



MÁSTER EN CIENCIAS ACTUARIALES Y FINANCIERAS

TRABAJO FIN DE MÁSTER:

**ANÁLISIS COMPARATIVO DE TRES
MODELOS DE DENSIDAD COMPUESTA
PARA EL ESTUDIO DE SINIESTROS DE
COLA PESADA CON R**

Autor:

OLEKSANDR ZAVERSHYNSKY

Tutores:

MARÍA PÉREZ MARTÍN
ADOLFO CABALLERO CARBONELL

A mis padres

Agradecimientos

En primer lugar, quiero agradecerles a mis tutores María Pérez Martín y Adolfo Caballero Carbonell toda la ayuda y apoyo prestados a lo largo de la elaboración del presente trabajo, sin olvidarme tampoco de la colaboración del profesor Adolfo Hernández Estrada.

A continuación, deseo mostrar mi gratitud a todos los profesores que he tenido a lo largo de mi formación académica, tanto en el Máster en Ciencias Actuariales y Financieras cursado en la Facultad de Ciencias Económicas y Empresariales de la Universidad Complutense de Madrid como a los que me impartieron clase en mi grado en la Facultad de Matemáticas de la Universidad de Murcia, sin omitir a los demás profesores de los que he tenido el placer de recibir clase. Me habéis equipado con un bagaje de conocimiento que nunca podré agradecerlos lo suficiente.

Seguidamente, he de dar las gracias a mi familia, especialmente a mi madre, por todo el trabajo y sacrificio que les ha costado ayudarme a llegar aquí y por su incondicional apoyo durante toda mi vida que nunca flaqueó. Os estaré eternamente agradecido por ello y por más.

Por supuesto, también merecen agradecimiento mis amigos –tanto los de siempre, como los nuevos–, pues sin ellos yo no sería quien soy. Citar a unos puede molestar a los que se quedaron sin mención, pero he de nombrar a algunos de ellos, pues también me ayudaron –y siguen ayudando– en mi camino. Gracias Álvaro, Joseph, Pablo y Vincent.

Simplemente, a todas las personas que me ayudan a seguir adelante –a las mencionadas explícitamente aquí y a las omitidas por mi negligencia–, gracias de todo corazón por estar ahí.

Resumen

En este trabajo se plantea el problema de la búsqueda de un modelo estocástico, en términos de una distribución de probabilidad, para explicar conjuntos de datos procedentes de siniestros de seguros no vida con colas pesadas.

Se estudian tres modelos distintos con la misma idea general: aplicar una densidad hasta un cierto umbral a determinar y otra a partir del mismo.

La densidad elegida para la primera parte del dominio de la variable aleatoria es la distribución lognormal, en vista de su forma asimétrica y con una moda estrictamente positiva, siendo ambas características muy frecuentes en los conjuntos de datos procedentes del sector asegurador no vida.

Para la segunda densidad del modelo se han considerado dos alternativas: una distribución de Pareto ordinaria y una distribución de Pareto generalizada.

Los modelos son requeridos de cumplir condiciones de continuidad y diferenciabilidad de distintos órdenes.

Todos los modelos son introducidos y desarrollados desde una perspectiva teórica, algunas de sus propiedades son estudiadas de forma analítica y otras por medio de simulación y representación gráfica.

Para verificar la utilidad de los modelos expuestos, éstos se aplican a los datos de siniestros derivados de incendios daneses en los años 1980 –un conjunto de datos muy utilizado en la literatura actuarial–.

Los ajustes se realizan sobre una parte de los datos, y los resultados obtenidos se contrastan con la parte complementaria de la muestra global.

Varias medidas para la comparación de modelos son introducidas –como la razón de verosimilitudes y el criterio de información de Akaike– y utilizadas para comparar los ajustes obtenidos con los diversos modelos estudiados en el presente trabajo.

Consideraciones previas

Todas las figuras que aparecen en esta memoria son de elaboración propia, a no ser que se especifique lo contrario.

Igualmente, salvo mención en contra, todos los gráficos se han realizado utilizando el programa *R* [R Core Team, 2015].

Índice general

Agradecimientos	V
Resumen	VII
Consideraciones previas	IX
1. Introducción	1
2. Modelo de Cooray & Ananda	3
2.1. Construcción del modelo	3
2.2. Expresión alternativa de la densidad	7
2.3. Comparación con las densidades	8
2.4. Variación de la densidad con los parámetros	9
2.4.1. Cambios en la densidad con el parámetro de forma α .	9
2.4.2. Cambios en la densidad con el parámetro de corte θ . .	11
2.5. Cálculo analítico de la moda	11
2.6. Función de distribución	12
2.7. Comparación con las distribuciones	13
2.8. Variación de la CDF con los parámetros	15
2.8.1. Cambios en la distribución con el parámetro de forma α	15
2.8.2. Cambios en la distribución con el parámetro de corte θ	16
2.9. Función cuantil	17
2.10. Estimación de los cuantiles	18
2.11. Análisis inicial de los datos daneses	19
2.12. Ajuste a los datos daneses	21
2.13. Prueba χ^2	26
2.14. Prueba G	33
2.15. Gráfico QQ	35

3. Modelo de Scollnik	39
3.1. Desarrollo teórico del modelo	40
3.1.1. Condición de continuidad	41
3.1.2. Estudio del parámetro de peso r	42
3.1.3. Condición de diferenciabilidad de primer orden	45
3.1.4. Restricción al modelo de Cooray & Ananda	47
3.1.5. Condición de diferenciabilidad de segundo orden	49
3.2. Variación de la densidad con los parámetros	51
3.2.1. Cambios en la densidad con el parámetro de forma α	51
3.2.2. Cambios en la densidad con el parámetro de corte θ	53
3.2.3. Cambios en la densidad con el parámetro de peso r	60
3.3. Función de distribución	61
3.4. Función cuantil	62
3.5. Ajuste a los datos daneses	64
3.6. Pruebas χ^2 y G	68
3.7. Gráfico QQ	71
3.8. Ratio de verosimilitudes	71
3.9. Criterio de información de Akaike	74
4. Modelos de Scollnik con GPD	77
4.1. Distribución de Pareto generalizada	77
4.2. Modelo compuesto	80
4.2.1. Condición de continuidad	81
4.2.2. Condición de diferenciabilidad	81
4.2.3. Condición de diferenciabilidad de segundo orden	82
4.3. Función de distribución	83
4.4. Función cuantil	84
4.5. Ajuste a los datos daneses	85
4.5.1. Ajuste del modelo con restricciones de continuidad y diferenciabilidad de primer orden	85
4.5.2. Ajuste del modelo con restricciones de continuidad y diferenciabilidad de segundo orden	88
4.6. Gráficos QQ	94
4.6.1. Modelo con restricciones de continuidad y diferencia- bilidad de primer orden	95
4.6.2. Modelo con restricciones de continuidad y diferencia- bilidad de segundo orden	97
4.7. Ratios de verosimilitud y AIC	97

