayores en los ajustes de la f.d. VEG que en los de la DPG, salvo en el caso de $\widehat{VaR}_{0.999}$ para bloques trimestrales ($u = 18.55$).

Max de Bloque			
Número de Máximos	132	44	22
$\hat{\xi}$	0.623	0.512	0.618
$\widehat{VaR}_{0.995}$	45.03	44.03	47.31
$\widehat{VaR}_{0.999}$	124.78	98.92	118.20
POT			
Umbral	7.99	18.55	25.95
$\hat{\xi}$	0.415	0.761	0.837
$\widehat{VaR}_{0.995}$	40.97	37.18	36.56
$\widehat{VaR}_{0.999}$	89.51	107.98	105.67

Comparación entre el POT y el Modelo MDB.

Intentaremos ver ahora cuál de los dos métodos arroja estimadores de VaR que se asemejen más con la información histórica de nuestra serie. Iniciamos con los estimadores de $VaR_{0.995}$: en la serie tenemos 11 pérdidas que han excedido los estimadores puntuales 36.56 y 37.18; hemos visto 10 excesos de 40.97; 9 pérdidas por encima de

las cantidades 44.03 y 45.03; y sólo 8 que han excedido 47.31 millones de DKK. Para los estimadores de $VaR_{0.999}$ hemos visto 3 excesos en todos los casos. Puesto que la serie consta de 2157 observaciones, esperamos haber visto alrededor de 10 u 11 excesos de $VaR_{0.995}$ y 2 de $VaR_{0.999}$. Utilizando este criterio, podemos decir que, en el caso de los datos daneses, los estimadores del $VaR_{0.995}$ obtenidos por el método POT reflejan mejor lo visto históricamente.

4.3 Caso 2: Datos acción BMW

La serie consta de 6146 rendimientos logarítmicos diarios del precio de la acción de BMW durante el período comprendido entre enero de 1973 y julio de 1996. Si el precio de la acción en el tiempo t es S_t , el rendimiento logarítmico (*log-return*) está dado por

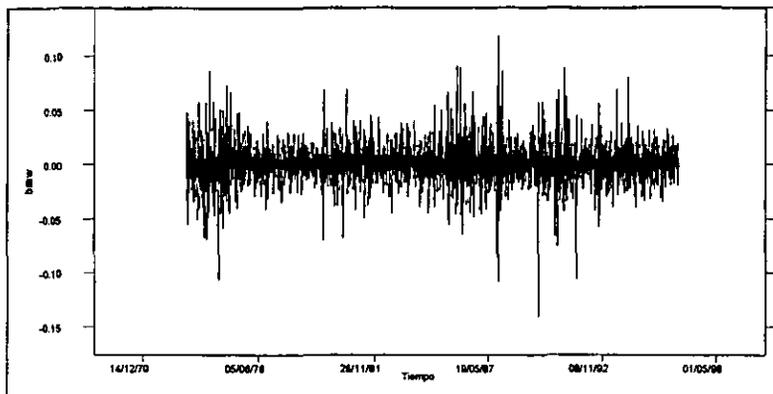
$$r_t = \log(S_t/S_{t-1}),$$

el cual, para movimientos pequeños del precio, es aproximadamente igual al rendimiento relativo $(S_t - S_{t-1})/S_{t-1}$. Nos interesan los movimientos de los precios a la baja, los cuales están representados por los rendimientos logarítmicos menores que cero.

4.3.1 Análisis Preliminar de Datos de Pérdidas en el Precio de la Acción BMW

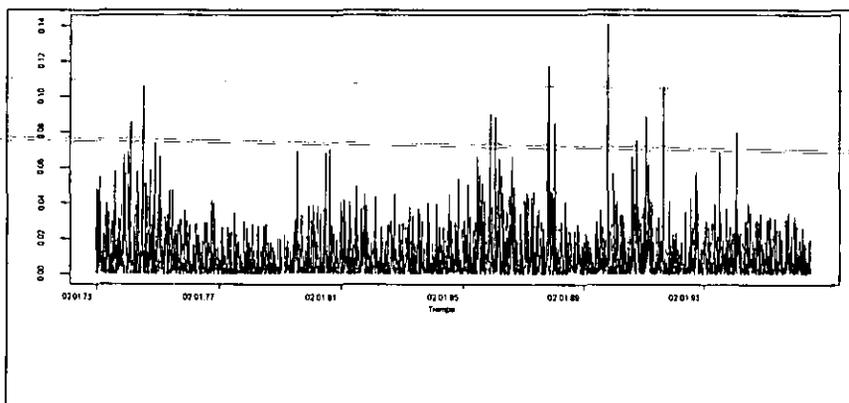
Comenzamos nuestro análisis examinando la serie de tiempo de los datos. Esto nos muestra que los valores más extremos se dieron en 1974, 1985, 1986, 1987, 1989, 1991

y 1994.



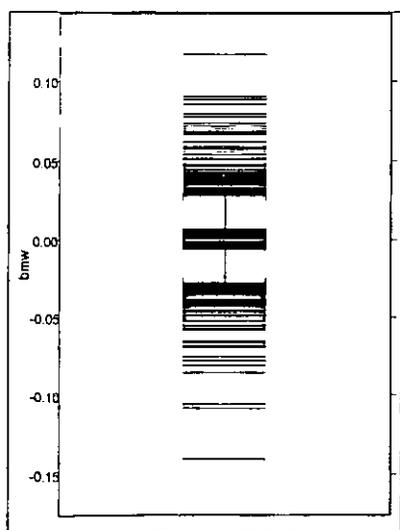
Serie de tiempo BMW.

La serie de tiempo de BMW no muestra particular evidencia de que los rendimientos no sean i.i.d. Sin embargo, si consideramos la serie de los valores absolutos de los rendimientos de los datos, podemos ver que hay evidencia de que los valores grandes se dan por “grupos” o aglomeraciones (*clusters*). Este comportamiento es típico para datos financieros.

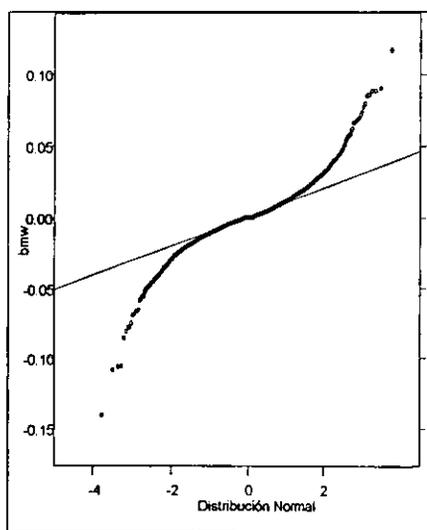


Serie de tiempo de los valores absolutos de los rendimientos BMW. Se observan *clusters* de valores grandes.

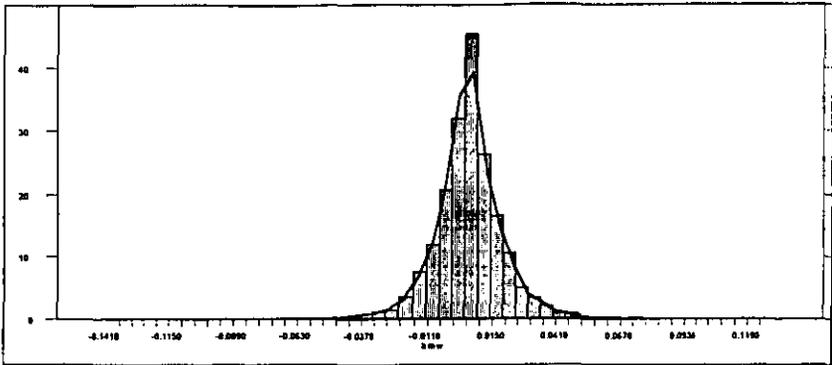
Realizamos a continuación el análisis del Boxplot, del QQ-Plot contra la distribución Normal y del histograma de la serie BMW. En el Boxplot vemos que las pérdidas extremas han sido mayores (en tamaño) que las ganancias extremas, si bien hay aproximadamente el mismo número de pérdidas que de ganancias; la distribución parece bastante simétrica con respecto al cero. Por su parte, el QQ-Plot nos deja ver claramente que tanto la cola inferior como la cola superior de nuestra distribución son mucho más pesadas que las colas de la distribución Normal, lo cual sugiere que deberíamos utilizar un modelo con cola más pesada. Finalmente, en el histograma de los datos se superimpuso la distribución Normal que, aunque aparentemente no es un mal ajuste para el centro de la distribución, sabemos (por el QQ-Plot) que subestima las colas de la distribución de nuestros datos, que es el área que nos interesa en este análisis.



Boxplot *log-returns* BMW.



QQ-Plot BMW contra distrib. Normal.



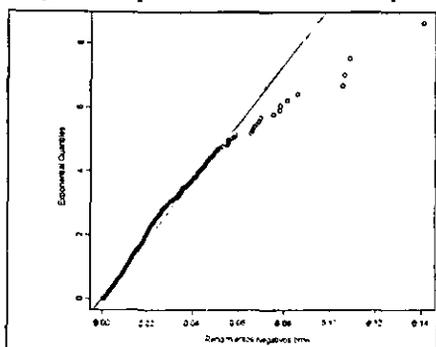
Histograma datos BMW, con distribución Normal sobreimpuesta.

La siguiente Tabla resume los principales datos estadísticos para nuestra serie de rendimientos de la acción BMW. El resumen estadístico se calcula para todos los datos (segunda columna), para los datos de ganancias (tercera columna) y para los datos de pérdidas (última columna).

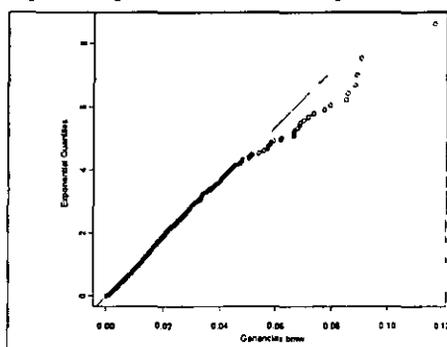
	BMW (Todos)	BMW (Ganancias)	BMW (Pérdidas)
Mínimo	-0.141	0.000	0.000
Primer cuartil	-0.007	0.004	0.004
Media	0.0003	0.011	0.011
Mediana	0.000	0.008	0.008
Tercer Cuartil	0.007	0.015	0.014
Máximo	0.117	0.117	0.141
Desv. Estándar	0.015	0.011	0.011
Total Datos	6146	2766	2769

Resumen estadístico datos BMW.

El siguiente paso de nuestro análisis preliminar es investigar el ajuste de la distribución exponencial a nuestros datos de pérdidas (izquierda) y de ganancias (derecha). El ajuste de la exponencial a los datos de ganancias es aceptable, pero nuestros datos de pérdidas parecen tener la cola un poco más pesada que la distribución exponencial.

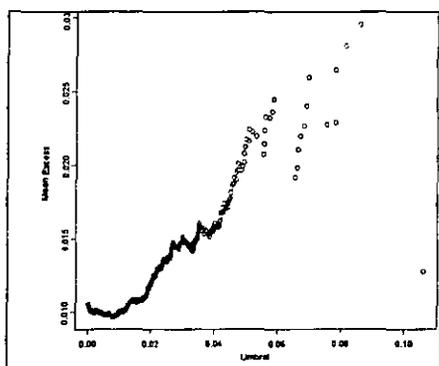


QQ-Plot Pérdidas BMW.

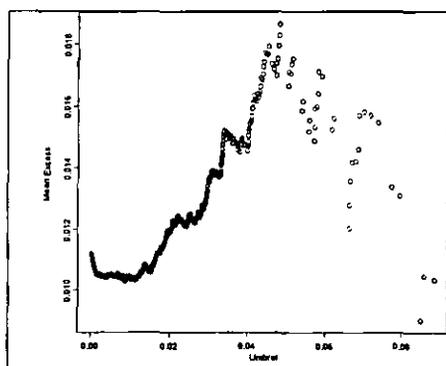


QQ-Plot Ganancias BMW.

Queda únicamente por trazar la gráfica de la Función Media de Exceso (F.M.E.) empírica, tanto para pérdidas como para ganancias. Recordemos que si la gráfica presenta pendiente positiva, esto es un indicio de que la cola de nuestros datos es pesada.



Gráfica F.M.E. Pérdidas BMW.



Gráfica F.M.E. Ganancias BMW.

La gráfica de la función muestral media de exceso de los valores absolutos de los rendimientos negativos (que corresponden, por tanto, a las pérdidas) muestra

claramente que los rendimientos tienen cola pesada. También es claro el cambio en la curvatura de la gráfica alrededor de 0.02 ó 0.025, por lo que podríamos considerar estos valores como umbrales iniciales para realizar el análisis POT.

4.3.2 TVE para Series Estacionarias

En los Capítulos 1 y 3 se presentó una gran cantidad de material sobre extremos, en el cual nos restringimos al caso de observaciones i.i.d.; sin embargo, en la realidad los eventos extremos ocurren a menudo en *clusters* (agrupamientos o aglomeraciones), ocasionados por dependencias locales de los datos. Por ejemplo, las grandes reclamaciones en seguros se deben principalmente a huracanes, tormentas, inundaciones, terremotos, etc. Las reclamaciones están por tanto asociadas a este evento y no ocurren de manera independiente. Lo mismo puede observarse con datos financieros, como los tipos de cambio y los precios de activos. Si ocurre un valor grande en una serie de tiempo de este tipo, usualmente podremos observar un *cluster* de valores grandes en un corto período de tiempo después de este evento.

Vamos qué sucede si relajamos nuestra suposición de i.i.d. y consideramos una serie de tiempo estrictamente estacionaria con dependencias seriales. De acuerdo con lo dicho sobre los *clusters* de datos financieros, las series de tiempo estacionarias deberían ser un mejor modelo para nuestros *log-returns*.

Por supuesto, no hay nada que nos impida simplemente ajustar la distribución de Valores Extremos Generalizada (VEG) a los máximos de bloque como se describió en la Sección 3.3 y como se hizo en el caso de los datos daneses; lo que queremos es tener

una justificación teórica para escoger esta distribución límite específica en el caso estacionario, así como la que teníamos para el caso i.i.d. Los resultados relevantes se detallan en [15] y en [6]. Aquí presentamos sólo un breve resumen.

Un concepto que cobra gran importancia es el de índice extremo de una serie estacionaria. El índice extremo es una cantidad que, de manera intuitiva, nos permite caracterizar la relación entre la estructura de dependencia de los datos y su comportamiento extremo. Para entender mejor este concepto, recordemos primero la Aproximación de Poisson (Proposición 1.1.6): Sean X_1, \dots, X_n v.a.i.i.d. y sea (u_n) una sucesión de números reales. Para $0 \leq \tau \leq \infty$

$$\lim_{n \rightarrow \infty} n\bar{F}(u_n) = \tau$$

si y sólo si

$$\lim_{n \rightarrow \infty} P\{M_n \leq u_n\} = \exp(-\tau). \quad (4.7)$$

Esto puede interpretarse de la siguiente manera: si la probabilidad de que el máximo de un bloque exceda (o no exceda) un umbral alto (u_n) tiene un límite asintótico cuando se incrementan el umbral y el tamaño del bloque, entonces el número esperado de observaciones dentro del bloque que exceden el umbral también tendrá un límite asintótico, y viceversa.

Necesitamos la siguiente notación: (X_n) será ahora nuestra serie de tiempo estacionaria, donde la distribución (marginal) de cada observación X_i es F y $M_n = \max(X_1, \dots, X_n)$. Denotaremos mediante (\widetilde{X}_n) a una serie i.i.d. asociada, con la misma distribución marginal F y hacemos $\widetilde{M}_n = \max(\widetilde{X}_1, \dots, \widetilde{X}_n)$.

Definición 4.3.1 Sea $0 \leq \theta \leq 1$ un número real y supongamos que para cada $\tau > 0$

existe una sucesión $u_n(\tau)$ tal que se da lo siguiente:

$$\begin{aligned}\lim_{n \rightarrow \infty} n\bar{F}(u_n(\tau)) &= \tau \\ \lim_{n \rightarrow \infty} P\{M_n \leq u_n(\tau)\} &= \exp(-\theta\tau).\end{aligned}\tag{4.8}$$

Diremos entonces que (X_n) tiene índice extremo θ .