<i>ÍNDICE GENERAL</i>	XIII
4.7.1. Ratios de verosimilitud	97
4.7.2. AIC	100
5. Conclusiones y futuros desarrollos	103

Índice de figuras

2.1. Funciones de densidad de una lognormal, de una Pareto y de su composición	8
2.2. Densidades del modelo compuesto para un rango de valores de α	10
2.3. Densidades del modelo compuesto para un rango de valores de θ	10
2.4. Funciones de distribución de una lognormal, de una Pareto y de su composición	14
2.5. Funciones de distribución del modelo compuesto para un rango de valores de α	15
2.6. Funciones de distribución del modelo compuesto para un rango de valores de θ	16
2.7. Histogramas de los datos de siniestros derivados de incendios en Dinamarca	22
2.8. Histograma parcial correspondiente a valores inferiores a 20 millones de los incendios daneses con la función de densidad del modelo de Cooray & Ananda ajustado	27
2.9. Función de distribución empírica junto con la función de distribución del modelo de Cooray & Ananda ajustado en el tramo de los valores inferiores a 20 millones de los incendios daneses	28
2.10. Gráfico QQ de los datos daneses frente al modelo de Cooray & Ananda ajustado	37
3.1. Valores del parámetro de peso r como función del parámetro de forma de la componente Pareto α	44
3.2. Densidades del modelo de Scollnik para dos formas de calcular σ , incluyendo un σ <i>infradisperso</i> con respecto a k/α , y para un rango de valores de α	54

3.3.	Densidades del modelo de Scollnik para dos formas de calcular σ , incluyendo un σ <i>sobredisperso</i> con respecto a k/α , y para un rango de valores de α	55
3.4.	Densidades del modelo de Scollnik con $\sigma = 2\frac{k}{\alpha}$ para un rango de valores de α	56
3.5.	Densidades del modelo de Scollnik para dos formas de calcular σ , una que proporciona un σ <i>infradisperso</i> y otra que proporciona un σ <i>sobredisperso</i> , con respecto a k/α , y para un rango de valores de θ	57
3.6.	Densidades del modelo de Scollnik para un rango de valores de r con $\alpha = 2$	58
3.7.	Densidades del modelo de Scollnik para un rango de valores de r con $\alpha = 0,5$	59
3.8.	Densidades del modelo de Scollnik para un rango de valores de r con $\alpha = 1$	59
3.9.	Histograma parcial correspondiente a valores inferiores a 20 millones de los incendios daneses con la función de densidad del modelo de Scollnik ajustado	68
3.10.	Función de distribución empírica junto con la función de distribución del modelo de Scollnik ajustado en el tramo de los valores inferiores a 20 millones de los incendios daneses	69
3.11.	Gráfico QQ de los datos daneses frente al modelo de Scollnik ajustado	72
4.1.	Histograma parcial correspondiente a valores inferiores a 20 millones de los incendios daneses con la función de densidad del modelo de Scollnik con cola GPD ajustado	89
4.2.	Función de distribución empírica junto con la función de distribución del modelo de Scollnik con cola GPD ajustado en el tramo de los valores inferiores a 20 millones de los incendios daneses	90
4.3.	Histograma parcial correspondiente a valores inferiores a 20 millones de los incendios daneses con la función de densidad del modelo de Scollnik ajustado con cola GPD y restricción de diferenciabilidad de segundo orden	94

4.4. Función de distribución empírica junto con la función de distribución del modelo de Scollnik ajustado con cola GPD y restricción de diferenciabilidad de segundo orden en el tramo de los valores inferiores a 20 millones de los incendios daneses	95
4.5. Gráfico QQ de los datos daneses frente al modelo de Scollnik ajustado con cola GPD y restricción de diferenciabilidad de primer orden	96
4.6. Gráfico QQ de los datos daneses frente al modelo de Scollnik ajustado con cola GPD y restricción de diferenciabilidad de segundo orden	98

Índice de tablas

2.1. Estadísticos descriptivos de los datos daneses	20
2.2. Cuantiles empíricos de los datos daneses	20
2.3. Frecuencias esperadas para los intervalos (2.26) según el modelo de Cooray & Ananda	33

Capítulo 1

Introducción

En la mayoría de los seguros de no vida, los datos sobre siniestros presentan dos características comunes: una cola derecha muy pesada y una fuerte asimetría positiva. Esto hace que entre los modelos paramétricos elegidos por los actuarios para ajustar a los datos siniestrales se encuentren distribuciones como la log-normal, gamma o inversa Gaussiana.

El problema de estas distribuciones es que sus colas son más ligeras que las que generalmente se observan empíricamente, y por tanto, un modelo basado únicamente en una distribución así no dará cuenta por completo de todo el riesgo subyacente en la cola.

Un posible enfoque se basa en el Teorema de Pickands–Balkema–de Haan el cual establece la distribución asintótica de los excesos de la variable aleatoria por encima de un cierto umbral. Bajo este planteamiento, *se elige* un umbral *adecuado*, se ajusta una distribución de Pareto generalizada a los excesos por encima de este umbral elegido para captar la peligrosidad de la cola, para seguidamente ajustar una distribución a los siniestros por debajo del umbral, eligiendo entre las mencionadas distribuciones log-normal, gamma, inversa Gaussiana o alguna otra distinta.

Los problemas que surgen a raíz de esta técnica son varios. El primero de ellos consiste en la mencionada elección del *umbral adecuado*. Esta elección es a menudo subjetiva, y los resultados finales son bastante sensibles a la elección del umbral [McNeil, 1997, 4.6].

Otro problema aparece con la función de densidad en el punto de corte. En general, usar un modelo para siniestros por debajo del umbral y una Pareto generalizada para los siniestros por encima del mismo, hace que la función de densidad del modelo global presente una discontinuidad en el umbral. Esto

tampoco es un problema crítico, pero parece poco realista que la probabilidad de observar siniestros se vea alterada de forma drástica en función de si son ligeramente inferiores o ligeramente superiores a un determinado umbral.

Por tanto, es interesante obtener modelos para las cuantías siniestras que combinen distribuciones asimétricas con una moda estrictamente positiva como la log-normal para siniestros de frecuencia junto con distribuciones de colas pesadas como la Pareto o la Pareto generalizada para captar el peligro de los siniestros de impacto, sin que hayan discontinuidades poco realistas en ningún punto.