Puede demostrarse que θ está bien definido de esta manera y no depende de la elección específica de la sucesión $(u_n(\tau))$. De hecho, si (X_n) tiene índice extremo θ , entonces para u_n en general y $\tau > 0$ puede demostrarse que las siguientes expresiones son equivalentes:

$$\begin{aligned}\lim_{n \rightarrow \infty} n\bar{F}(u_n) &= \tau \\ \lim_{n \rightarrow \infty} P\{\widetilde{M}_n \leq u_n\} &= \exp(-\tau) \\ \lim_{n \rightarrow \infty} P\{M_n \leq u_n\} &= \exp(-\theta\tau).\end{aligned}$$

Estas tres expresiones equivalentes para el caso de series de tiempo estacionarias con índice extremo θ pueden compararse con las dos expresiones equivalentes dadas en (4.7) para series i.i.d.

Una importante consecuencia de esta definición es el siguiente resultado para series estacionarias:

Teorema 4.3.2 Si (X_n) es estacionaria con índice extremo $\theta \in (0, 1]$, entonces

$$\lim_{n \rightarrow \infty} P\left\{\left(\widetilde{M}_n - b_n\right)/a_n \leq x\right\} = H(x)$$

para alguna f.d. $H(x)$ no-degenerada si y sólo si

$$\lim_{n \rightarrow \infty} P\{(M_n - b_n)/a_n \leq x\} = H^\theta(x),$$

con $H^\theta(x)$ también no-degenerada.

Puesto que las distribuciones de valores extremos son max-estables (ver Definición 1.2.6), para $F \in DAM(H_\xi)$, la distribución asintótica de los máximos normalizados de la serie estacionaria (X_n) con índice extremo θ también es una distribución de valores extremos (pues $(H_\xi)^\theta$ es del mismo tipo que H_ξ). El parámetro de forma ξ es el mismo que en el caso i.i.d., puesto que elevar una f.d. a la potencia θ sólo afecta los parámetros de localización y escala.

A partir de la definición del índice extremo θ (relación (4.8)), podemos deducir que, para valores grandes de n ,

$$P\{M_n \leq u_n\} \approx P^\theta\{\tilde{M}_n \leq u_n\} = F^{n\theta}(u_n), \quad (4.9)$$

siempre que $n\bar{F}(u_n) \rightarrow \tau > 0$.

Esta es una especie de definición informal de θ . Así, diremos que si una serie estacionaria exhibe la propiedad (4.9), la serie tiene índice extremo θ .

Esta definición nos da cierta intuición acerca de la relación entre el efecto de aglomeración y la naturaleza de θ . Vemos que el máximo de n observaciones de una serie estacionaria con índice extremo θ se comporta como el máximo de $n\theta$ observaciones de la serie i.i.d. asociada. De esta manera, podemos pensar que $n\theta$ cuenta el número de *clusters* pseudo-independientes dentro de un total de n observaciones, por lo que θ puede interpretarse como el recíproco del tamaño medio de las aglomeraciones.

Esta es la justificación que necesitábamos para ajustar la distribución VEG a los máximos de bloque de una serie de tiempo estacionaria con tendencia a formar *clusters* de valores grandes, como es el caso de las series de tiempo financieras. La

dependencia de las X_i tiene el efecto de que la convergencia a la distribución VEG es más lenta, puesto que el tamaño efectivo de la muestra es $n\theta$, que es más pequeño que n . Así, debe tenerse aún más cuidado de que los bloques sean suficientemente grandes como para que sus máximos puedan admitir un modelo de VEG. Del mismo modo, los bloques deben ser suficientemente grandes como para poder suponer que los máximos de distintos bloques son efectivamente independientes.

En la Sección 3.3 y en el caso de los datos daneses mencionamos que el nivel de retorno $R_{n,k}$ es un cuantil específico de la distribución marginal. Hemos visto que, en los casos en que se tienen datos i.i.d., la probabilidad asociada es

$$(1 - 1/k)^{1/n} = P^{1/n} \{M_n \leq R_{n,k}\} = F(R_{n,k}), \quad (4.10)$$

de manera que $R_{n,k} = x_{(1-1/k)^{1/n}}$. Sin embargo, cuando nuestros datos no son i.i.d., en lugar de la identidad (4.10) podemos utilizar nuestra definición informal del índice extremo (4.9) para obtener la aproximación asintótica

$$(1 - 1/k)^{1/(n\theta)} = P^{1/(n\theta)} \{M_n \leq R_{n,k}\} \approx F(R_{n,k}),$$

de manera que

$$R_{n,k} \approx x_{(1-1/k)^{1/(n\theta)}}. \quad (4.11)$$

Tenemos pues que en el caso de series de tiempo estacionarias es más difícil decir exactamente cuál es la probabilidad asociada a cierto nivel de retorno. Si deseamos responder esta pregunta debemos conocer, o poder estimar, el índice extremo θ de la serie. Si no conocemos el índice extremo, sabemos que el nivel de retorno $R_{n,k}$ se excede en un n -bloque de entre cada k n -bloques, en promedio, pero no sabemos

cuántos excesos ocurrirán en ese n -bloque extremo o período acentuado, puesto que eso depende de la propensidad de la serie a formar *clusters*. En este sentido, $R_{n,k}$ nos da una idea acerca de la frecuencia de los períodos acentuados, pero no de la verdadera frecuencia de los rendimientos extremos promediada a largo plazo. Para obtener información más precisa necesitamos el índice extremo θ .

La relación (4.9) sugiere cómo podría construirse un estimador de θ mediante el llamado método de bloques. Este método está basado en argumentos asintóticos que pueden encontrarse en [6, Cap. 8]. El método de bloques para estimar el índice extremo consiste en dividir los datos X_1, \dots, X_r en m bloques de tamaño n , como se hace para ajustar la distribución VEG, y poner umbrales altos, digamos en u . Así, el estimador asintótico natural para θ es

$$\hat{\theta} = n^{-1} \frac{\log(1 - K_u/m)}{\log(1 - N_u/(mn))},$$

donde N_u es el número de excesos del umbral y K_u es el número de bloques en los que hay uno o más excesos del umbral. Cuando K_u/m y $N_u/(mn)$ son pequeños, este estimador se reduce a K_u/N_u .

También para aplicar el Modelo de Picos Sobre el Umbral deben hacerse ligeras modificaciones. La idea de aproximar $F_u(x)$ mediante una Distribución Pareto Generalizada (DPG) $G_{\xi, \beta}(x)$ para algún umbral u alto y para los parámetros ξ y β sigue teniendo sentido. El problema es estadístico: ¿cómo estimar los parámetros de la DPG a partir de una muestra dependiente de pérdidas que exceden el umbral u ?

Dos posibles enfoques para este problema son:

1. Hacer algunas suposiciones sobre la naturaleza de la dependencia de los datos y

maximizar una verosimilitud adecuada para la muestra de excesos dependientes.

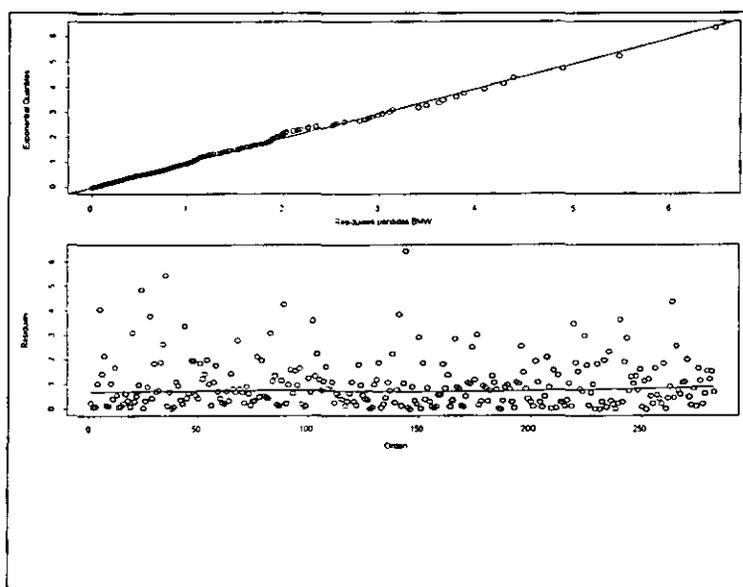
2. Intentar hacer que los excesos sean “más independientes”. Puesto que n observaciones extremas de un proceso con índice extremo θ proporcionan aproximadamente la misma cantidad de información sobre la cola de F que la que obtendríamos de $n\theta$ observaciones independientes, podemos intentar reducir los datos a observaciones i.i.d. deshaciendo los *clusters* (*declustering*). Para ello, identificamos los *clusters* de excesos del umbral u y reducimos cada *cluster* a un único representante, como podría ser el máximo dentro del *cluster*. Uno de los métodos que puede utilizarse para llevar a cabo este proceso es el método de *runs* (corridas): cuando decimos que $run = r$, consideramos que cualquier sucesión de r observaciones consecutivas por debajo del umbral separa dos *clusters*.

Tenemos ya la información necesaria acerca de la importancia del índice extremo de una serie estacionaria y tenemos también un método para calcular, a partir de los datos, un estimador del índice extremo. Podemos por tanto comenzar nuestro análisis de la serie de pérdidas en el precio de la acción BMW.

4.3.3 Análisis de Máximos de Bloque para Pérdidas en el Precio de la Acción BMW

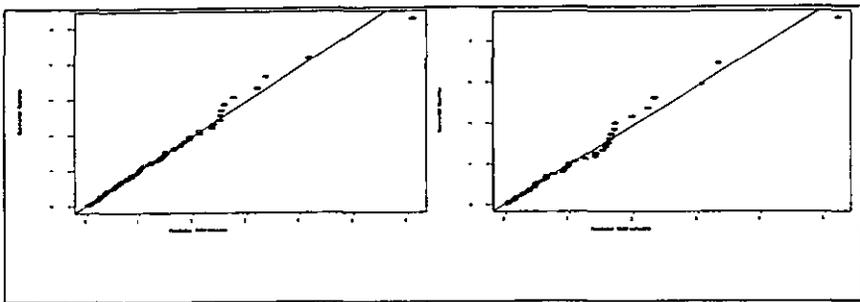
El primer análisis se realiza para bloques mensuales. Es importante recalcar que estamos interesados en los rendimientos negativos, de manera que nuestros 6146 datos se reducen a 2769 pérdidas, que a su vez se convierten en 283 máximos de bloque, para

bloques que contienen aproximadamente 22 observaciones cada uno. Los estimadores de los parámetros son $\hat{\xi} = 0.232$, $\hat{\sigma} = 0.009$ y $\hat{\mu} = 0.019$. El ajuste del modelo Fréchet a estos máximos de bloque se verificó utilizando, como en el caso de los datos daneses, residuales exponenciales (4.6). Tanto el QQ-Plot como la gráfica de dispersión de estos nos indican que el ajuste es adecuado.



QQ-Plot y Dispersión de Residuales pérdidas BMW mensuales.

Se realiza también el análisis para bloques de mayor tamaño. Los estimadores del parámetro de forma ξ son 0.241, 0.274 y 0.207 para bloques trimestrales, semestrales y anuales, respectivamente. También en estos casos se investigó el ajuste Fréchet mediante el uso de residuales. Los QQ-plots de la siguiente Figura muestran que el ajuste parece adecuado.



QQ-Plots para residuales de las pérdidas BMW para máximos trimestrales (izquierda) y semestrales (derecha).

La siguiente Tabla muestra los estimadores de los parámetros obtenidos al ajustar la distribución VEG a máximos de bloque mensuales, trimestrales, semestrales y anuales. Se incluyen asimismo los niveles de retorno de 20-periodos. Así, puede verse que el nivel de retorno de 20 trimestres es 0.076, de manera que en un bloque trimestral cada 5 años esperamos ver pérdidas diarias de alrededor del 7.6%. El estimador del nivel de retorno de 80 trimestres (marcado en la columna R_{20}) es 0.115, lo que significa que cada 20 años veremos rendimientos logarítmicos negativos de, en promedio, -0.115.

	$\approx n$	m	$\hat{\xi}$	E.E.	$\hat{\sigma}$	E.E.	$\hat{\mu}$	E.E.	$R_{n,20}$	R_{20}
Mensual	22	283	0.232	0.05	0.009	0.00	0.019	0.00	0.057	0.118
Trimestral	65	95	0.241	0.08	0.011	0.00	0.027	0.00	0.076	0.115
Semestral	128	48	0.274	0.13	0.014	0.00	0.033	0.00	0.096	0.120
Anual	256	24	0.207	0.19	0.019	0.00	0.044	0.00	0.120	0.120

Estimadores de parámetros VEG y niveles de retorno Pérdidas BMW.

En la columna R_{20} se muestran los niveles de retorno de 240-meses, 80-trimestres, 40-semestres y 20-años. Obsérvese que los estimadores de R_{20} para los diferentes tamaños de bloque son prácticamente iguales, lo que muestra que nuestro modelo para máximos es consistente.

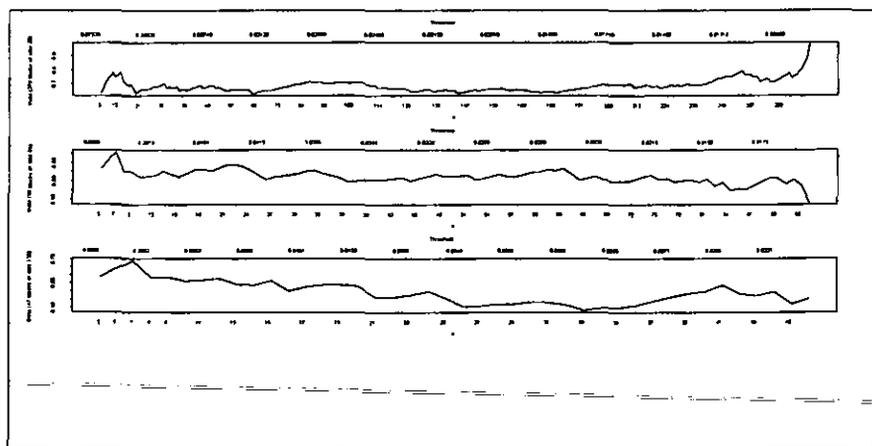
Nos interesa ahora estimar los cuantiles correspondientes a los niveles de retorno calculados. Puesto que suponemos que nuestros datos son una serie de tiempo estacionaria, sabemos por (4.11) que necesitamos conocer el índice extremo de la serie para pasar de la información expresada como niveles de retorno a información en términos de cuantiles. Intentamos entonces estimar el índice extremo θ para los distintos tamaños de bloque que hemos considerado. La siguiente Tabla muestra los estimadores de θ obtenidos mediante el método de bloques, para máximos mensuales, trimestrales y semestrales, conforme se varía el número N_u de excesos del umbral.

(m, n)	N_u	15	20	25	30	40	50	100	150	200
	u	0.065	0.056	0.051	0.049	0.045	0.043	0.034	0.029	0.026
mes	K_u	11	14	17	19	26	32	59	90	110
(283,22)	$\hat{\theta}$	0.748	0.718	0.701	0.655	0.681	0.678	0.658	0.717	0.689
trim.	K_u	8	10	13	15	21	25	40	55	65
(95,65)	$\hat{\theta}$	0.554	0.525	0.555	0.540	0.589	0.575	0.513	0.589	0.536
sem.	K_u	8	10	13	14	17	21	28	34	42
(48,128)	$\hat{\theta}$	0.587	0.565	0.611	0.556	0.529	0.558	0.425	0.400	0.521

Estimadores del Índice Extremo θ para *log-returns* negativos BMW.

En la Tabla anterior vemos que los estimadores del índice extremo θ para bloques

mensuales están entre 0.655 y 0.748, para un promedio de 0.694. Por su parte, los bloques trimestrales dan estimadores $\hat{\theta}$ entre 0.513 y 0.589, para un promedio de 0.555; mientras que en los bloques semestrales los estimadores de θ están entre 0.400 y 0.611, para un promedio de 0.528. En el caso de bloques trimestrales y semestrales, los promedios de los estimadores de θ obtenidos son suficientemente parecidos; para bloques mensuales el valor de n es demasiado pequeño como para utilizar un argumento asintótico. Las siguientes gráficas muestran que el estimador del índice extremo θ es más estable cuando se trabaja con bloques trimestrales; esto era de esperarse, pues tanto m como n son suficientemente grandes. En el eje- x superior se indica el umbral u y en el eje- x inferior se marca el número de excesos.