Capítulo 2

Modelo basado en las distribuciones lognormal y Pareto

En 2005 Cooray & Ananda construyeron un modelo basado en las distribuciones lognormal y Pareto que no requiere elegir el umbral *a priori*, sino que lo deja libre dentro del algoritmo de elección de parámetros utilizado, como la maximización de la verosimilitud, y que no provoca ningún tipo de discontinuidad en la función de densidad en el umbral elegido.

La distribución de Pareto utilizada para modelar los siniestros más extremos *no* es la distribución de Pareto generalizada, sino la distribución de Pareto usual. A pesar de no tener el apoyo teórico del Teorema de Pickands–Balkema–de Haan detrás, como la Pareto generalizada, la distribución de Pareto clásica ha probado ser un buen modelo para los siniestros de mayores cuantías, sobre todo en seguros de automóvil e incendios [Seal, 1980].

2.1. Construcción del modelo

Más concretamente, Cooray & Ananda definen el siguiente modelo:

$$f(x) = \begin{cases} cf_1(x), & 0 < x \leq \theta \\ cf_2(x), & \theta < x < \infty \end{cases} \quad (2.1)$$

donde c es una constante de normalización tal que

$$\int_0^{\infty} f(x)dx = \int_0^{\theta} cf_1(x)dx + \int_{\theta}^{\infty} cf_2(x)dx = 1, \quad (2.2)$$

f_1 es la función de densidad de la distribución lognormal de parámetros μ y σ ,

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp \left\{ -\frac{1}{2} \left(\frac{\ln(x) - \mu}{\sigma} \right)^2 \right\}, \quad x > 0, \quad (2.3)$$

y f_2 es la función de densidad de la distribución de Pareto clásica de parámetros α y θ ,

$$f_2(x) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, \quad x > \theta. \quad (2.4)$$

Por tanto, en principio el modelo (2.1) tiene cuatro parámetros, dos de la parte de la lognormal, y otros dos de la parte de la Pareto. Además, el parámetro μ puede tomar cualquier valor real, pero los restantes tres parámetros σ , α y θ han de ser estrictamente positivos.

Para solventar los mencionados problemas de continuidad, se imponen al modelo (2.1) dos condiciones: una de continuidad y otra de suavidad. Concretamente,

$$\lim_{x \rightarrow \theta^-} f(x) = \lim_{x \rightarrow \theta^+} f(x), \quad (2.5)$$

y

$$\lim_{x \rightarrow \theta^-} f'(x) = \lim_{x \rightarrow \theta^+} f'(x). \quad (2.6)$$

Puesto que tanto f_1 como f_2 son funciones continuas, se tiene que

$$\lim_{x \rightarrow \theta^-} f(x) = cf_1(\theta), \quad \lim_{x \rightarrow \theta^+} f(x) = cf_2(\theta), \quad (2.7)$$

$$\lim_{x \rightarrow \theta^-} f'(x) = cf_1'(\theta), \quad \lim_{x \rightarrow \theta^+} f'(x) = cf_2'(\theta). \quad (2.8)$$

Utilizando (2.7), (2.5) se transforma en

$$\frac{1}{\sqrt{2\pi}\sigma\theta} \exp \left\{ -\frac{1}{2} \left(\frac{\ln(\theta) - \mu}{\sigma} \right)^2 \right\} = \frac{\alpha}{\theta},$$

o de forma más compacta,

$$\exp \left\{ -\frac{1}{2} \left(\frac{\ln(\theta) - \mu}{\sigma} \right)^2 \right\} = \sqrt{2\pi}\alpha\theta. \quad (2.9)$$

Derivando y simplificando, se tiene que

$$f_1'(x) = -\frac{\exp\left\{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2\right\}[\ln(x)-\mu+\sigma^2]}{\sqrt{2\pi}\sigma^3x^2}, \quad (2.10)$$

mientras que

$$f_2'(x) = -\frac{(\alpha+1)\alpha\theta^\alpha}{x^{\alpha+2}}. \quad (2.11)$$

Similarmente, utilizando (2.8) y estas derivadas, (2.6) se transforma en

$$\frac{\exp\left\{-\frac{1}{2}\left(\frac{\ln(\theta)-\mu}{\sigma}\right)^2\right\}(\ln(\theta)-\mu+\sigma^2)}{\sqrt{2\pi}\sigma^3\theta^2} = \frac{(\alpha+1)\alpha}{\theta^2},$$

donde usando (2.9) y operando se tiene

$$\frac{\ln(\theta)-\mu}{\sigma} = \alpha\sigma. \quad (2.12)$$

Llamando k al producto $\alpha\sigma$, y utilizando la recién obtenida relación (2.12) en (2.9) se llega a la ecuación

$$\exp\left\{-\frac{1}{2}k^2\right\} = \sqrt{2\pi}k, \quad (2.13)$$

la cual, al resolver numéricamente para k , proporciona $k \approx 0,3722388980356$.

Con estas restricciones, el número de parámetros del modelo (2.1) se reduce de los inicialmente cuatro a solamente dos. Así, si se eligen los parámetros μ y σ de la lognormal, los parámetros α y θ de la Pareto son necesariamente

$$\alpha = \frac{k}{\sigma}, \quad \theta = \exp\{\mu + k\sigma\}.$$

Recíprocamente, si se establecen los parámetros deseados para la cola gobernada por los parámetros α y θ de la Pareto, los parámetros μ y σ de la cabeza¹ lognormal del modelo global han de ser

$$\sigma = \frac{k}{\alpha}, \quad \mu = \ln(\theta) - k\sigma. \quad (2.14)$$

¹Por analogía con el término *cola* para denotar la parte de la distribución de probabilidad correspondiente a los valores más altos, se utiliza el término *cabeza* para denotar la parte correspondiente a los valores más bajos. Así, en el modelo de Cooray & Ananda, la *cola* es la parte de la distribución de Pareto, que empieza a partir de $x = \theta$, mientras que la *cabeza* es la parte correspondiente a la distribución lognormal, que gobierna desde $x = 0$ hasta $x = \theta$.

Volviendo al argumento de (2.2) para calcular la constante de normalización c se tiene:

$$\begin{aligned}
 1 &= \int_0^{\infty} f(x) dx \\
 &= \int_0^{\theta} c f_1(x) dx + \int_{\theta}^{\infty} c f_2(x) dx \\
 &= c \left(\int_0^{\theta} f_1(x) dx + \int_{\theta}^{\infty} \frac{\alpha \theta^{\alpha}}{x^{\alpha+1}} dx \right) \\
 &= c \left(\int_0^{\theta} f_1(x) dx + \alpha \theta^{\alpha} \left[\frac{-1}{\alpha x^{\alpha}} \right]_{x=\theta}^{x=\infty} \right) \\
 &= c \left(\int_0^{\theta} f_1(x) dx + 1 \right).
 \end{aligned}$$

Por otro lado, llamando X a una variable aleatoria que siga una distribución lognormal de parámetros μ y σ , se tiene

$$\begin{aligned}
 \int_0^{\theta} f_1(x) dx &= \mathbb{P}[X \leq \theta] = \mathbb{P}[\ln(X) \leq \ln(\theta)] = \mathbb{P}\left[\frac{\ln(X) - \mu}{\sigma} \leq \frac{\ln(\theta) - \mu}{\sigma}\right] \\
 &= \mathbb{P}\left[Z \leq \frac{\ln(\theta) - \mu}{\sigma}\right] = \Phi\left(\frac{\ln(\theta) - \mu}{\sigma}\right) = \Phi(\alpha\sigma) = \Phi(k),
 \end{aligned}$$

donde, como es usual, Z denota a una variable aleatoria que siga una distribución normal estándar, y $\Phi(\cdot)$ es su función de distribución.

Despejando de las dos últimas identidades, se obtiene el valor de c :

$$c = \frac{1}{1 + \Phi\left(\frac{\ln(\theta) - \mu}{\sigma}\right)} = \frac{1}{1 + \Phi(\alpha\sigma)} = \frac{1}{1 + \Phi(k)} \approx 0,607850077484. \quad (2.15)$$

2.2. Expresión alternativa de la densidad

En [Cooray & Ananda, 2005] se afirma que el modelo (2.1) puede ser escrito como

$$f(x) = \begin{cases} \frac{\alpha\theta^\alpha}{[1 + \Phi(k)]x^{\alpha+1}} \exp\left\{-\frac{\alpha^2}{2k^2} \left[\ln\left(\frac{x}{\theta}\right)\right]^2\right\}, & 0 < x \leq \theta \\ \frac{\alpha\theta^\alpha}{[1 + \Phi(k)]x^{\alpha+1}}, & \theta < x < \infty \end{cases}. \quad (2.16)$$

En vista de (2.4) y (2.15) es inmediato que la segunda componente de la definición de (2.1) se puede escribir como la segunda componente de (2.16).

Para ver la igualdad entre las primeras componentes de (2.1) y (2.16) es suficiente ver que

$$\frac{c}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{1}{2} \left(\frac{\ln(x) - \mu}{\sigma}\right)^2\right\} = \frac{\alpha\theta^\alpha}{(1 + \Phi(k))x^{\alpha+1}} \exp\left\{-\frac{\alpha^2}{2k^2} \left[\ln\left(\frac{x}{\theta}\right)\right]^2\right\}.$$

Por lo obtenido en (2.15), esta última igualdad es equivalente a

$$\exp\left\{-\frac{1}{2} \left(\frac{\ln(x) - \mu}{\sigma}\right)^2\right\} = \sqrt{2}\alpha\sigma \left(\frac{\theta}{x}\right)^\alpha \exp\left\{-\frac{\alpha^2}{2k^2} \left[\ln\left(\frac{x}{\theta}\right)\right]^2\right\}.$$

Puesto que $k = \alpha\sigma$, y usando (2.13), lo que se persigue demostrar se puede escribir como

$$\exp\left\{-\frac{1}{2} \left(\frac{\ln(x) - \mu}{\sigma}\right)^2\right\} = \exp\left\{-\frac{1}{2}k^2 + \alpha \ln\left(\frac{\theta}{x}\right) - \frac{\alpha^2}{2k^2} \left[\ln\left(\frac{x}{\theta}\right)\right]^2\right\}.$$

A raíz de la inyectividad de la exponencial, esto es equivalente a

$$\left[\frac{\ln(x) - \mu}{\sigma}\right]^2 = k^2 - 2\alpha \ln\left(\frac{\theta}{x}\right) + \left[\frac{\alpha}{k} \ln\left(\frac{x}{\theta}\right)\right]^2,$$

lo cual a su vez se puede escribir como

$$[\ln(x) - \mu]^2 = \alpha^2\sigma^4 + 2\alpha\sigma^2 \ln\left(\frac{x}{\theta}\right) + \left[\ln\left(\frac{x}{\theta}\right)\right]^2.$$

Usando las propiedades de los logaritmos, desarrollando los cuadrados, reordenando y simplificando, se obtiene

$$2\ln(x)[\ln(\theta) - \mu] - \alpha^2\sigma^4 - 2\alpha\sigma^2 [\ln(x) - \ln(\theta)] = [\ln(\theta)]^2 - \mu^2.$$

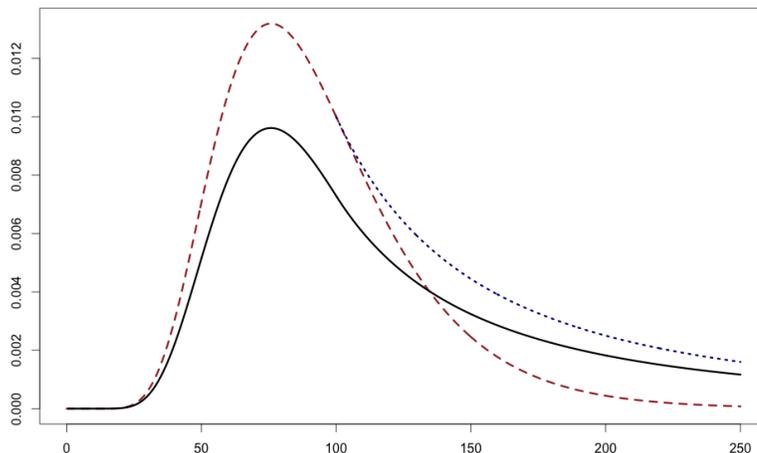


Figura 2.1: Funciones de densidad de una lognormal (línea discontinua marrón), de una Pareto (línea punteada azul) y de su composición (línea continua negra)

En vista de que $\ln(\theta) - \mu = \alpha\sigma^2$ (véase (2.12)), la anterior ecuación se simplifica a

$$\alpha\sigma^2[2\ln(\theta) - \alpha\sigma^2] = [\ln(\theta)]^2 - \mu^2.$$

Volviendo a usar la relación (2.12), esto último se escribe como

$$[\ln(\theta) - \mu][\ln(\theta) + \mu] = [\ln(\theta)]^2 - \mu^2,$$

lo cual es cierto por ser una famosa identidad notable. Siguiendo las equivalencias anteriores en el orden inverso, se consigue demostrar la deseada igualdad entre las primeras componentes de (2.1) y (2.16), y por tanto se tiene demostrada por completo la equivalencia de estos dos modelos.