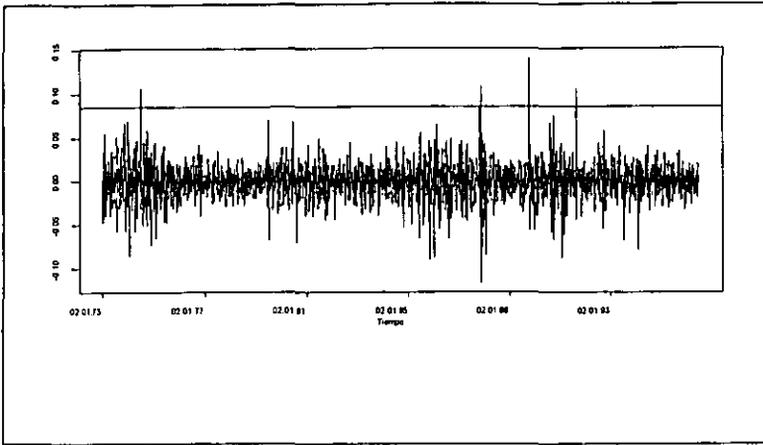
Estabilidad del índice extremo θ de la serie de pérdidas BMW al variar el umbral para bloques de tamaño: 22 (meses), arriba; 65 (trimestres), centro; y 128 (semestres), abajo.

Supondremos entonces que, como nos indica nuestro análisis de bloques trimestrales, la serie de los *log-returns* negativos de BMW tiene índice extremo 0.55, de

modo que el tamaño promedio de los *clusters* es 1.8. Esto significa que podemos utilizar la ecuación (4.11) para calcular que nuestro nivel de retorno de 20-trimestres corresponde aproximadamente al cuantil 0.9985. Así, se estima que la probabilidad de exceder este nivel de retorno es 0.015.

Podemos ahora expresar nuestros resultados en términos de la frecuencia tanto de los bloques extremos como de los rendimientos extremos. En un período de 5 años esperamos tener un trimestre acentuado en el que las pérdidas serán de alrededor del 7.6%, en promedio. Cuando ocurran pérdidas de esta magnitud, esperamos ver *clusters* de rendimientos negativos grandes, con un tamaño promedio de 1.8. Más aún, puesto que en estos 5 años hay alrededor de 1305 días hábiles, esperamos que haya aproximadamente $0.0015 \times 1305 \approx 1.9575$ días en los que veremos pérdidas de esta magnitud.

Recíprocamente, supongamos que necesitamos un estimador del cuantil 0.999 de la distribución de los rendimientos logarítmicos negativos. Utilizando de nuevo la expresión (4.11) y el valor de $\theta = 0.55$, obtenemos que esto es equivalente a estimar el nivel de retorno de 28.89-trimestres. Cerramos este número y estimamos el nivel de retorno de 29-trimestres. Obtenemos el estimador puntual 0.085. La siguiente Figura muestra que este valor corresponde aceptablemente con lo observado históricamente. Sobreimpuesta en la gráfica de la serie de tiempo de los rendimientos negativos se marcó el estimador del cuantil 0.999 o nivel de retorno de 29-trimestres mediante una recta horizontal. La longitud total de la serie BMW es 6146, de manera que esperamos que alrededor de 6.1 rendimientos excedan el cuantil 0.999. Tenemos que 5 rendimientos lo exceden en la práctica.

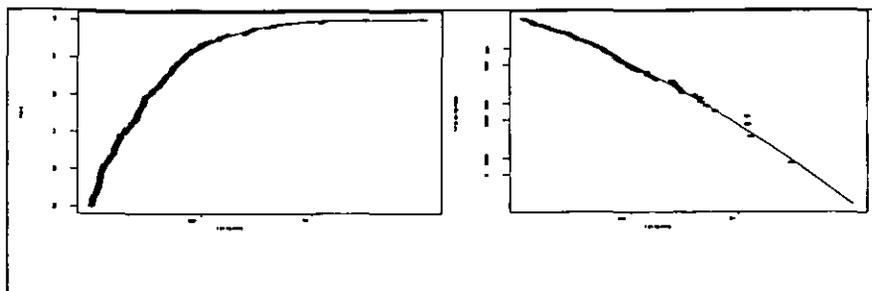


Serie (con pérdidas positivas) BMW, con el cuantil 0.999 marcado.

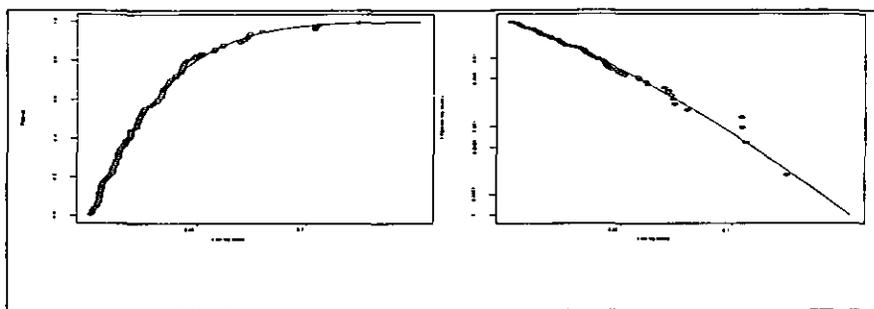
4.3.4 Análisis POT para Pérdidas en el Precio de la Acción BMW

Al igual que en el caso del ajuste de la distribución VEG, al aplicar este método también debemos tomar en cuenta que nuestros datos no son i.i.d., sino que se presentan en *clusters*. Para ello, además de elegir un umbral, debemos elegir un *run* (corrida) para nuestros datos.

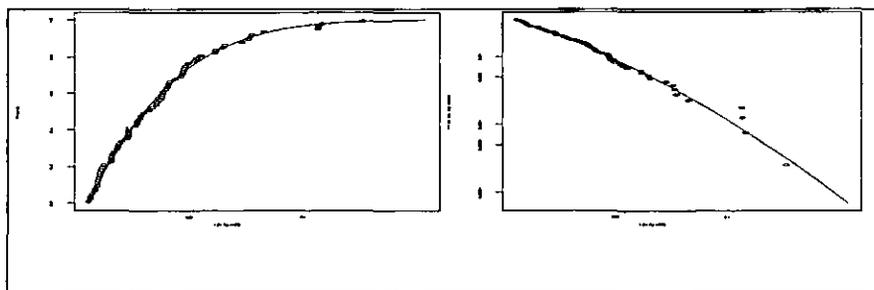
La gráfica de la función media de exceso nos sugiere elegir como umbral inicial una cantidad de alrededor de 0.02 ó 0.03. Elegimos pues, para iniciar, $u = 0.025$. Las siguientes gráficas muestran el ajuste de la DPG a la distribución de exceso F_u (izquierda) y a la cola \bar{F} de la distribución subyacente (derecha), cuando los *clusters* están separados por 0 (es decir, no están separadas), 20 y 30 observaciones por debajo del umbral, respectivamente.



$Run = 0, u = 0.025$. Se obtiene $\hat{\xi} = 0.178$.



$Run = 20, u = 0.025$. Se obtiene $\hat{\xi} = 0.220$.



$Run = 30, u = 0.025$. Se obtiene $\hat{\xi} = 0.215$.

Las gráficas anteriores muestran que el ajuste es bueno en todos los casos. La siguiente Tabla muestra los estimadores de los parámetros de la DPG para umbrales desde 0.02 hasta 0.04, cuando $run = 0, 20$ y 30 .

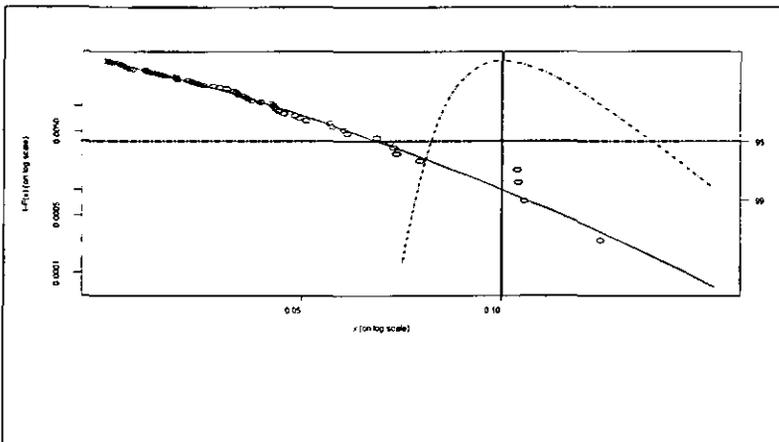
u	Excesos	ξ	E.E.	β	run
0.02	83	0.2373	0.137	0.0133	30
0.025	70	0.2149	0.153	0.0142	30
0.03	58	0.2351	0.173	0.0139	30
0.04	33	0.2082	0.153	0.0148	30
0.02	111	0.2007	0.116	0.0131	20
0.025	89	0.2201	0.130	0.0126	20
0.03	66	0.2624	0.163	0.0124	20
0.04	35	0.3030	0.252	0.0136	20
0.02	354	0.2232	0.069	0.0093	0
0.025	212	0.1778	0.082	0.0110	0
0.03	136	0.1428	0.095	0.0126	0
0.04	65	0.2644	0.172	0.0118	0

Estimadores DPG para distintos umbrales, BMW.

Nótese que, particularmente cuando $run = 30$, los estimadores del parámetro de forma ξ son bastante estables, con un promedio de 0.2239. En el caso en que $run = 20$, el promedio de ξ es 0.2461.

Ilustraremos el cálculo del VaR para $u = 0.025$, $run = 30$. Sin tomar en cuenta que los excesos se presentan en *clusters*, estaríamos hablando de aproximadamente 200 excesos del umbral; pero al escoger $run = 30$, se toman en cuenta solamente 70 excesos del umbral. En este caso, obtenemos que el estimador del $VaR_{0.999}$ para las pérdidas BMW es 0.1004. Esto se muestra en la siguiente Figura: la recta punteada

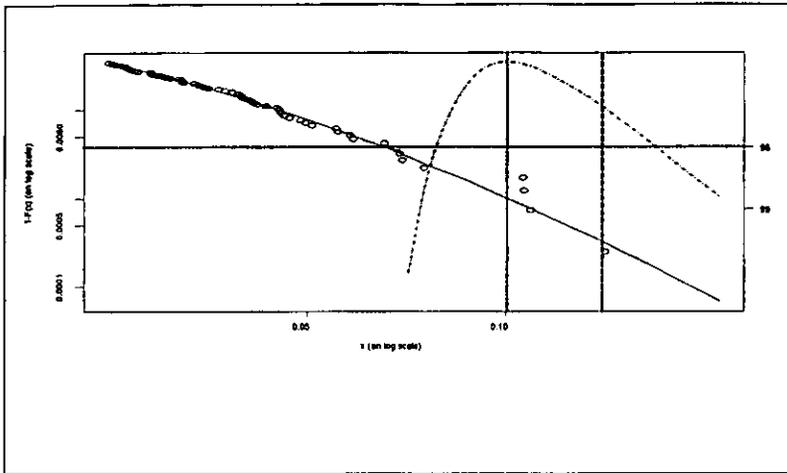
vertical se intersecta con el estimador de la cola en el punto $(0.1004, 0.001)$, lo que nos permite leer el estimador del VaR en el eje- x . La curva punteada es una herramienta que nos permite calcular intervalos de confianza (mediante el método de verosimilitud de perfiles) para el VaR. El eje- y del lado derecho de la gráfica es una escala de confianza (no una escala de cuantiles). La línea punteada horizontal corresponde a una confianza del 95%; las coordenadas en x de los dos puntos donde la curva punteada intersecta a la línea horizontal son los extremos del intervalo de confianza del 95%: $(0.07924, 0.16823)$. Es importante notar que el intervalo obtenido es asimétrico, lo cual refleja el problema de estimar altos cuantiles para datos de cola pesada: es más fácil acotar el intervalo por abajo que hacerlo por arriba. Obsérvese que se obtiene un intervalo de confianza del 99% más amplio si se baja la línea horizontal al valor 99 del eje de confianza.



Estimador de $VaR_{0.999}$ para $u = 0.025$ y $run = 30$.

El estimador de $ES_{0.999}$ se ilustra en la siguiente Figura mediante una segunda recta vertical, a la derecha de la correspondiente al $VaR_{0.999}$. El estimador puntual

que se obtiene es 0.1392.



Estimador de $ES_{0.999}$ con $u = 0.025$ y $run = 30$.

La siguiente Tabla muestra los estimadores de VaR y de ES para dos distintos valores del umbral (0.02 y 0.025), donde separamos los *clusters* por 0, 20 y 30 observaciones por debajo del umbral.

u	run	$VaR_{0.995}$	$VaR_{0.999}$	$VaR_{0.9999}$	$ES_{0.995}$	$ES_{0.999}$
0.02	0	0.05007033	0.08098086	0.1497941	0.0706189	0.11041
0.025	0	0.05039095	0.07933166	0.138161	0.06928087	0.1044779
0.02	20	0.06136126	0.1020376	0.1886009	0.08814847	0.1390375
0.025	20	0.05541479	0.09275413	0.1753751	0.08020455	0.1280845
0.02	30	0.0640145	0.1105334	0.2170801	0.09513455	0.15613
0.025	30	0.05902148	0.1004137	0.191061	0.08644162	0.1391701

Estimadores VaR y ES para Pérdidas BMW con distintos umbrales y *runs*.

La Tabla anterior muestra que todos los estimadores de VaR y ES son más altos para $u = 0.02$ y $run = 30$, mientras que para los demás modelos los valores que se obtienen son muy similares.

En la práctica, los estimadores de $VaR_{0.999}$ que parecen más acertados son aquellos en los que se ignora que la serie forma *clusters* y se toma $run = 0$; en este caso, obtenemos 6 observaciones (tanto para $u = 0.25$ como para $u = 0.2$) que han excedido este valor, lo cual es de esperarse para una serie de 6146 rendimientos. En cambio, para $run = 20$ y $run = 30$, se obtiene que el estimador de $VaR_{0.999}$ ha sido excedido solamente en cuatro ocasiones (excepto para $u = 0.02$ y $run = 30$, donde $VaR_{0.999}$ sólo se ha excedido una vez), lo que nos indica que posiblemente hemos sobreestimado ligeramente el valor de $VaR_{0.999}$. El caso de $VaR_{0.995}$ es similar: aquí para $u = 0.025$, $run = 20$ se tienen 22 excesos, y en los demás casos ($run = 20$ y $run = 30$) hay 16 excesos. Para $run = 0$ hay 27 excesos para los umbrales $u = 0.02$ y $u = 0.25$. En teoría, deberíamos haber visto ya alrededor de 30 ó 31 excesos de $VaR_{0.995}$.

En cuanto a los estimadores de ES, se tiene lo siguiente: en el caso de $run = 0$, el promedio de los 27 excesos de $VaR_{0.995}$ es 0.07159, lo cual es muy parecido a los estimadores de $ES_{0.995}$ obtenidos tanto para $u = 0.02$ como para $u = 0.025$. Para $u = 0.025$ y $run = 20$, el promedio de los 22 excesos que hemos visto de $VaR_{0.995}$ es 0.07596, cantidad que el estimador de $ES_{0.995}$ parece sobreestimar ligeramente. En los demás casos, el promedio histórico de los 16 excesos es 0.08314, que es bastante similar a los estimadores obtenidos para $ES_{0.995}$, excepto en el caso $u = 0.02$ y $run = 30$. Veamos ahora lo que ocurre con $ES_{0.999}$: para $run = 0$, el promedio de los 6 excesos que hemos visto de $VaR_{0.999}$ es 0.10468, lo cual es casi exactamente el estimador de

$ES_{0.999}$ obtenido para $u = 0.025$. Para $u = 0.02$ y $run = 30$, el tamaño del único exceso que contiene nuestra serie histórica es 0.1406, que resulta ligeramente inferior al estimador de $ES_{0.999}$ obtenido. En los demás casos, los 4 excesos de $VaR_{0.999}$ promedian 0.11527, de manera que $ES_{0.999}$ parece estar ligeramente sobreestimada.