2.3. Comparación con las densidades lognormal y Pareto

En la Figura 2.1 se muestran tres funciones de densidad: la de una Pareto de parámetros $\theta = 100$ y $\alpha = 1$; la de una lognormal con los parámetros calculados como en (2.14) para que se verifiquen las condiciones de continuidad

y diferenciabilidad en el modelo lognormal-Pareto compuesto (2.16), esto es,

$$\sigma = \frac{k}{\alpha} \approx \frac{0,3722389}{1} = 0,3722389,$$

$$\mu = \ln(\theta) - k\sigma \approx \ln(100) - 0,3722389 \cdot 0,3722389 \approx 4,466608;$$

y finalmente, la propia función de densidad del modelo compuesto (2.16). La densidad de la lognormal se ha representado en línea discontinua, la de la Pareto con una línea punteada y la de la composición por una línea continua.

Nótese que se consigue obtener uno de los objetivos buscados con este modelo: obtener una función de densidad que tenga una forma parecida a la densidad de una distribución lognormal, esto es, con la característica campana asimétrica hacia la derecha, pero que a su vez posea una cola más pesada que la distribución lognormal. La cola del modelo compuesto parece encontrarse entre las dos colas de las distribuciones que conforman el modelo (2.16), siendo más pesada que la cola de la lognormal, pero más ligera que la de la Pareto.

2.4. Variación de la densidad con los parámetros

2.4.1. Cambios en la densidad con el parámetro de forma α

En la Figura 2.2 se representan las funciones de densidad del modelo compuesto para un rango de valores de α , mientras que el valor de θ se mantuvo constante. Concretamente, se han dibujado densidades para $\theta = 100$ y $\alpha = \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}$ y 3.

Se puede apreciar que a medida que aumenta α , la moda de la función de densidad de la distribución compuesta se desplaza a la derecha, aumentando también el valor de la función de densidad evaluada en esta moda, mientras que la distribución en sí se vuelve más leptocúrtica, esto es, más aguda en torno a su moda.

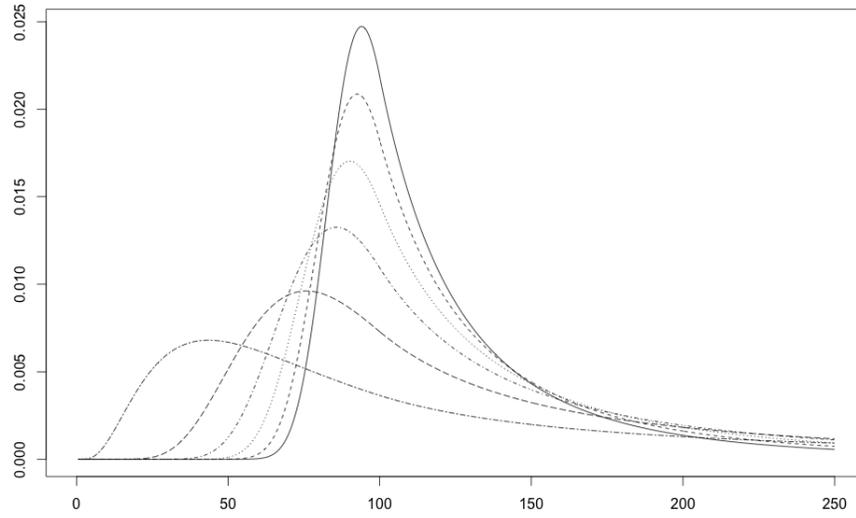


Figura 2.2: Densidades del modelo compuesto para un rango de valores de α

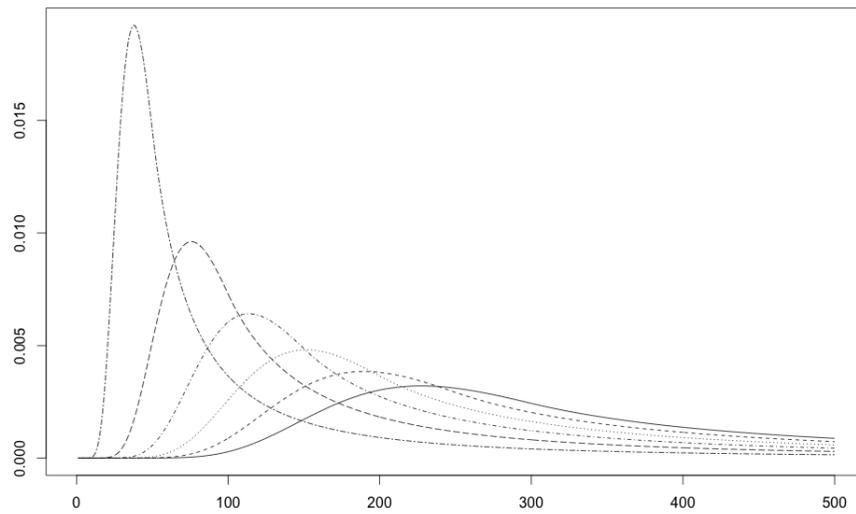


Figura 2.3: Densidades del modelo compuesto para un rango de valores de θ

2.4.2. Cambios en la densidad con el parámetro de corte θ

En la Figura 2.3 se han dibujado funciones de densidad del modelo (2.16) para distintos valores de θ , manteniendo ahora constante el parámetro α . Su valor se fijó en $\alpha = 1$, mientras que el otro parámetro tomó los siguientes valores: $\theta = 50, 100, 150, 200, 250$ y 300 .

Se observa que a medida que aumenta θ la moda de la distribución se desplaza a la derecha, igual que en el caso de que aumente el otro parámetro α , pero el valor de la función de densidad evaluada en esta moda disminuye y además la distribución se vuelve más platocúrtica, o aplanada, cuando crece el parámetro θ .

2.5. Cálculo analítico de la moda

Es posible calcular la moda de forma analítica, la cual además es única, tal y como se verá a continuación. Puesto que la moda será aquel valor en el que la función de densidad sea máxima, la derivada de ésta se tendrá que anular en este punto. Derivando primero la segunda componente de (2.16) se tiene:

$$\frac{d}{dx} \frac{\alpha\theta^\alpha}{[1 + \Phi(k)] x^{\alpha+1}} = -\frac{(\alpha + 1)\alpha\theta^\alpha}{[1 + \Phi(k)] x^{\alpha+2}} \neq 0, \quad \forall x \in \mathbb{R},$$

y por tanto la moda, caso de existir, se tiene que dar en el intervalo $[0, \theta]$.

Derivando ahora la primera componente de (2.16) se tiene:

$$\begin{aligned} & \frac{d}{dx} \left(\frac{\alpha\theta^\alpha}{[1 + \Phi(k)] x^{\alpha+1}} \exp \left\{ -\frac{\alpha^2}{2k^2} \left[\ln \left(\frac{x}{\theta} \right) \right]^2 \right\} \right) \\ &= -\frac{(\alpha + 1)\alpha\theta^\alpha}{[1 + \Phi(k)] x^{\alpha+2}} \exp \left\{ -\frac{\alpha^2}{2k^2} \left[\ln \left(\frac{x}{\theta} \right) \right]^2 \right\} \\ &+ \frac{\alpha\theta^\alpha}{[1 + \Phi(k)] x^{\alpha+1}} \exp \left\{ -\frac{\alpha^2}{2k^2} \left[\ln \left(\frac{x}{\theta} \right) \right]^2 \right\} \frac{-\alpha^2}{2k^2} 2 \ln \left(\frac{x}{\theta} \right) \frac{\theta}{x} \frac{1}{\theta} \\ &= \frac{\alpha\theta^\alpha}{[1 + \Phi(k)] x^{\alpha+2}} \exp \left\{ -\frac{\alpha^2}{2k^2} \left[\ln \left(\frac{x}{\theta} \right) \right]^2 \right\} \left(-(1 + \alpha) - \frac{\alpha^2}{k^2} \ln \left(\frac{x}{\theta} \right) \right). \end{aligned}$$

Si esta derivada se anula, es porque

$$-(1 + \alpha) - \frac{\alpha^2}{k^2} \ln \left(\frac{x}{\theta} \right) = 0,$$

y despejando x , se obtiene la expresión para la moda:

$$x = \theta \exp \left\{ -k^2 \frac{1 + \alpha}{\alpha^2} \right\}. \quad (2.17)$$

Nótese que esa es la única solución posible, y por tanto, se ha demostrado que la distribución compuesta es unimodal.

Otra consecuencia inmediata es que a medida que crece el parámetro θ , la moda también lo hace, algo que se apreció en la Figura 2.3. En vista de que

$$\frac{1 + \alpha}{\alpha^2} = \frac{1}{\alpha^2} + \frac{1}{\alpha}$$

se concluye rápidamente que a mayores valores del parámetro α , le corresponden mayores valores de la exponencial en (2.17), y por tanto la moda aumentará también si crece el parámetro α , algo que también se observó ya en la Figura 2.2.

2.6. Función de distribución

En [Cooray & Ananda, 2005] también se afirma que el modelo desarrollado en este capítulo posee la siguiente función de distribución:

$$F(x) = \begin{cases} \frac{1}{1 + \Phi(k)} \Phi \left(\frac{\alpha}{k} \ln \left(\frac{x}{\theta} \right) + k \right), & 0 < x \leq \theta \\ 1 - \frac{1}{1 + \Phi(k)} \left(\frac{\theta}{x} \right)^\alpha, & \theta < x < \infty \end{cases}. \quad (2.18)$$

Para verlo, sea x tal que $0 < x \leq \theta$. Entonces,

$$\begin{aligned} F(x) &= \int_0^x c f_1(\xi) d\xi = c \mathbb{P} [X \leq x] \\ &= c \mathbb{P} \left[Z \leq \frac{\ln(x) - \mu}{\sigma} \right] = \frac{\Phi \left(\frac{\ln(x) - \mu}{\sigma} \right)}{1 + \Phi(k)} \end{aligned} \quad (2.19)$$

donde, como antes, X es una variable aleatoria que siga una distribución lognormal de parámetros μ y σ y Z denota a una variable aleatoria que siga una distribución normal estándar, y $\Phi(\cdot)$ es su función de distribución.

Nótese que esto implica que

$$F(\theta) = \frac{\Phi\left(\frac{\ln(\theta)-\mu}{\sigma}\right)}{1 + \Phi(k)} = \frac{\Phi(k)}{1 + \Phi(k)} \approx 0,39214992251571, \quad (2.20)$$

o lo que es lo mismo, que aproximadamente el 39,215 % de la muestra procede de la componente lognormal, y por tanto, el restante $1 - F(\theta) \approx 60,785$ % de la muestra se distribuye según la Pareto.

Volviendo a la prueba de la validez de (2.18), sumando $\ln(x)/\sigma$ a ambos lados de (2.12) se tiene que

$$\frac{\ln(x) - \mu}{\sigma} + \frac{\ln(\theta)}{\sigma} = k + \frac{\ln(x)}{\sigma}.$$

Reordenando, y teniendo en cuenta que $k = \alpha\sigma$ se llega a

$$\frac{\ln(x) - \mu}{\sigma} = \frac{\alpha}{k} \ln\left(\frac{x}{\theta}\right) + k.$$

Sustituyendo esto en (2.19), se tiene demostrada la primera igualdad de (2.18).

Sea ahora un x tal que $\theta < x < \infty$. Entonces,

$$\begin{aligned} F(x) &= F(\theta) + \int_{\theta}^x c f_2(\xi) d\xi = F(\theta) + \int_{\theta}^x c \frac{\alpha\theta^\alpha}{\xi^{\alpha+1}} d\xi = F(\theta) + c\theta^\alpha \left[\frac{1}{-\xi^\alpha} \right]_{\xi=\theta}^{\xi=x} \\ &= F(\theta) + c\theta^\alpha \left(\frac{1}{\theta^\alpha} - \frac{1}{x^\alpha} \right) = \frac{\Phi(k)}{1 + \Phi(k)} + \frac{1}{1 + \Phi(k)} \left(1 - \left(\frac{\theta}{x} \right)^\alpha \right) \\ &= 1 - \frac{1}{1 + \Phi(k)} \left(\frac{\theta}{x} \right)^\alpha. \end{aligned}$$

Con esto termina la prueba de la validez de (2.18).