4.3.5 Comparación de Cuantiles Obtenidos con el Método Máximos de Bloque y con el Método POT para Pérdidas en el Precio de la Acción BMW

La siguiente Tabla presenta una comparación entre los estimadores de $VaR_{0.999}$ obtenidos por el Método de Máximos de Bloque y los obtenidos por el Método POT. El umbral del método POT se escoge de manera que haya el mismo número de excesos de ese umbral que el número de máximos con que se está comparando. Así, por ejemplo, hay 283 excesos del umbral $u = 0.0218$. Cuando $run \neq 0$, el número efectivo de excesos se reduce para tomar en cuenta la tendencia de la serie a formar *clusters*; así, por ejemplo, para $run = 20$, los 283 excesos se convierten en 101 excesos efectivos; 95 excesos se transforman en 45 efectivos y 48 excesos se vuelven 27.

Los estimadores de $VaR_{0.999}$ obtenidos mediante el método POT son más parecidos a los obtenidos a través del método de Máximos de Bloque cuando se toma $run = 20$, es decir, cuando 20 observaciones por debajo del umbral separan 2 *clusters*.

En la práctica, para el estimador $\widehat{VaR}_{0.999} = 0.079$ hay 6 excesos; para 0.085 hay 5 excesos; para los estimadores entre 0.086 y 0.104 hay 4 excesos.

Max de Bloque	Mes	Trimestre	Semestre
Número de Máximos	283	95	48
$\hat{\xi}$	0.232	0.241	0.274
$\widehat{VaR}_{0.999}$	0.089	0.085	0.087
POT			
u	0.0218	0.0347	0.0433
$(Run = 0) \hat{\xi}$	0.171	0.189	0.217
$\widehat{VaR}_{0.999}$	0.079	0.079	0.079
$(Run = 20) \hat{\xi}$	0.170	0.207	0.258
$\widehat{VaR}_{0.999}$	0.097	0.090	0.086
$(Run = 30) \hat{\xi}$	0.217	0.181	0.211
$\widehat{VaR}_{0.999}$	0.087	0.091	0.104

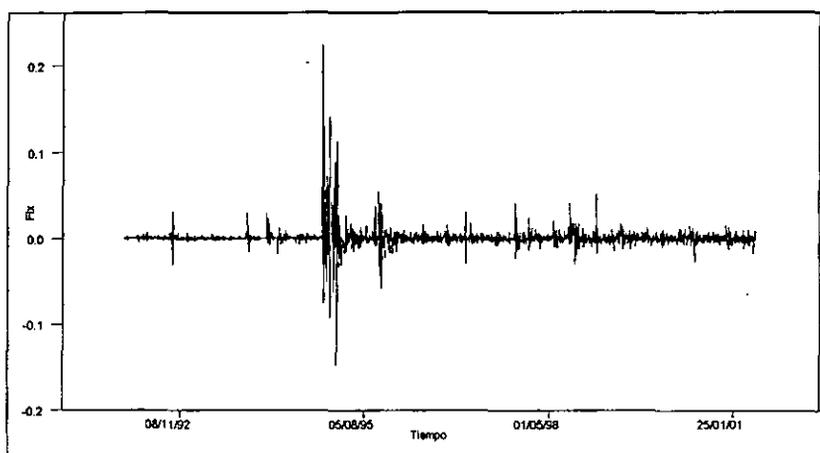
Comparación Métodos POT y MDB, Pérdidas BMW.

4.4 Caso 3: Tipo de cambio FIX peso/dólar

En este caso, nuestros datos corresponden a la variación porcentual del tipo de cambio FIX de pesos mexicanos a dólares americanos (MXN/USD), durante el período comprendido entre enero de 1992 y mayo del 2001. Tenemos pues 2357 datos para realizar nuestros análisis. En este análisis tomamos como pérdidas las devaluaciones del peso con respecto al dólar, y les asignaremos su variación porcentual positiva.

4.4.1 Análisis Preliminar de Datos de Pérdidas en el Tipo de Cambio FIX

Como en los dos casos anteriores, comenzamos por realizar el análisis preliminar, mediante gráficas, de los datos. Iniciamos con la gráfica de la serie de tiempo.



Serie de tiempo del tipo de cambio FIX, con pérdidas positivas.

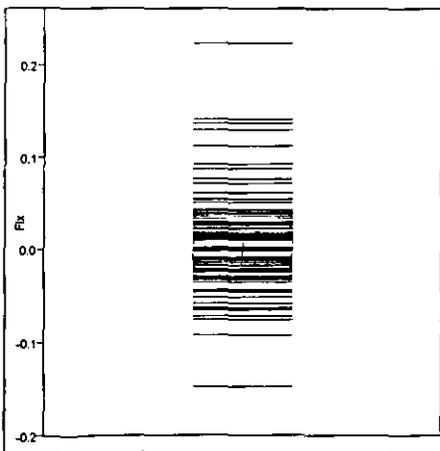
La serie de tiempo (nótese que las pérdidas aparecen como valores positivos) nos permite observar que las pérdidas y las ganancias más extremas ocurrieron entre finales de 1994 y principios de 1995 y que sus magnitudes fueron mucho mayores que las de las otras pérdidas y ganancias extremas que podemos identificar. Asimismo, el valor de la pérdida más grande es mayor que el valor absoluto de la ganancia más grande que se registró en el período 1992-2001.

La siguiente Tabla muestra los principales datos estadísticos de la serie del tipo de cambio FIX. En la segunda columna se toman todos los datos (con pérdidas positivas), en la tercera únicamente las ganancias y en la última sólo las pérdidas.

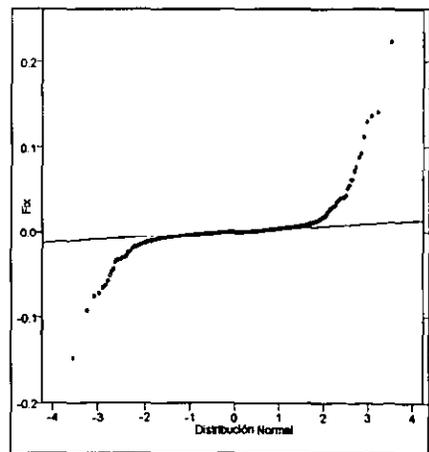
	FIX (Todos)	FIX (Ganancias)	FIX (Pérdidas)
Mínimo	-0.148	0.000	0.000
Primer cuartil	-0.002	0.001	0.001
Media	0.001	0.004	0.005
Mediana	0.000	0.002	0.002
Tercer Cuartil	0.002	0.005	0.005
Máximo	0.223	0.148	0.223
Desv. Estándar	0.012	0.008	0.013
Total Datos	2357	1161	1155

Resumen estadístico datos Tipo de Cambio FIX.

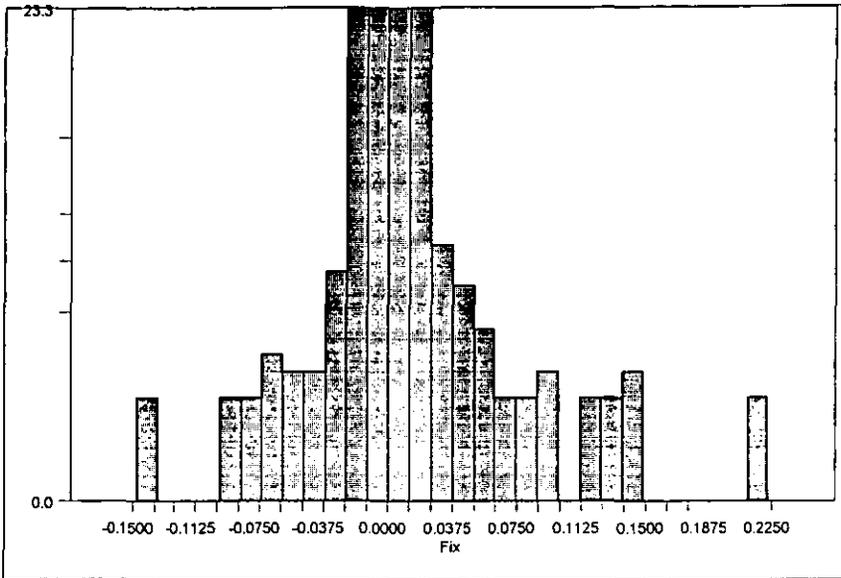
Veamos ahora el Boxplot y el histograma de la serie FIX, así como un QQ-Plot de nuestros datos contra la distribución Normal.



Boxplot Tipo de Cambio FIX.



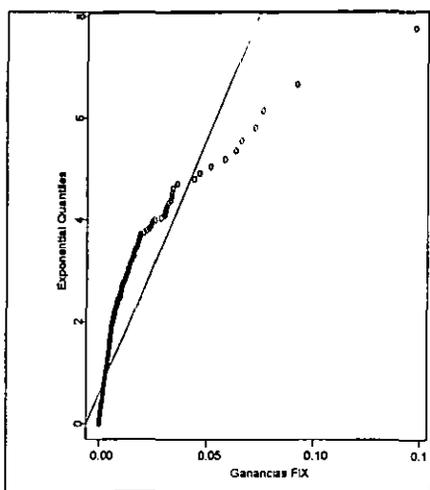
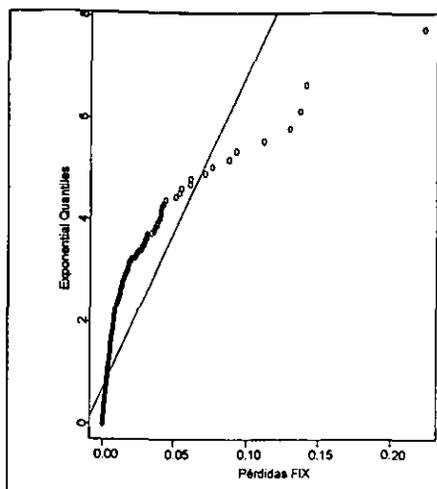
QQ-Plot distrib. Normal contra FIX.



Histograma tipo de cambio FIX.

El Boxplot nos muestra que hay cinco pérdidas por encima de 0.1, mientras que sólo tenemos una ganancia que excede esa cantidad. La distribución parece bastante simétrica respecto al cero, si bien, como aparentemente confirma el histograma, hay cierta evidencia de que la cola derecha (pérdidas) sea más pesada que la cola izquierda. Por su parte, el QQ-Plot contra la distribución Normal nos muestra clara evidencia de que las dos colas de nuestros datos son mucho más pesadas que las colas de la distribución Normal.

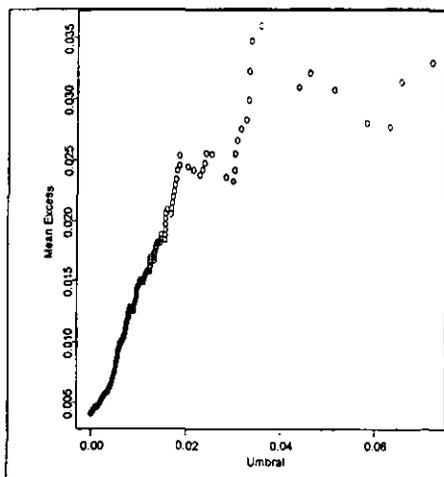
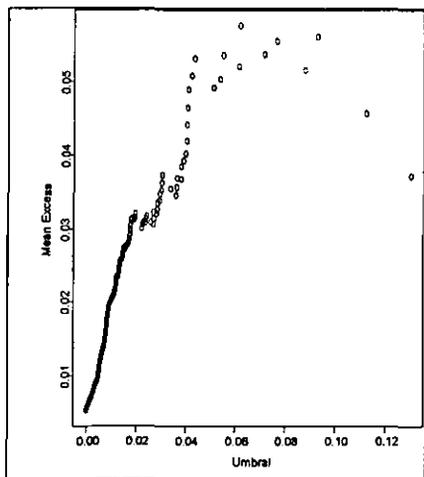
Investigamos a continuación el hecho de que nuestros datos provengan de una distribución exponencial, mediante la creación de dos QQ-Plots, en los cuales se analizará esta posibilidad, por separado, para pérdidas y para ganancias del tipo de cambio FIX.



QQ-Plot Pérdidas FIX vs. Exponencial. QQ-Plot Ganancias FIX vs. Exponencial.

Estos QQ-Plots nos indican que tanto la cola derecha (que, en nuestro caso son las pérdidas) como la cola izquierda (las ganancias) de nuestra serie del tipo de cambio FIX son más pesadas que la cola de una distribución exponencial.

Finalmente, analizamos las gráficas de la función media de exceso (F.M.E.) empírica para pérdidas y para ganancias por separado.



Gráfica F.M.E. para Pérdidas FIX.

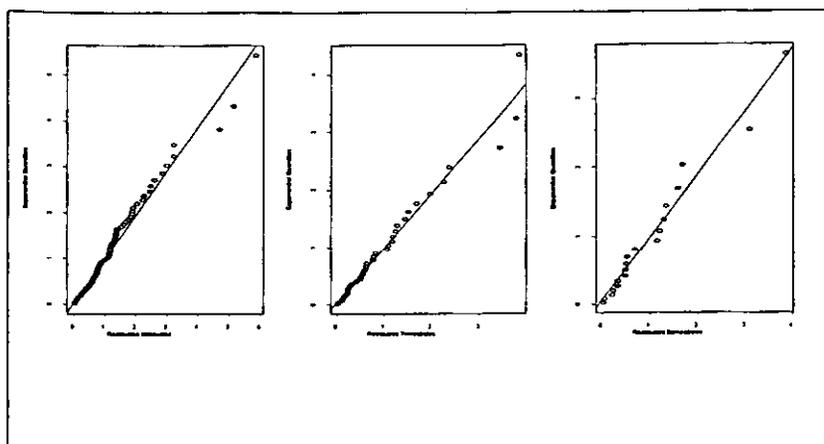
Gráfica F.M.E. para Ganancias FIX.

Las gráficas de la función muestral media de exceso tanto para pérdidas como para ganancias del tipo de cambio FIX muestran claramente que nuestros datos tienen cola pesada. Más aún, la diferencia en las pendientes parece mostrar que la cola de las pérdidas es más pesada que la de las ganancias. La gráfica de la F.M.E. para pérdidas es bastante recta en general, aunque hay un ligero cambio en la pendiente alrededor de 0.008; esta podría ser, por tanto, nuestra primera elección para el umbral en el método POT.

4.4.2 Análisis de Máximos de Bloque para el Tipo de Cambio FIX

El análisis se realiza para bloques mensuales, trimestrales y semestrales. De esta manera, nuestros 2357 datos se convierten en: 113 máximos de bloque, donde el número aproximado de observaciones en cada uno de los bloques es $n \approx 21$ para bloques mensuales; 38 máximos de bloque de tamaño aproximado 62 para bloques trimestrales; y 19 máximos de bloque de alrededor de 124 observaciones cada uno en el caso semestral. Los estimadores del parámetro de forma ξ son, respectivamente, 0.5926, 0.7224 y 0.6292, lo que nos indica que la distribución GEV que se ajusta es del tipo Fréchet y la hipótesis de varianza infinita parece plausible en todos los casos.

Se investigó el ajuste de la Fréchet a los tres distintos tamaños de bloque utilizando los residuales descritos en (4.6). La suposición de exponencialidad de los residuales parece aceptable en todos los casos, como los siguientes QQ-Plots nos muestran.



QQ-Plot de residuales pérdidas FIX vs. distribución Exponencial para bloques mensuales (izq.), trimestrales (centro) y semestrales (der.).

La siguiente Tabla muestra los estimadores de los parámetros obtenidos al ajustar la distribución VEG a máximos de bloque mensuales, trimestrales y semestrales de las pérdidas en el tipo de cambio FIX. Se incluyen asimismo los niveles de retorno de 20-periodos.

	$\approx n$	m	ξ	E.E.	σ	E.E.	μ	E.E.	$R_{n,20}$	R_{20}
Mes	21	113	0.593	0.09	0.004	0.00	0.006	0.00	0.0427	0.1283
Trim.	62	38	0.722	0.20	0.007	0.00	0.010	0.00	0.0886	0.1479
Sem.	124	19	0.629	0.27	0.014	0.00	0.018	0.00	0.1366	0.1366

Estimadores y niveles de retorno (VEG) para Pérdidas en Tipo de Cambio FIX .