2.7. Comparación con las distribuciones lognormal y Pareto

En la Figura 2.4 se han representado tres funciones de distribución: la de una Pareto de parámetros $\alpha = 1$ y $\theta = 100$; la de una lognormal, con los

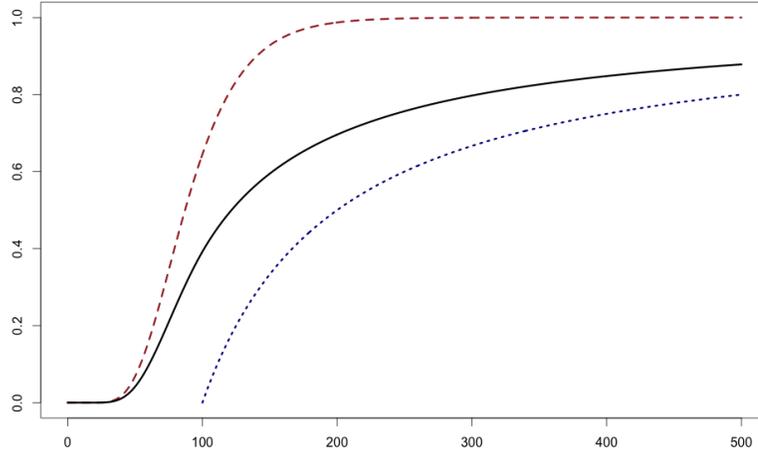


Figura 2.4: Funciones de distribución de una lognormal (línea discontinua marrón), de una Pareto (línea punteada azul) y de su composición (línea continua negra)

parámetros calculados como en (2.14), obteniendo de nuevo

$$\sigma = \frac{k}{\alpha} \approx \frac{0,3722389}{1} = 0,3722389,$$

$$\mu = \ln(\theta) - k\sigma \approx \ln(100) - 0,3722389 \cdot 0,3722389 \approx 4,466608;$$

y la función de distribución del modelo compuesto, usando la expresión (2.18) recién demostrada, con los estos cuatro valores de para los parámetros del modelo compuesto.

Obsérvese que con el modelo compuesto se consigue obtener probabilidad no nula antes de $x = \theta$, a la vez que se consigue una cola más pesada que la que proporcionaría el modelo de la distribución lognormal solamente. La cola del modelo compuesto queda entre la cola de la lognormal y la cola de la Pareto, siendo más pesada que la primera, y más ligera que la segunda.

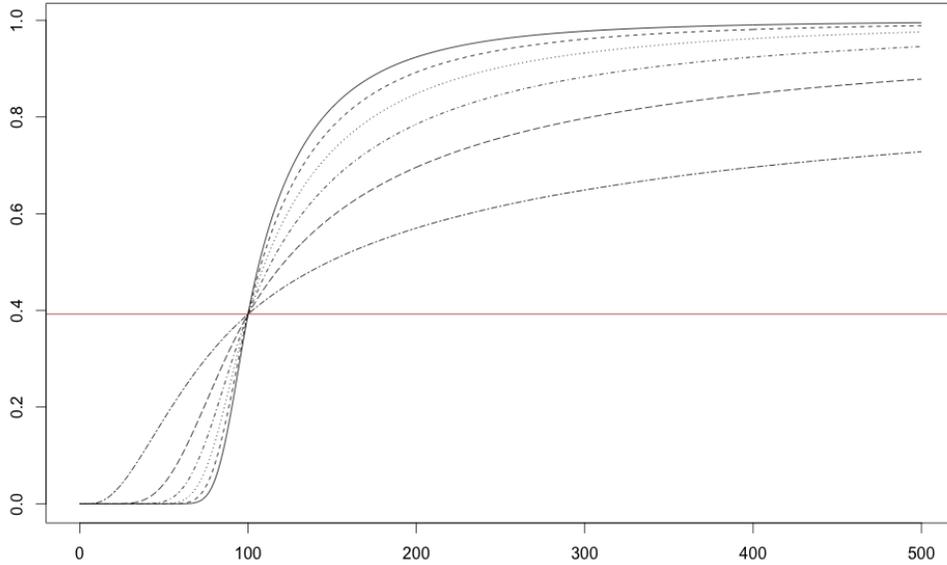


Figura 2.5: Funciones de distribución del modelo compuesto para un rango de valores de α . La línea horizontal marrón marca el valor $\frac{\Phi(k)}{1+\Phi(k)}$.

2.8. Variación de la distribución con los parámetros

2.8.1. Cambios en la distribución con el parámetro de forma α

En la Figura 2.5 se muestran las funciones de distribución del modelo compuesto con el valor de θ fijado en 100, y para un rango de valores de α , concretamente, para $\alpha = \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}$ y 3, igual que antes.

Nótese que independientemente del valor de α , el valor de la función de la distribución en $x = \theta$ es constante e igual a $\frac{\Phi(k)}{1+\Phi(k)}$, tal y como ya se vio de forma analítica en (2.20). Obsérvese también que a medida que aumenta α , la masa probabilística contenida en el intervalo $[0, \theta]$ se desplaza hacia el extremo derecho de dicho intervalo, y la cola de la distribución desde $x = \theta$ se hace cada vez más ligera, desplazándose la probabilidad concentrada en

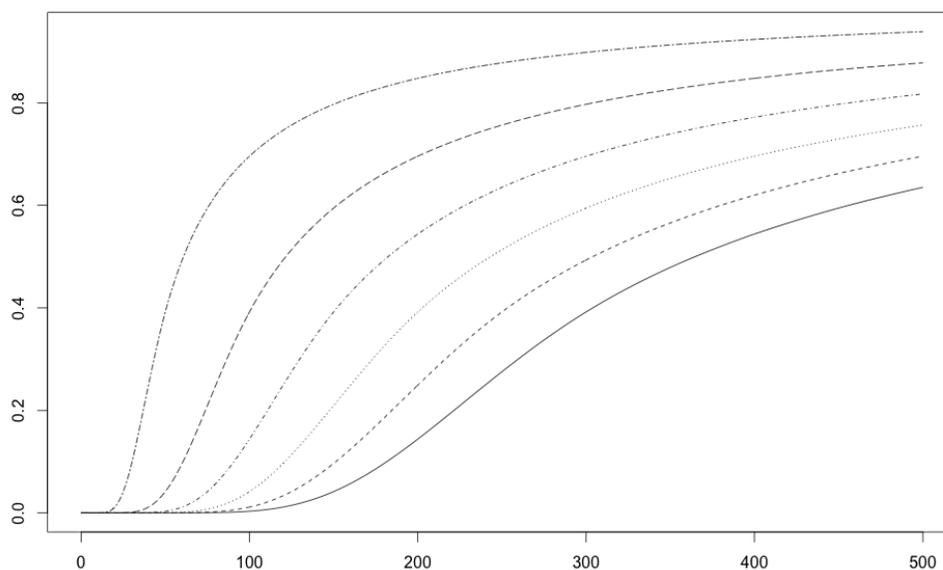


Figura 2.6: Funciones de distribución del modelo compuesto para un rango de valores de θ

el intervalo $[\theta, +\infty)$ hacia el extremo izquierdo, teniendo estos dos hechos por consecuencia que la distribución sea cada vez más leptocúrtica cuando crece α , tal y como ya se vio al discutir sobre el efecto de la variación de este parámetro sobre la forma de las funciones de densidad.

2.8.2. Cambios en la distribución con el parámetro de corte θ

En la Figura 2.6 se representan las funciones de distribución (2.18) con un valor de α fijado en 1, y para un rango de valores de θ , concretamente, para $\theta = 50, 100, 150, 200, 250$ y 300 , de nuevo, igual que en caso de las figuras con las funciones de densidad.