Así, puede verse que el nivel de retorno de 20 semestres es 0.1366, de manera que en un semestre cada 10 años esperamos ver pérdidas diarias de alrededor de -0.1366. El estimador del nivel de retorno de 40 trimestres (marcado en la columna R_{20}) es

0.1479, lo que significa que cada 10 años veremos rendimientos logarítmicos negativos de, en promedio, -0.148. Los estimadores de R_{20} para los diferentes tamaños de bloque son similares, si bien no son tan cercanos como los obtenidos en el caso de las pérdidas de BMW.

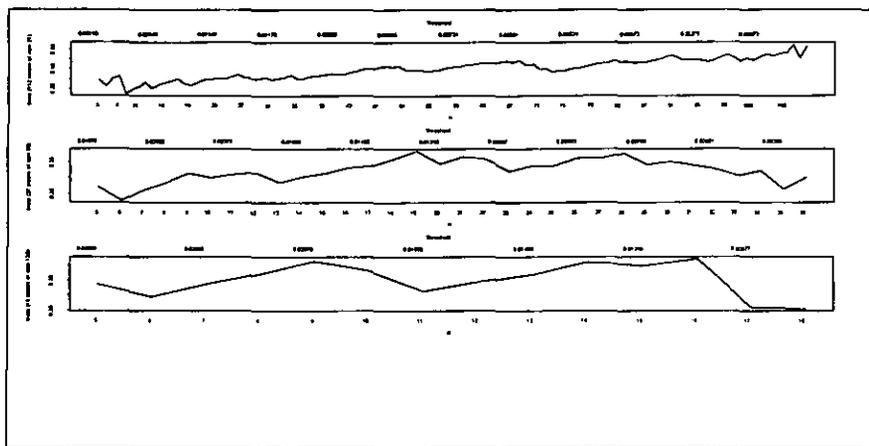
Queremos ahora estimar los cuantiles correspondientes a los niveles de retorno calculados. Para ello, utilizaremos la aproximación (4.11), de manera que necesitamos conocer el índice extremo de la serie para pasar de la información expresada como niveles de retorno a información en términos de cuantiles. Intentamos entonces estimar el índice extremo θ para bloques mensuales, trimestrales y semestrales. La siguiente Tabla muestra los estimadores de θ obtenidos mediante el método de bloques, para máximos mensuales, trimestrales y semestrales, conforme se varía el número N_u de excesos del umbral u .

(m, n)	N_u	15	20	25	30	40	50	100	150	200
	u	0.042	0.040	0.036	0.030	0.024	0.018	0.011	0.008	0.007
mensual	K_u	6	8	9	10	14	18	35	52	67
(113,21)	$\hat{\theta}$	0.411	0.414	0.374	0.348	0.372	0.389	0.412	0.452	0.490
trimestral	K_u	4	5	6	7	11	13	23	29	32
(38,62)	$\hat{\theta}$	0.285	0.270	0.263	0.260	0.327	0.321	0.356	0.370	0.358
semestral	K_u	4	5	6	7	10	11	17	17	18
(19,124)	$\hat{\theta}$	0.296	0.287	0.285	0.287	0.349	0.323	0.415	0.274	0.266

Estimadores del Índice Extremo para Pérdidas en el tipo de cambio FIX.

En la Tabla anterior vemos que los bloques mensuales dan estimadores de entre

0.348 y 0.490, para un promedio de 0.407. Por su parte, los bloques trimestrales dan estimadores entre 0.260 y 0.370, para un promedio de 0.312; mientras que en los bloques semestrales, los estimadores están entre 0.266 y 0.415, para un promedio de 0.309. En el caso de bloques trimestrales y semestrales, los promedio de los estimadores de θ obtenidos son suficientemente parecidos; para bloques mensuales el valor de n ($= 21$) es demasiado pequeño como para utilizar un argumento asintótico. Las siguientes gráficas (nótense las diferentes escalas) muestran que el estimador del índice extremo es más estable cuando se trabaja con bloques trimestrales; esto era de esperarse pues tanto m como n son suficientemente grandes.



Estabilidad del estimador del índice extremo θ para bloques mensuales (arriba), trimestrales (centro) y semestrales (abajo) de las pérdidas del Tipo de Cambio FIX, conforme se varía el umbral u .

Supondremos que nuestros estimadores del índice extremo para datos trimestrales son los más aproximados a la realidad, de manera que tomaremos $\theta = 0.31$ para nuestra serie de pérdidas en el tipo de cambio FIX. Esto significa que el tamaño promedio

de los *clusters* es 3.23. Así, utilizando nuestro modelo para bloques trimestrales, el cuantil 0.995 corresponde al nivel de retorno de 10.887-trimestres. Redondeamos este número 11 y obtenemos que el nivel de retorno de 11-trimestres está dado por 0.0565; de la misma manera, el cuantil 0.999 corresponde aproximadamente al nivel de retorno de 53-trimestres y es 0.1817, mientras que el estimador del cuantil 0.9999 —o el nivel de retorno de 521-trimestres— es 0.9546.

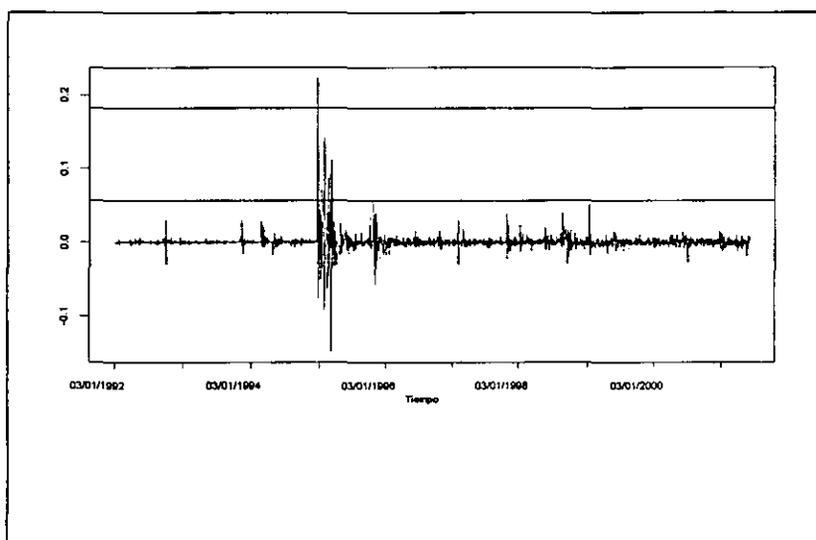
La siguiente Tabla muestra los estimadores de VaR obtenidos para la serie de pérdidas en el tipo de cambio FIX, con los correspondientes niveles de retorno. En todos los casos se considera $\theta = 0.31$.

	$VaR_{0.995}$	$VaR_{0.999}$	$VaR_{0.9999}$
Mensuales	0.0587	0.1537	0.6064
$R_{n,t}$	$R_{21,33}$	$R_{21,162}$	$R_{21,1614}$
Trimestrales	0.0565	0.1817	0.9546
$R_{n,t}$	$R_{62,11}$	$R_{62,53}$	$R_{62,521}$
Semestrales	0.0594	0.1665	0.7140
$R_{n,t}$	$R_{124,6}$	$R_{124,27}$	$R_{124,261}$

Estimadores del VaR (VEG) para Pérdidas FIX.

Sobreimpuestos en la gráfica de la serie de tiempo de las pérdidas en el tipo de cambio FIX se marcaron los estimadores de los cuantiles 0.995 y 0.999 (correspondientes a los niveles de retorno de 11 y de 53-trimestres, respectivamente) mediante rectas horizontales. La longitud total de la serie FIX es 2357, de manera que esperamos que alrededor de 11 o 12 pérdidas excedan el cuantil 0.995 y que aproximadamente 2

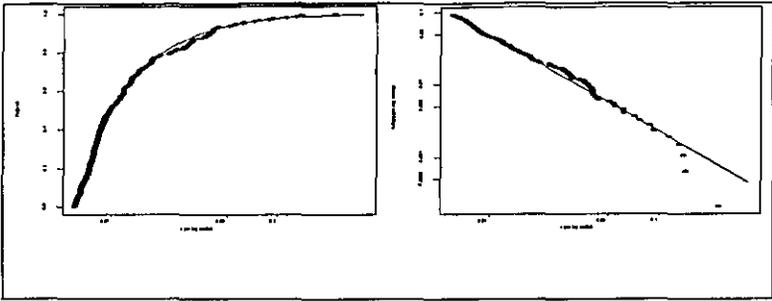
rendimientos excedan el cuantil 0.999. En la práctica, tenemos que 11 rendimientos exceden el cuantil 0.995, lo cual está de acuerdo con nuestra estimación, mientras que hay un rendimiento que excede el cuantil 0.999, que es solamente una vez menos que nuestra estimación.



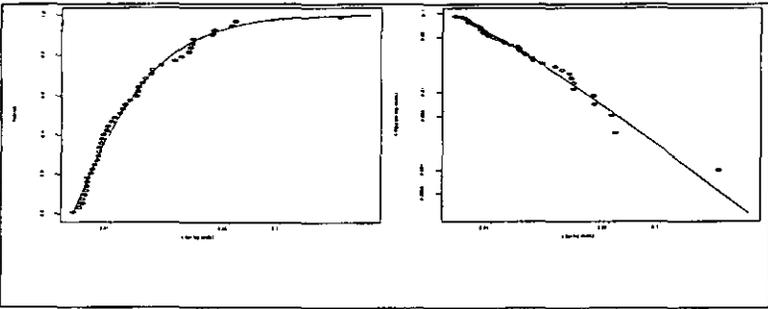
Serie de tiempo pérdidas FIX con estimadores de $\text{VaR}_{0.995}$ y $\text{VaR}_{0.999}$ sobreimpuestos.

4.4.3 Análisis POT para Pérdidas en el Tipo de Cambio FIX

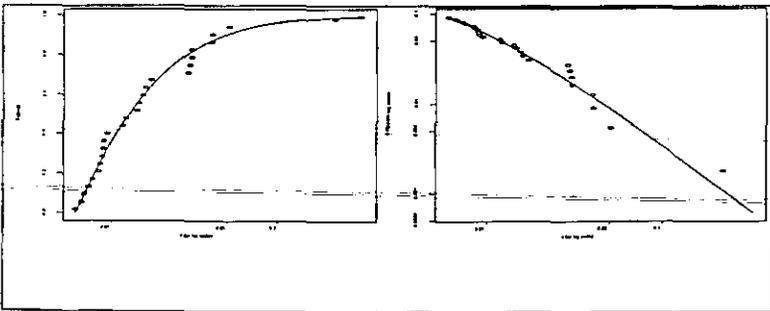
La gráfica de la función media de exceso sugiere que podemos empezar este análisis con un umbral de alrededor de 0.008. Sin embargo, este umbral reduce demasiado el número de excesos cuando se toman $run = 20$ y $run = 30$, de manera que decidimos empezar el análisis con un umbral ligeramente más bajo. Tomaremos pues $u = 0.006$.



$u = 0.006, run = 0$. El estimador de ξ es 0.745.



$u = 0.006, run = 20$. El estimador de ξ es 0.598.



$u = 0.006, run = 30$. El estimador de ξ es 0.674.

Las gráficas anteriores muestran que el ajuste de la DPG parece aceptable para los casos $run = 0$ y $run = 20$. Cuando $run = 30$ son únicamente 26 los excesos que efectivamente se toman en cuenta; esta es una cantidad de datos demasiado pequeña,

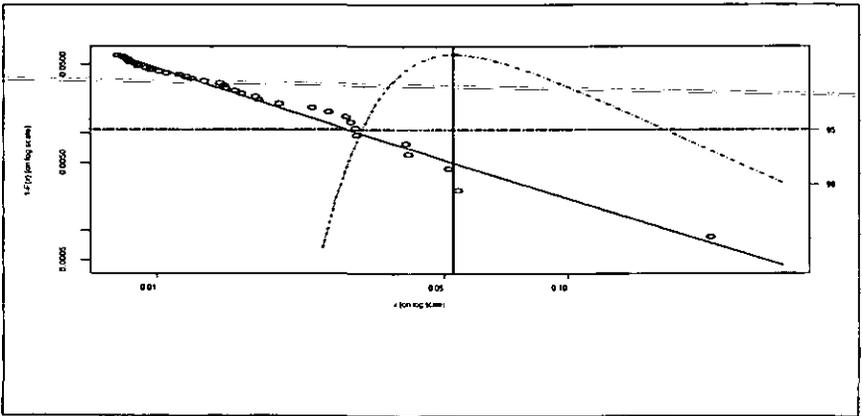
por lo que el ajuste obtenido en este caso no es tan confiable. Los estimadores del parámetro de forma ξ no son muy estables para $u = 0.006$ al variar run . La Tabla que se presenta a continuación nos permite observar que los estimadores del parámetro ξ son más estables para umbrales más altos, a pesar de que se usan muy pocos excesos para realizar los ajustes a la DPG cuando $run \neq 0$.

u	0.005	0.006	0.008	0.009	0.01	0.013
run = 0						
ξ	0.8179	0.7449	0.7778	0.5609	0.5686	0.5036
E.E.	0.110	0.120	0.165	0.152	0.164	0.191
β	0.003	0.005	0.006	0.009	0.010	0.013
Excesos	292	219	148	118	107	79
run = 20						
ξ	0.9041	0.5981	0.7692	0.4778	0.5294	0.5013
E.E.	0.318	0.235	0.339	0.270	0.295	0.313
β	0.004	0.006	0.005	0.011	0.010	0.012
Excesos	48	46	36	27	26	20
run = 30						
ξ	0.5653	0.6741	0.7619	0.4737	0.4517	0.5266
E.E.	0.332	0.340	0.395	0.300	0.300	0.361
β	0.010	0.008	0.007	0.013	0.014	0.013
Excesos	23	26	26	22	20	17

Estimadores de los Parámetros DPG para varios umbrales y $runs$.

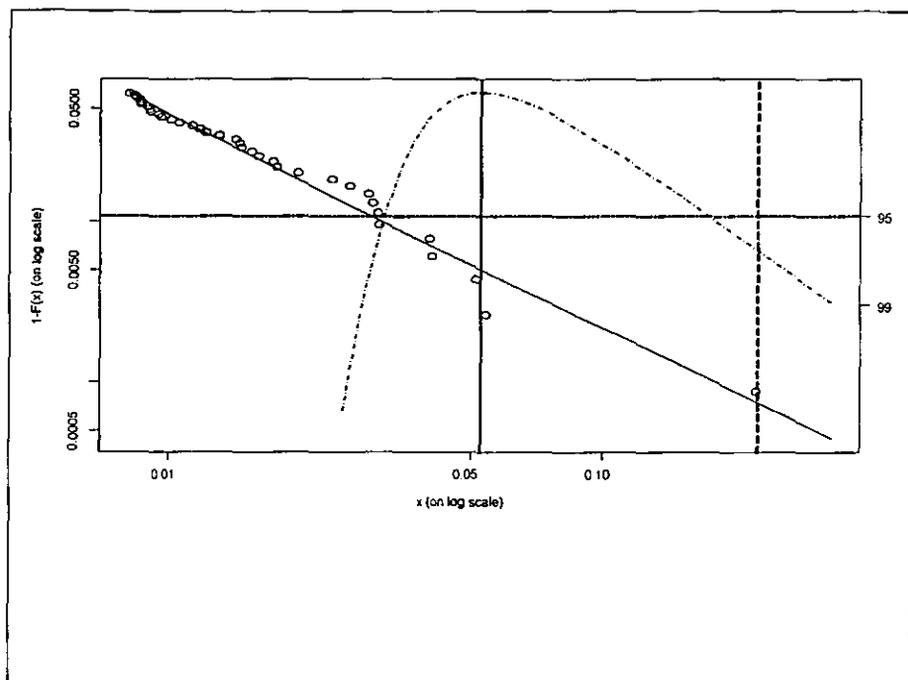
Al fijar *run* —es decir, el número de observaciones por debajo del umbral que separarán dos *clusters*— y variar el umbral, podemos notar que en ninguno de los casos puede decirse que hay estabilidad de los estimadores del parámetro de forma ξ . Para *run* = 30, el promedio es de 0.5756, para *run* = 20 se obtiene 0.6299, mientras que el promedio para *run* = 0 es 0.6623.

Ilustraremos el cálculo del VaR para $u = 0.008$, *run* = 20; sin tomar en cuenta que los excesos se presentan en *clusters*, estaríamos hablando de aproximadamente 148 excesos del umbral, pero al escoger *run* = 20, se toman en cuenta solamente 36 excesos del umbral. En este caso, obtenemos que el estimador del $VaR_{0.999}$ para las pérdidas en el tipo de cambio FIX es 0.0524. Esto se muestra en la siguiente figura mediante una recta punteada vertical. El intervalo de confianza (asimétrico) del 95% puede leerse en las intersecciones de la curva punteada y la línea punteada horizontal: (0.0315, 0.1717). Obsérvese que se obtiene un intervalo de confianza del 99% más amplio si se baja la línea horizontal al valor 99 del eje de confianza (eje-*y*, lado derecho).



Estimador de $VaR_{0.995}$ para $u = 0.008$ y *run* = 20.

El estimador de $ES_{0.995}$ se ilustra en la siguiente Figura mediante una segunda recta vertical, a la derecha de la correspondiente al $VaR_{0.995}$. El estimador puntual que se obtiene es 0.2251.



Estimador de $ES_{0.995}$ con $u = 0.008$ y $run = 20$.

La siguiente Tabla muestra los estimadores de VaR y de ES para tres distintos valores del umbral (0.006, 0.008 y 0.010), donde separamos los *clusters* por 0, 20 y por 30 observaciones por debajo del umbral. Obsérvese que todos los estimadores son muy altos para $u = 0.008$ y $u = 0.006$ cuando $run = 30$.

u	run	$VaR_{0.995}$	$VaR_{0.999}$	$VaR_{0.9999}$	$ES_{0.995}$	$ES_{0.999}$
0.006	0	0.05464107	0.1817237	1.010989	0.2148426	0.7130112
0.008	0	0.05576991	0.1944254	1.164584	0.2502175	0.8744287
0.010	0	0.05419454	0.1467676	0.5642077	0.1356973	0.3502925
0.006	20	0.05539596	0.1522037	0.6163804	0.1444086	0.3852844
0.008	20	0.05241205	0.1792757	1.05077	0.2250805	0.774745
0.010	20	0.05094708	0.1308515	0.4630129	0.1178088	0.2876071
0.006	30	0.07556453	0.2339585	1.124417	0.2428026	0.7288832
0.008	30	0.05907985	0.2030319	1.176713	0.2503317	0.8548645
0.010	30	0.06127548	0.1481615	0.4558557	0.1282469	0.2867228

Estimadores de VaR y ES para serie FLX, usando distintos umbrales y $runs$.

Los estimadores de $VaR_{0.995}$ son bastante estables para variaciones del umbral y de run . En la práctica, hemos visto entre 14 y 11 excesos de $VaR_{0.995}$, excepto en el caso en que $u = 0.006$ y $run = 30$, en el cual sólo hay 8 excesos. Puesto que la serie FLX consta de 2357 observaciones, esperamos aproximadamente 11 ó 12 excesos del cuantil 0.995. En cuanto a los estimadores de $VaR_{0.999}$, hemos visto 3 excesos del correspondiente a $u = 0.010$ y $run = 20$ y sólo uno en los demás casos, excepto para el estimador $\widehat{VaR}_{0.999}$ obtenido cuando $u = 0.006$ y $run = 30$, que no ha sido excedido históricamente. Por su parte, los estimadores de $VaR_{0.9999}$ parecen tener sentido solamente cuando se toma $u = 0.010$, y tal vez también cuando se tiene $u = 0.006$ y $run = 20$. Aún así, en la serie FLX nunca hemos visto una pérdida que exceda a ninguno de los estimadores de $VaR_{0.9999}$, lo cual es de esperarse, puesto que

estamos hablando del tamaño que excederá, en promedio, una pérdida entre 10,000.

Los estimadores que mejor se adaptan a lo visto en la serie histórica son los correspondientes a $u = 0.010$, $run = 20$. En este caso incluso la Esperanza Condicional de la Cola (que nos indica los tamaños esperados de las pérdidas dado que se ha excedido el VaR) no parece estar tan sobreestimada como en los otros casos: el promedio de los 14 excesos de $VaR_{0.995}$ es 0.0967, contra un pronóstico de 0.1178, y el promedio de los 3 excesos de $VaR_{0.999}$ es 0.1671, contra un pronóstico de 0.2876. Nótese que la mayor observación dentro de la serie de pérdidas en el tipo de cambio FIX toma el valor 0.2228.

Los estimadores de $ES_{0.999}$ parecen estar menos sobreestimados cuando se toma un umbral suficientemente alto (en este caso, al tomar $u = 0.010$). Sin embargo, aún en esta situación, los estimadores $\widehat{ES}_{0.999}$ que se obtienen son, aparentemente, bastante elevados en comparación con lo que hemos visto en la serie FIX.

4.4.4 Comparación de Cuantiles Obtenidos con el Método Máximos de Bloque y con el Método POT para Pérdidas en el Tipo de Cambio FIX

La siguiente Tabla presenta una comparación entre los estimadores de $VaR_{0.995}$ y $VaR_{0.999}$ obtenidos por el Método de Máximos de Bloque y los obtenidos por el Método POT, para pérdidas en el Tipo de Cambio FIX. El umbral del método POT se escoge de manera que haya el mismo número de excesos de ese umbral que el número de máximos con que se está comparando. Así, por ejemplo, hay 113 excesos

del umbral $u = 0.0094$, de manera que los cuantiles obtenidos mediante el método MDB para bloques mensuales se comparan con los obtenidos a través del método POT usando el umbral $u = 0.0094$. Cuando $run \neq 0$, el número efectivo de excesos se reduce para tomar en cuenta la tendencia de la serie a formar *clusters*; así, por ejemplo, para $run = 20$, los 113 excesos se convierten en 27 excesos efectivos, y para $run = 30$ en sólo 21 excesos efectivos.

Max de Bloque	Mes	Trimestre	Semestre
Número de Máximos	113	38	19
$\hat{\xi}$	0.593	0.722	0.629
$\widehat{VaR}_{0.995}$	0.0587	0.0565	0.0594
$\widehat{VaR}_{0.999}$	0.1537	0.1817	0.1665
POT	$u = 0.0094$	$u = 0.027$	$u = 0.040$
$(Run = 0) \hat{\xi}$	0.561	0.491	0.257
$\widehat{VaR}_{0.995}$	0.0542	0.0546	0.0566
$\widehat{VaR}_{0.999}$	0.1457	0.1302	0.1287
$(Run = 20) \hat{\xi}$	0.503	0.816	1.331
$\widehat{VaR}_{0.995}$	0.0517	0.0465	0.0474
$\widehat{VaR}_{0.999}$	0.1294	0.1325	0.1578
$(Run = 30) \hat{\xi}$	0.460	0.816	1.331
$\widehat{VaR}_{0.995}$	0.0614	0.0465	0.0474
$\widehat{VaR}_{0.999}$	0.1495	0.1325	0.1578

Comparación entre el POT y el Modelo MDB, serie FIX.

Los estimadores de $VaR_{0,995}$ con $run = 0$ son muy similares a los obtenidos mediante el método de Máximos de Bloque, pero los de $VaR_{0,999}$ de Máximos de Bloque se parecen más a los obtenidos en el POT con $run = 30$. En el caso de bloques semestrales hay que recordar que, al tomar 19 máximos con $run = 20$ o $run = 30$, en realidad estamos reduciendo los datos a una cantidad mínima (4 excesos efectivos), por lo que los estimadores no deberían ser demasiado confiables.

En la práctica, para los estimadores $\widehat{VaR}_{0,999} = 0.1287$ y 0.1294 hay 4 excesos; para 0.1302 y 0.1325 hay 3 excesos; y para los estimadores entre 0.1457 y 0.1817 hay 1 exceso. Por su parte, para el estimador $\widehat{VaR}_{0,995} = 0.0614$ hay 10 excesos históricos; para los estimadores entre 0.0565 y 0.0594 hay 11 excesos; para 0.0542 y 0.0546 hay 12 excesos en nuestros datos; y para 0.0465 y 0.0474 hay 14 excesos. De acuerdo con la longitud de la serie FIX (2357 datos), esperamos alrededor de 2 excesos del cuantil 0.999 y aproximadamente 11 o 12 del cuantil 0.995 .

4.5 Conclusión

Los desastres naturales o causados por el hombre, los *crashes* de la Bolsa y otros eventos extremos forman parte de la sociedad actual. La metodología expuesta en este trabajo puede ser útil para llevar a cabo el análisis de estos eventos.

El reconocimiento de que una distribución tiene cola pesada es un hecho clave para la medición de los riesgos extremos. Más aún, existe la necesidad de evaluar los riesgos extremos dentro de los mercados financieros. El BIS (*Bank for International Settlement*) ha decretado reglas que los bancos deben seguir para controlar sus riesgos,

pero la mayoría de los modelos actuales para la medición de riesgos se basan en la suposición de que los activos financieros siguen una distribución Gaussiana.

En este trabajo hemos dicho que, si nos interesan las colas de alguna distribución (es decir, los eventos extremos), es natural considerar la aplicación de los métodos de la Teoría de Valores Extremos (TVE). Los métodos basados en la suposición de normalidad de las distribuciones tienden a subestimar el riesgo del área de la cola. Los métodos basados en Simulación Histórica nos ofrecen estimadores muy imprecisos del riesgo extremo. En cambio, tanto el Modelo de Máximos de Bloque (MDB) como el Modelo de Picos Sobre el Umbral (POT) tienen la base científica de la TVE; esta base teórica nos permite enfrentarnos a un problema inherentemente difícil: la predicción del tamaño de un evento extremo.

En este Capítulo hemos demostrado la aplicación de los Modelos MDB y POT a tres conjuntos de datos; uno de ellos pertenece al área de seguros, mientras que los otros dos corresponden a datos financieros. En todos los casos vimos que el ajuste de la Distribución Pareto Generalizada (DPG) a las pérdidas que exceden un umbral alto es un método útil para la estimación de las colas, los altos cuantiles (VaR), y la Esperanza Condicional de la Cola (ES) de las distribuciones. Vimos también que podemos utilizar el ajuste de la distribución de Valores Extremos Generalizada (VEG) a los máximos de bloques de tamaño suficientemente grande para obtener estimadores de la cola y de los altos cuantiles (VaR) de la distribución subyacente. Más aún, pudimos constatar que la implementación de los Modelos MDB y POT a un sistema de medición de riesgos es bastante sencilla.

Como hemos dicho antes, tanto el Modelo MDB como el POT están sólidamente

fundamentados en la teoría matemática del comportamiento de los extremos; no es simplemente cuestión de ajustar una curva *ad hoc*. Bien podría suceder que, al tanteo, se encuentre alguna otra distribución que se ajuste aún mejor al área de la cola de los datos que estemos considerando. Pero tal distribución sería una elección arbitraria, de manera que tendríamos mucho menos confianza al momento de extrapolar más allá del rango de los datos para obtener medidas de riesgo extremo.

Esperamos haber mostrado la utilidad de los Modelos de la Teoría de Valores Extremos para el análisis de datos de cola pesada y la obtención de medidas de riesgo extremo; queremos recalcar particularmente la posibilidad de obtener estimadores de la Esperanza Condicional de la Cola (ES) que, como se mencionó anteriormente, es una medida de riesgo coherente.

Apéndice A

Convergencia

Definición A.1 (Variable aleatoria) Una variable aleatoria (v.a) X es una función que asocia un número $X(\omega)$ a cada resultado ω , elemento del espacio muestral Ω .

La definición de los conceptos de convergencia para v.a. se basa en la manipulación de sucesiones de eventos que requieren límites de conjuntos. Sea $A_n \subset \Omega$. Definimos

$$\inf_{k \geq n} A_k := \bigcap_{k=n}^{\infty} A_k, \quad \sup_{k \geq n} A_k := \bigcup_{k=n}^{\infty} A_k.$$

Definición A.2 (liminf) Diremos que

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

Puede demostrarse que

$$\liminf_{n \rightarrow \infty} A_n = \lim_{n \rightarrow \infty} \left(\inf_{k \geq n} A_k \right).$$

Para \liminf se tiene la siguiente interpretación:

$$\liminf_{n \rightarrow \infty} A_n = \{ \omega : \omega \in A_n \text{ para toda } n \text{ excepto por un número finito} \}$$

$$\begin{aligned}
&= \left\{ \omega : \sum_n \mathbf{1}_{A_n^c}(\omega) < \infty \right\} \\
&= \{ \omega : \omega \in A_n, \forall n \geq n_0(\omega) \}.
\end{aligned}$$

Definición A.3 (*limsup*) Se tiene que

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Puede demostrarse que

$$\limsup_{n \rightarrow \infty} A_n = \lim_{n \rightarrow \infty} \left(\sup_{k \geq n} A_k \right).$$

Para lim sup se tiene la interpretación siguiente:

$$\begin{aligned}
\limsup_{n \rightarrow \infty} A_n &= \left\{ \omega : \sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega) = \infty \right\} \\
&= \{ \omega : \omega \in A_{n_k}, k = 1, 2, \dots \}
\end{aligned}$$

para alguna subsucesión n_k que depende de ω .

Definición A.4 (*Convergencia Débil o en Distribución*) Sea U_0, U_1, U_2, \dots una sucesión de variables aleatorias (v.a.). Decimos que (U_n) converge en distribución débilmente a la v.a. U_0 (lo cual denotamos $U_n \xrightarrow{d} U_0$) si para toda función continua y acotada f se da la relación

$$E[f(U_n)] \rightarrow E[f(U_0)], \quad n \rightarrow \infty.$$

La convergencia en distribución también puede describirse en términos de las funciones de distribución (f.d.) F_n y F_0 de U_n y U_0 , respectivamente: $U_n \xrightarrow{d} U_0$ se da si y sólo si para todos los puntos de continuidad y de la f.d. F_0 se satisface la relación

$$F_n(y) \rightarrow F_0(y), \quad n \rightarrow \infty. \tag{A.1}$$

Más aún, si F_0 es continua entonces (A.1) puede incluso fortalecerse a convergencia uniforme:

$$\sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)| \rightarrow 0, \quad n \rightarrow \infty. \quad (\text{A.2})$$

En efecto, sabemos que si f_n , $n \geq 0$ son funciones de valores reales en \mathbb{R} (o en cualquier otro espacio métrico) entonces f_n converge uniformemente a f_0 en $A \subset \mathbb{R}$ si

$$\sup_{x \in A} |f_n(x) - f_0(x)| \rightarrow 0$$

cuando $n \rightarrow \infty$.

Si U_n , $n \geq 0$ son funciones reales no-decrecientes y si U_0 es continua y $U_n(x) \rightarrow U_0(x)$ cuando $n \rightarrow \infty$ para todo $x \in \mathbb{R}$, entonces $U_n \rightarrow U_0$ localmente uniformemente; es decir, para cualesquiera $a < b$

$$\sup_{x \in [a, b]} |U_n(x) - U_0(x)| \rightarrow 0. \quad (\text{A.3})$$

El siguiente argumento muestra que si U_0 es continua en $[a, b]$, entonces es uniformemente continua: para cualquier x existe una vecindad 0_x que contiene a x y en la cual U oscila por menos que ε . De esta manera, podemos construir una cubierta abierta \mathcal{C} de $[a, b]$. Por la compacidad de $[a, b]$ podemos obtener una subcubierta finita a partir de $\mathcal{C} = \{0_x : x \in [a, b]\}$. Usando esta subcubierta finita y la monotonía de las funciones se obtiene directamente la convergencia uniforme en $[a, b]$.

Si F_n , $n \geq 0$ son funciones de distribución (no-degeneradas) en \mathbb{R} , entonces la convergencia puntual de F_n a F_0 y la continuidad de F_0 implican la convergencia uniforme en \mathbb{R} . La convergencia uniforme local se obtiene de (A.3) y afuera de un intervalo grande $[a, b]$ no hay muchas posibilidades de oscilación: dado $\varepsilon > 0$ se escoge

b tal que $F_0(b) > 1 - \varepsilon$; por la convergencia puntual, existe $n_0 \in \mathbb{N}$ tal que para $n \geq n_0$

$$|F_n(b) - F_0(b)| < \varepsilon.$$

De este modo, para $x \geq b$

$$|F_n(x) - F_n(b)| \leq 1 - F_n(b) \leq 1 - F_0(b) + |F_0(b) - F_n(b)| < 2\varepsilon.$$

Así, para $n \geq n_0$

$$\begin{aligned} \sup_{x > b} |F_n(x) - F_0(x)| &\leq \sup_{x > b} |F_n(x) - F_n(b)| + |F_n(b) - F_0(b)| + |F_0(b) - F_0(x)| \\ &\leq 2\varepsilon + \varepsilon + \varepsilon. \end{aligned}$$

Para $x < a$ se procede de manera similar. Al combinar estos resultados con el de convergencia uniforme en $[a, b]$ se obtiene la convergencia uniforme en \mathbb{R} , es decir, la expresión (A.2).

Definición A.5 (Convergencia en Probabilidad) Decimos que la sucesión de v.a. (A_n) converge en probabilidad a la v.a. A (lo cual denotamos $A_n \xrightarrow{P} A$) si para todo $\varepsilon > 0$ se da la relación

$$P(|A_n - A| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty.$$

La convergencia en probabilidad implica la convergencia en distribución. El recíproco es cierto si y sólo si $A = a$ casi seguramente para alguna constante a .

La relación $A_n \xrightarrow{P} \infty$ debe interpretarse como $1/A_n \xrightarrow{P} 0$.

En un espacio métrico S la convergencia $a_n \rightarrow a$ para los elementos a, a_1, a_2, \dots es equivalente al *principio de la subsucesión*: toda subsucesión (a_{n_k}) contiene a su vez

una subsucesión $(a_{n_{k_j}})$ que converge a a . Así, $A_n \xrightarrow{P} A$ si y sólo si toda subsucesión $(A_{n_{k_j}})$ contiene una subsucesión $(A_{n_{k_{j_l}}})$ que converge en probabilidad a A .

Definición A.6 (Convergencia Casi Segura) Decimos que (A_n) converge casi seguramente (c.s.) o con probabilidad 1 a la v.a. A (lo cual denotamos $A_n \xrightarrow{c.s.} A$) si la relación

$$A_n(\omega) \rightarrow A(\omega), \quad n \rightarrow \infty$$

se da para P -casi todo $\omega \in \Omega$.

Esto significa que

$$P(A_n \rightarrow A) = P[\{\omega : A_n(\omega) \rightarrow A(\omega)\}] = 1.$$

La convergencia casi segura es equivalente a la relación

$$\sup_{k \geq n} |A_k - A| \xrightarrow{P} 0.$$

Así, la convergencia casi segura implica la convergencia en probabilidad y, por tanto, la convergencia en distribución.

La relación $A_n \xrightarrow{c.s.} \infty$ debe interpretarse como $1/A_n \xrightarrow{c.s.} 0$.

Definición A.7 (O(1)) $a(x) = O(b(x))$ cuando $x \rightarrow x_0$ significa que

$$\limsup_{x \rightarrow x_0} \left| \frac{a(x)}{b(x)} \right| < \infty.$$

Definición A.8 (o(1)) $a(x) = o(b(x))$ cuando $x \rightarrow x_0$ significa que

$$\lim_{x \rightarrow x_0} \frac{a(x)}{b(x)} = 0.$$

Teorema A.9 (del Límite Central) Sean X_1, X_2, \dots, X_n observaciones mutuamente independientes de una variable aleatoria X con media μ_X y varianza $\sigma_X^2 < \infty$. Sea

$$Z_n = \frac{\left(\frac{1}{n}\right) \sum_{i=1}^n X_i - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} = \frac{\left(\frac{1}{n}\right) S_n - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} = \frac{S_n - n\mu_X}{\sigma_X \sqrt{n}}$$

y sea $F_{Z_n}(z)$ la función de distribución (acumulativa) de la variable aleatoria Z_n .

Entonces, para $z \in (-\infty, \infty)$ se tiene que

$$\lim_{n \rightarrow \infty} F_{Z_n}(z) = F_Z(z)$$

donde $F_Z(z)$ es la función de distribución acumulativa de $Z \sim N(0, 1)$.

Teorema A.10 (Convergencia a Tipos). Sean X_1, X_2, \dots v.a. con función de distribución F_n para $n \geq 1$, y sean U, V v.a. con f.d. $U(x)$ y $V(x)$ respectivamente, que no se concentran en un punto. Sean $a_n > 0$, $\alpha_n > 0$, $b_n \in \mathbb{R}$, $\beta_n \in \mathbb{R}$ constantes.

(a) Si

$$F_n(a_n x + b_n) \xrightarrow{d} U(x) \quad \text{y} \quad F_n(\alpha_n x + \beta_n) \xrightarrow{d} V(x) \quad (\text{A.4})$$

o, equivalentemente

$$\frac{X_n - b_n}{a_n} \xrightarrow{d} U \quad \text{y} \quad \frac{X_n - \beta_n}{\alpha_n} \xrightarrow{d} V \quad (\text{A.5})$$

entonces existen constantes $A > 0$, $B \in \mathbb{R}$ tales que

$$\lim_{n \rightarrow \infty} \frac{\alpha_n}{a_n} = A > 0 \quad \text{y} \quad \lim_{n \rightarrow \infty} \frac{\beta_n - b_n}{a_n} = B \in \mathbb{R} \quad (\text{A.6})$$

y

$$V(x) = U(Ax + B), \quad V \stackrel{d}{=} \frac{U - B}{A}. \quad (\text{A.7})$$

(b) Si se da (A.6), entonces cualquiera de las relaciones de (A.4) implica la otra y se da (A.7).

Demostración:

b) Supongamos que $G_n(x) := F_n(a_n x + b_n) \xrightarrow{d} U(x)$ y $\frac{\alpha_n}{a_n} \xrightarrow{n \rightarrow \infty} A > 0$, $\frac{\beta_n - b_n}{a_n} \xrightarrow{n \rightarrow \infty} B \in \mathbb{R}$. Entonces $G_n\left(\frac{\alpha_n}{a_n}x + \frac{\beta_n - b_n}{a_n}\right) = F_n\left(a_n\left(\frac{\alpha_n}{a_n}x + \frac{\beta_n - b_n}{a_n}\right) + b_n\right) = F_n((\alpha_n x + \beta_n - b_n) + b_n) = F_n(\alpha_n x + \beta_n)$. Sea x un punto de continuidad de la función U al aplicarse en $(A \cdot + B)$, es decir, $x \in C(U(A \cdot + B))$.

Supongamos que $x > 0$ (el razonamiento es similar para $x \leq 0$). Sea $\epsilon > 0$.

Entonces, para n grande, se tiene que

$$\begin{aligned} \left| \frac{\alpha_n}{a_n} - A \right| &\leq \epsilon, & \left| \frac{\beta_n - b_n}{a_n} - B \right| &\leq \epsilon \\ \Rightarrow -\epsilon &\leq \frac{\alpha_n}{a_n} - A \leq \epsilon, & -\epsilon &\leq \frac{\beta_n - b_n}{a_n} - B \leq \epsilon \\ \Rightarrow A - \epsilon &\leq \frac{\alpha_n}{a_n} \leq A + \epsilon, & B - \epsilon &\leq \frac{\beta_n - b_n}{a_n} \leq B + \epsilon. \end{aligned}$$

De donde

$$(A - \epsilon)x + (B - \epsilon) \leq \frac{\alpha_n}{a_n}x + \frac{\beta_n - b_n}{a_n} \leq (A + \epsilon)x + (B + \epsilon).$$

Entonces

$$\limsup_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) = \limsup_{n \rightarrow \infty} G_n\left(\frac{\alpha_n}{a_n}x + \frac{\beta_n - b_n}{a_n}\right) \leq \limsup_{n \rightarrow \infty} G_n((A + \epsilon)x + (B + \epsilon)).$$

Así, $\forall z \in C(U(\cdot))$ tal que $z > (A + \epsilon)x + (B + \epsilon)$, se tiene que

$$\limsup_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) \leq \limsup_{n \rightarrow \infty} G_n(z) = U(z),$$

donde la última igualdad se da por la hipótesis (A.4). Entonces

$$\limsup_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) \leq \inf_{z > (A + \epsilon)x + (B + \epsilon), z \in C(U(\cdot))} U(z).$$

Como $\epsilon > 0$ es arbitrario, puede hacerse tan pequeño como se desee, por lo que

$$\limsup_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) \leq \inf_{z > Ax+B, z \in C(U(\cdot))} U(z) = U(Ax+B)$$

por ser U función de distribución y, por tanto, continua por la derecha. Además

$$\liminf_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) = \liminf_{n \rightarrow \infty} G_n\left(\frac{\alpha_n}{a_n}x + \frac{\beta_n - b_n}{a_n}\right) \geq \liminf_{n \rightarrow \infty} G_n((A-\epsilon)x + (B-\epsilon)).$$

Entonces, $\forall z \in C(U(\cdot))$ tal que $z < (A-\epsilon)x + (B-\epsilon)$, se tiene que

$$\liminf_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) \geq \liminf_{n \rightarrow \infty} G_n(z) = U(z),$$

donde la última igualdad se obtiene de la hipótesis (A.4). Puesto que esto se da para toda $\epsilon > 0$, y ya que $Ax+B \in C(U(\cdot))$

$$\liminf_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) \geq \sup_{z < (A-\epsilon)x + (B-\epsilon), z \in C(U(\cdot))} U(z) = U(Ax+B).$$

Tenemos entonces

$$\limsup_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n) \leq U(Ax+B) \leq \liminf_{n \rightarrow \infty} F_n(\alpha_n x + \beta_n).$$

Por tanto,

$$F_n(\alpha_n x + \beta_n) \xrightarrow{d} U(Ax+B) = V(x).$$

a) Supongamos que $F_n(a_n x + b_n) \xrightarrow{d} U(x)$ y $F_n(\alpha_n x + \beta_n) \xrightarrow{d} V(x)$. Sabemos que si $G_n \xrightarrow{d} G$, entonces también $G_n^- \xrightarrow{d} G^-$. Así, se tiene que

$$\begin{aligned} \frac{F_n^-(y) - b_n}{a_n} &\longrightarrow U^-(y), \quad y \in C(U^-) \\ \frac{F_n^-(y) - \beta_n}{\alpha_n} &\longrightarrow V^-(y), \quad y \in C(V^-). \end{aligned}$$

Puesto que $U(x)$ y $V(x)$ no se concentran en un punto, podemos encontrar $y_1 < y_2$ tales que $y_i \in [C(U^-) \cap C(V^-)]$ para $i = 1, 2$ y $-\infty < U^-(y_1) < U^-(y_2) < \infty$, $-\infty < V^-(y_1) < V^-(y_2) < \infty$. Entonces, para $i = 1, 2$ se tiene

$$\frac{F_n^-(y_i) - b_n}{a_n} \rightarrow U^-(y_i), \quad \frac{F_n^-(y_i) - \beta_n}{\alpha_n} \rightarrow V^-(y_i). \quad (\text{A.8})$$

Al restar las expresiones de (A.8) con $i = 1$ de las de $i = 2$, se obtiene

$$\frac{F_n^-(y_2) - b_n}{a_n} - \frac{F_n^-(y_1) - b_n}{a_n} \rightarrow U^-(y_2) - U^-(y_1),$$

es decir,

$$\frac{F_n^-(y_2) - F_n^-(y_1)}{a_n} \rightarrow U^-(y_2) - U^-(y_1)$$

y

$$\frac{F_n^-(y_2) - F_n^-(y_1)}{\alpha_n} \rightarrow V^-(y_2) - V^-(y_1).$$

Al dividir

$$\frac{\frac{F_n^-(y_2) - F_n^-(y_1)}{a_n}}{\frac{F_n^-(y_2) - F_n^-(y_1)}{\alpha_n}} = \frac{a_n}{\alpha_n} \rightarrow \frac{U^-(y_2) - U^-(y_1)}{V^-(y_2) - V^-(y_1)}.$$

También de (A.8) obtenemos

$$\frac{F_n^-(y_1) - b_n}{a_n} \rightarrow U^-(y_1) \quad \text{y} \quad \frac{F_n^-(y_1) - \beta_n}{\alpha_n} = \frac{F_n^-(y_1) - \beta_n}{\alpha_n} \cdot \frac{\alpha_n}{a_n} \rightarrow V^-(y_1) A.$$

Restando estas dos últimas expresiones se tiene

$$\frac{F_n^-(y_1) - b_n}{a_n} - \frac{F_n^-(y_1) - \beta_n}{\alpha_n} = \frac{\beta_n - b_n}{\alpha_n} \rightarrow U^-(y_1) - V^-(y_1) A := B.$$

Así, vemos que se da (A.6). Por la parte b), el que se den (A.6) y (A.4) implica que se da (A.7). ■

Observación A.11 (Ecuaciones de Cauchy y de Hamel) Si $f(x)$, $x > 0$ es finita, medible y de valores reales y satisface la ecuación de Cauchy

$$f(x+y) = f(x) + f(y), \quad x > 0, y > 0,$$

entonces f es necesariamente de la forma

$$f(x) = cx, \quad x > 0,$$

para algún $c \in \mathbb{R}$. Una variante de esto es la ecuación de Hamel. Si $\phi(x)$, $x > 0$ es finita, medible, de valores reales y satisface la ecuación de Hamel

$$\phi(xy) = \phi(x)\phi(y), \quad x > 0, y > 0,$$

entonces ϕ es de la forma

$$\phi(x) = x^\rho,$$

para algún $\rho \in \mathbb{R}$.

Apéndice B

Variación Regular

Definición B.1 (*Variación Lenta*) Una función positiva, Lebesgue-medible L en $(0, \infty)$ es de variación lenta en ∞ si

$$\lim_{x \rightarrow \infty} \frac{L(xt)}{L(x)} = 1, \quad t > 0.$$

Esto se denota $L \in \mathcal{R}_0$.

Ejemplo B.2 Las constantes positivas, las funciones que convergen a una constante positiva, los logaritmos y los logaritmos iterados son elementos de \mathcal{R}_0 .

Definición B.3 (*Variación Regular*) Una función positiva, Lebesgue-medible h en $(0, \infty)$ es de variación regular en ∞ si

$$\lim_{x \rightarrow \infty} \frac{h(xt)}{h(x)} = t^\alpha, \quad t > 0.$$

Esto se denota $h \in \mathcal{R}_\alpha$.

Ejemplo B.4 x^α , $x^\alpha \ln(1+x)$, $(x \ln(1+x))^\alpha$, $x^\alpha \ln(\ln(e+x))$ son elementos de \mathcal{R}_α .

Definición B.5 (Cola de Variación Regular) La cola \bar{F} de una distribución es de variación regular con índice $-\alpha$ para alguna $\alpha \geq 0$ (escribimos $\bar{F} \in \mathcal{R}_{-\alpha}$) si

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(xt)}{\bar{F}(x)} = t^{-\alpha}, \quad t > 0.$$

Teorema B.6 (Representación de funciones de variación regular) Decimos que $h \in \mathcal{R}_\alpha$ para algún $\alpha \in \mathbb{R}$, si y sólo si

$$h(x) = c(x) \exp \left\{ \int_z^x \frac{\delta(u)}{u} du \right\}, \quad x \geq z,$$

para alguna $z > 0$, donde c y δ son funciones medibles, $c(x) \rightarrow c_0 \in (0, \infty)$, $\delta(x) \rightarrow \alpha$ cuando $x \rightarrow \infty$.

Una consecuencia inmediata del Teorema anterior es el siguiente Corolario.

Corolario B.7 Si $h \in \mathcal{R}_\alpha$ para algún $\alpha \neq 0$, entonces, cuando $x \rightarrow \infty$

$$h(x) \rightarrow \begin{cases} \infty & \text{si } \alpha > 0 \\ 0 & \text{si } \alpha < 0 \end{cases}$$

La Demostración puede encontrarse en [22], Prop 0.8.

Proposición B.8 (Variación Regular para colas de f.d.) Suponga que F es una f.d. con $F(x) < 1$ para toda $x \geq 0$.

1. Si las sucesiones (a_n) y (x_n) satisfacen $a_n/a_{n+1} \rightarrow 1$, $x_n \rightarrow \infty$, y si para alguna función real g y todo λ de un subconjunto denso de $(0, \infty)$

$$\lim_{n \rightarrow \infty} a_n \bar{F}(\lambda x_n) = g(\lambda) \in (0, \infty),$$

entonces $g(\lambda) = \lambda^{-\alpha}$ para alguna $\alpha \geq 0$ y \bar{F} es de variación regular.

2. Suponga que F es absolutamente continua con densidad f tal que, para alguna $\alpha > 0$, $\lim_{x \rightarrow \infty} xf(x)/\overline{F}(x) = \alpha$. Entonces $f \in \mathcal{R}_{-1-\alpha}$ y, en consecuencia, $\overline{F} \in \mathcal{R}_{-\alpha}$.
3. Suponga que $f \in \mathcal{R}_{-1-\alpha}$ para alguna $\alpha > 0$. Entonces $\lim_{x \rightarrow \infty} xf(x)/\overline{F}(x) = \alpha$. Esto último también es válido si $\overline{F} \in \mathcal{R}_{-\alpha}$ para alguna $\alpha > 0$ y la densidad f es finalmente monótona.
4. Suponga que X es una v.a. no negativa y que para la cola \overline{F} de su f.d. se tiene que $\overline{F} \in \mathcal{R}_{-\alpha}$ para algún $\alpha > 0$. Entonces

$$E[X^\beta] < \infty \quad \text{si } \beta < \alpha$$

$$E[X^\beta] = \infty \quad \text{si } \beta > \alpha$$

Definición B.9 (Sucesiones de variación regular) Una sucesión (c_n) de números positivos es de variación regular con índice $\alpha \in \mathbb{R}$ si

$$\lim_{n \rightarrow \infty} \frac{c_{[tn]}}{c_n} = t^\alpha, \quad t > 0.$$

Siempre que (c_n) es de variación regular con índice α , entonces $c(x) = c_{[x]}$ pertenece a \mathcal{R}_α . A través de esta propiedad, la mayoría de los resultados de \mathcal{R}_α se extienden al caso de sucesiones.

Teorema B.10 (Karamata) (a) Si $\alpha \geq -1$, entonces $U \in \mathcal{R}_\alpha$ implica que $\int_0^x U(t) dt \in \mathcal{R}_{\alpha+1}$ y

$$\lim_{x \rightarrow \infty} \frac{xU(x)}{\int_0^x U(t) dt} = \alpha + 1.$$

Si $\alpha < -1$ (o si $\alpha = -1$ y $\int_x^\infty U(s) ds < \infty$) entonces $U \in \mathcal{R}_\alpha$ implica que $\int_x^\infty U(t) dt$ es finita, $\int_x^\infty U(t) dt \in \mathcal{R}_{\alpha+1}$ y

$$\lim_{x \rightarrow \infty} \frac{xU(x)}{\int_x^\infty U(t) dt} = -\alpha - 1.$$

(b) Si U satisface

$$\lim_{x \rightarrow \infty} \frac{xU(x)}{\int_0^x U(t) dt} = \lambda \in (0, \infty)$$

entonces $U \in \mathcal{R}_{\lambda-1}$. Si $\int_x^\infty U(t) dt < \infty$ y

$$\lim_{x \rightarrow \infty} \frac{xU(x)}{\int_x^\infty U(t) dt} = \lambda \in (0, \infty)$$

entonces $U \in \mathcal{R}_{-\lambda-1}$.

La Demostración puede encontrarse en [22], Teorema 0.6.

Corolario B.11 (La Representación de Karamata) L es de variación lenta si y sólo si L puede representarse como

$$L(x) = c(x) \exp \left\{ \int_1^x t^{-1} \varepsilon(t) dt \right\}$$

para $x > 0$, donde $c: \mathbb{R}^+ \rightarrow \mathbb{R}^+$, $\varepsilon: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ y

$$\lim_{x \rightarrow \infty} c(x) = c \in (0, \infty)$$

$$\lim_{t \rightarrow \infty} \varepsilon(t) = 0.$$

Definición B.12 (Π -Variación) Una función no-negativa y no-decreciente $V(x)$ definida en un intervalo semi-infinito (z, ∞) es de Π -Variación (lo cual se denota $V \in \Pi$) si existen funciones $a(t) > 0$, $b(t) \in \mathbb{R}$ tales que para $x > 0$

$$\lim_{t \rightarrow \infty} \frac{V(tx) - b(t)}{a(t)} = \ln x. \quad (\text{B.1})$$

Nótese que en la relación (B.1) podemos tomar $b(t) = V(t)$, puesto que

$$\frac{V(tx) - V(t)}{a(t)} = \frac{V(tx) - b(t)}{a(t)} - \frac{V(t) - b(t)}{a(t)} \rightarrow \ln x - \ln 1 = \ln x. \quad (\text{B.2})$$

Más aún, haciendo $x = e$ en la expresión (B.2) muestra que podemos tomar

$$a(t) = V(te) - V(t).$$

La función $a(\cdot)$ es única excepto por equivalencia asintótica: si $a(\cdot)$ satisface la expresión (B.1) y $a(t) \sim a_1(t)$, entonces $a_1(t)$ satisface (B.1). Cualquier función $a(t)$ que satisfaga (B.1) se llama función auxiliar.

Proposición B.13 (*Relación entre el DAM(Λ) y Π -Variación*) Sea F una función de distribución y sea

$$U := 1/(1 - F),$$

de manera que U^- está definido en $(1, \infty)$. Son equivalentes:

(i) $F \in \text{DAM}(\Lambda)$

(ii) $U^- \in \Pi$.

Para la Demostración, ver [22], Prop. 0.10.

Proposición B.14 Sea $V \in \Pi$ y sean $V_1(t) = t \int_t^\infty V(u)u^{-2}du$ y $V_2(t) = t \int_t^\infty V_1(u)u^{-2}du$.

Entonces V_1 es continua, y estrictamente creciente, V_2 es dos veces diferenciable y se satisface que

$$V_1(t) > V(t),$$

$$V_2(t) > V(t)$$

Apéndice C

Estadística

Cuando se debe tomar una decisión entre dos o más estimadores para un mismo parámetro es importante tener criterios para compararlos. A continuación se presentan los criterios que se utilizan más comunmente. Sea X_1, \dots, X_n una muestra aleatoria de una distribución con parámetro θ cuyo valor es desconocido y se debe estimar.

Definición C.1 (*Estimador Insesgado*) *Un estimador $\delta(X_1, \dots, X_n)$ es un estimador insesgado de un parámetro θ si $E[\delta(X_1, \dots, X_n)] = \theta$ para todo valor posible de θ .*

En otras palabras, un estimador de un parámetro θ es insesgado si su esperanza es igual al verdadero valor desconocido de θ .

Si un estimador es insesgado, entonces el promedio de los valores que toma en un número que tiende a infinito de muestras es el parámetro estimado; sin embargo, esto no implica que el estimador tenga muy alta probabilidad de estar cerca del parámetro desconocido para cualquier muestra dada. Conviene entonces considerar la varianza

de los estimadores insesgados, pues el estimador insesgado que tiene la menor varianza tiene mayor probabilidad de estar "cerca" del parámetro desconocido. Esto se precisa en la siguiente definición.

Definición C.2 (Eficiencia Relativa de un Estimador) Si tanto δ_1 como δ_2 son estimadores insesgados de θ para la misma muestra, entonces δ_1 es más eficiente que δ_2 si

$$\text{Var}(\delta_1) < \text{Var}(\delta_2).$$

Tenemos pues que el estimador más eficiente es el que tiene la menor varianza.

Definición C.3 (Eficiencia de un Estimador) Sea Θ un intervalo abierto en la recta real, y sea $\{f_\theta : \theta \in \Theta\}$ una familia de funciones de densidad. Sea $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Supongamos que el conjunto $\{\mathbf{x} : f_\theta(\mathbf{x}) > 0\}$ es independiente de θ y que, para cada θ , está definida $\partial f_\theta(\mathbf{x}) / \partial \theta$. Sea ψ definida y diferenciable en Θ , y sea T un estimador insesgado de ψ tal que $E_\theta[T^2] < \infty$ para toda θ . Supongamos que

$$\frac{\partial}{\partial \theta} \int f_\theta(\mathbf{x}) d\mathbf{x} = \int \frac{\partial}{\partial \theta} f_\theta(\mathbf{x}) d\mathbf{x} = 0 \quad (\text{C.1})$$

y

$$\frac{\partial}{\partial \theta} \int T(\mathbf{x}) f_\theta(\mathbf{x}) d\mathbf{x} = \int T(\mathbf{x}) \frac{\partial}{\partial \theta} f_\theta(\mathbf{x}) d\mathbf{x} \quad (\text{C.2})$$

para toda $\theta \in \Theta$ se satisfacen para la familia de f.d. $\{F_\theta : \theta \in \Theta\}$. Bajo estas condiciones, diremos que un estimador insesgado T para el parámetro θ es eficiente para la familia $\{F_\theta\}$ si

$$\text{Var}_\theta(T) = \left[E_\theta \left\{ \frac{\partial \ln f_\theta(\mathbf{X})}{\partial \theta} \right\}^2 \right]^{-1}.$$

Observación C.4 Las condiciones (C.1) y (C.2) se conocen como *Condiciones de Regularidad*. Es claro que $E_{\theta} \{ \partial \ln f_{\theta}(\mathbf{X}) / \partial \theta \}^2$ está bien definida y que satisface

$$0 \leq E_{\theta} \left\{ \frac{\partial \ln f_{\theta}(\mathbf{X})}{\partial \theta} \right\}^2 \leq \infty.$$

Definición C.5 (Consistencia de un Estimador) Sea Θ_n un estimador del parámetro θ para una muestra de tamaño n . Θ_n es un estimador consistente de θ si

$$\lim_{n \rightarrow \infty} P [|\Theta_n - \theta| > \varepsilon] = 0$$

o, en forma equivalente, si

$$\lim_{n \rightarrow \infty} P [|\Theta_n - \theta| < \varepsilon] = 1,$$

para todo $\varepsilon > 0$.

En primer lugar, es preciso notar que un estimador Θ_n es consistente si una sucesión de probabilidades converge a cero (o a uno) conforme crece el tamaño de la muestra. Por tanto, siendo rigurosos, la consistencia de un estimador depende únicamente de su comportamiento en el límite conforme crece el tamaño de la muestra, y no implica que el valor observado de Θ_n esté necesariamente cercano a θ para cualquier tamaño específico n de la muestra. Por ello, la consistencia de un estimador es una propiedad asintótica.

Teorema C.6 Si

$$\lim_{n \rightarrow \infty} E [\Theta_n] = \theta$$

y

$$\lim_{n \rightarrow \infty} \text{Var} (\Theta_n) = 0,$$

donde n es el tamaño de la muestra, entonces Θ_n es un estimador consistente de θ .

Teorema C.7 (Glivenko-Cantelli) Sean $\{X_n, n \geq 1\}$ v.a.i.i.d. con f.d. F . Suponemos que F es desconocida y en base a la muestra X_1, \dots, X_n queremos estimar F . El estimador será la función de distribución empírica (f.d.e.) definida mediante

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{X_j \leq x\}}.$$

Una aplicación de la Ley Fuerte de los Grandes Números (LFGN) nos dice que

$$\widehat{F}_n(x) \rightarrow E[\mathbf{1}_{\{X \leq x\}}] = F(x)$$

para todo $x \in \mathbb{R}$. Esto último es equivalente a

$$\Delta_n = \sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F(x) \right| \xrightarrow{c.s.} 0, \quad n \rightarrow \infty.$$

Demostración:

La demostración se hará para el caso de una f.d. F continua. Sean

$$-\infty = x_0 < x_1 < \dots < x_k < x_{k+1} = \infty$$

puntos tales que $F(x_{i+1}) - F(x_i) < \varepsilon$ para $\varepsilon > 0$ dado, $i = 0, \dots, k$. $F(\pm\infty)$ se interpretan como límites en la manera natural. Por la monotonía de F y \widehat{F}_n se obtiene

$$\begin{aligned} \Delta_n &= \max_{i=0, \dots, k} \sup_{x \in (x_i, x_{i+1})} \left| \widehat{F}_n(x) - F(x) \right| \\ &\leq \max_{i=0, \dots, k} \left(\widehat{F}_n(x_{i+1}) - F(x_i), F(x_{i+1}) - \widehat{F}_n(x_i) \right). \end{aligned}$$

Al aplicar la LFGN al lado derecho se obtiene

$$\limsup_{n \rightarrow \infty} \Delta_n \leq \max_{i=0, \dots, k} (F(x_{i+1}) - F(x_i)) < \varepsilon \quad \text{c.s.}$$

■

Bibliografía

- [1] ADLER, R., FELDMAN, R., TAQQU, M. (editores) (1998), *A Practical Guide to Heavy Tails*, Birkhäuser, New York.
- [2] ARTZNER, P. (1999), "*Application of Coherent Risk Measures to Capital Requirements in Insurance*", Preprint, Society of Actuaries.
- [3] ARTZNER, P., DELBAEN, F., EBER, J.-M. y HEATH, D. (1997), "*Thinking Coherently*", RISK, Nov. 1997, pp. 68-71.
- [4] ARTZNER, P., DELBAEN, F., EBER, J.-M. y HEATH, D. (1999), "*Coherent Measures of Risk*", *Mathematical Finance*, 9(3), pp. 203-228.
- [5] BASSI, F., EMBRECHTS, P. y KAFETZAKI, M. (1997) "*A Survival Kit on Quantile Estimation*", UBS Quant Workshop, Zurich.
- [6] EMBRECHTS, P., KLÜPPELBERG, C. y MIKOSCH, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer Verlag, Berlín.
- [7] EMBRECHTS, P., RESNICK, S. y SAMORODNITSKY, G. (1998), "*Extreme Value Theory as a Risk Management Tool*", *North American Actuarial Journal*.
- [8] EMBRECHTS, P., MC.NEIL, A., STRAUMANN, D. y KAUFMANN, R. (1999), "*Modelling Extremal Events for Insurance and Finance*", ETH Zurich.
- [9] EMBRECHTS, P., MC.NEIL, A. y STRAUMANN, D. (1999), "*Correlation and Dependence in Risk Management: Properties and Pitfalls*", Preprint, ETH Zurich.

- [10] EMBRECHTS, P., MC.NEIL, A. y STRAUMANN, D. (1999), "Pitfalls and Alternatives", RISK, Mayo 1999, pp. 69-71.
- [11] FANG, K.-T., KOTZ, S. y NG, K.-W. (1987), *Symmetric Multivariate and Related Distributions*, Chapman & Hall, Londres.
- [12] FELLER, W. (1983), *Introducción a la Teoría de Probabilidades y sus Aplicaciones*, Vol. 1, Limusa, México.
- [13] FELLER, W. (1985), *Introducción a la Teoría de Probabilidades y sus Aplicaciones*, Vol. 2, Limusa, 2a. ed., México.
- [14] GALAMBOS, J., LECHNER, J. y SIMIU, E. (editores) (1994), *Extreme Value Theory and Applications*, Kluwer Academic Publishers, Países Bajos.
- [15] LEADBETTER, M.R., LINDGREN, G. y ROOTZÉN, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.
- [16] MC.NEIL, A.J. (1997), "Estimating the Tails of Loss Severity Distributions using Extreme Value Theory", ASTIN Bulletin 27, pp. 117-137.
- [17] MC.NEIL, A.J. (1998), "Calculating Quantile Risk Measures for Financial return Series Using Extreme Value Theory", Preprint, ETH Zurich.
- [18] MC.NEIL, A.J. (1999a), "Extreme Value Theory for Risk Managers", Preprint, ETH Zurich.
- [19] MC.NEIL, A.J. (1999b), "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach", Preprint, ETH Zurich.
- [20] NELSEN, R. B. (1999), *An Introduction to Copulas*, Springer, New York.
- [21] OLKIN, I., GLEESER, L. y DERMAN, C. (1980), *Probability Models and Applications*, Macmillan Publishing Co., New York.
- [22] RESNICK, SIDNEY I. (1987) *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag, New York.