En este caso, se aprecia como el aumentar el parámetro θ hace que toda la distribución se desplace a la derecha, haciendo que la distribución sea cada vez más pesada, tanto a la izquierda como a la derecha del valor $x = \theta$.

2.9. Función cuantil

A partir de la función de distribución (2.18) es posible calcular la función cuantil. Nótese que en vista de que en este caso particular F es monótona y estrictamente creciente, la definición general de la función cuantil,

$$Q(p) = \inf\{x \in \mathbb{R} : p \leq F(x)\}, \quad (2.21)$$

se simplifica en este caso a

$$Q(p) = F^{-1}(p),$$

pues F resulta invertible a raíz de las dos propiedades mencionadas.

Puesto que funciones de distribución son funciones no decrecientes, como consecuencia de (2.20), dado un $p \in [0, \frac{\Phi(k)}{1+\Phi(k)}]$, la función cuantil aplicada a p , $Q(p)$, esto es, aquel valor de x tal que $F(x) = p$ es la solución en x a la ecuación

$$\frac{1}{1 + \Phi(k)} \Phi \left(\frac{\alpha}{k} \ln \left(\frac{x}{\theta} \right) + k \right) = p.$$

A raíz de que la función de distribución de la normal estándar $\Phi(\cdot)$ es una función también monótona y estrictamente creciente, resulta posible calcular su inversa, y vale

$$\frac{\alpha}{k} \ln \left(\frac{x}{\theta} \right) + k = \Phi^{-1} (p (1 + \Phi(k))).$$

La inyectividad del logaritmo permite escribir

$$x = \theta \exp \left\{ \frac{k}{\alpha} \Phi^{-1} (p (1 + \Phi(k))) \right\}.$$

Por otro lado, si $p \in (\frac{\Phi(k)}{1+\Phi(k)}, 1)^2$, es necesario usar la segunda componente de (2.18), y la función cuantil $Q(p)$ será la solución en x a la ecuación

$$1 - \frac{1}{1 + \Phi(k)} \left(\frac{\theta}{x} \right)^\alpha = p.$$

²Nótese que es imposible calcular el cuantil correspondiente a $p = 1$: dado que el soporte de la distribución es toda la semirrecta positiva, la probabilidad completa se alcanza solamente en el infinito.

Con un poco de álgebra se llega a que

$$x = \theta [(1 + \Phi(k))(1 - p)]^{-1/\alpha}.$$

Reuniendo estos resultados, se escribe la función cuantil del modelo (2.16) de Cooray & Ananda para $p \in [0, 1)$:

$$Q(p) = \begin{cases} \theta \exp \left\{ \frac{k}{\alpha} \Phi^{-1} (p(1 + \Phi(k))) \right\}, & 0 \leq p \leq \theta \\ \theta [(1 + \Phi(k))(1 - p)]^{-1/\alpha}, & \theta < p < 1 \end{cases}. \quad (2.22)$$

2.10. Estimación de los cuantiles

Dada una función de distribución teórica el cálculo de sus cuantiles se realiza acorde a la definición (2.21).

Sin embargo, cuando en vez de disponer de la función de distribución se tiene solamente una muestra de la misma, estos cuantiles han de ser estimados.

Existen multitud de métodos para hacerlo, y en [Hyndman & Fan, 1996] se comparan hasta 9 métodos distintos. Aquí, se limitará a introducir la definición 7 del artículo de Hyndman & Fan, la cual es la elegida para efectuar todas las estimaciones necesarias de los cuantiles a lo largo del trabajo.

Esta séptima definición del estimador de los cuantiles, consiste en lo siguiente: dada una muestra aleatoria simple (x_1, x_2, \dots, x_n) y un $p \in (0, 1)$, el cuantil p -ésimo se estima del siguiente modo:

$$\widehat{Q}(p) = x_{(\lfloor h \rfloor)} + (h - \lfloor h \rfloor) (x_{(\lfloor h \rfloor + 1)} - x_{(\lfloor h \rfloor)}),$$

donde h es tal que

$$p = \frac{h - 1}{n - 1},$$

o lo que es lo mismo, $h = (n - 1)p + 1$; $\lfloor h \rfloor$ es la función *piso*, también conocida como función *suelo*, que asigna a un $h \in \mathbb{R}$ el número entero más próximo por defecto, es decir,

$$\lfloor h \rfloor = \text{máx} \{k \in \mathbb{Z} : k \leq h\};$$

y $x_{(i)}$ es el i -ésimo estadístico de orden, esto es, el i -ésimo valor más pequeño de la muestra (x_1, x_2, \dots, x_n) , de tal suerte que

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

2.11. Análisis inicial de los datos daneses

Los *datos daneses* se refieren a un famoso conjunto de datos utilizado ampliamente [Embrechts, Klüppelberg & Mikosch, 1997; McNeil, 1997; Cooray & Ananda, 2005] en investigaciones actuariales consistente en las pérdidas asociadas a incendios extraordinarios ocurridos en Dinamarca entre el año 1980 y el año 1990, expresados en millones de coronas danesas, y ajustados por inflación para que reflejen valores del año 1985. Los datos han sido reescalados por razones comerciales. Este conjunto de datos se puede encontrar en [Davison, 2013].

El tamaño de la muestra es 2492 siniestros, con una única variable: la cuantía de la pérdida en millones de coronas danesas. No hay ningún valor perdido. Los datos se han estudiado sin realizar ninguna transformación adicional. Algunos estadísticos descriptivos se muestran en la Tabla 2.1. Tal y como en [Joanes & Gill, 1998], los estimadores para la asimetría y curtosis se han tomado como

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} \cdot \frac{m_3}{m_2^{3/2}}$$

y

$$G_2 = \frac{n-1}{(n-2)(n-3)} \left[(n+1) \left(\frac{m_4}{m_2^2} - 3 \right) + 6 \right]$$

respectivamente, donde

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r$$

es el estimador natural, aunque sesgado, del r -ésimo momento centrado de una muestra aleatoria simple (x_1, x_2, \dots, x_n) , donde como es usual, \bar{x} denota a la media muestral

$$\frac{1}{n} \sum_{i=1}^n x_i.$$

En la Tabla 2.2 se muestran algunos cuantiles empíricos de los datos daneses. La estimación de los cuantiles se ha realizado utilizando la definición 7 de [Hyndman & Fan, 1996] introducida en 2.10.

En la Figura 2.7 se muestran dos histogramas elaborados con *R* [R Core Team, 2015] sobre estos datos de siniestros de incendios daneses: un histograma completo y otro parcial, en el que solamente se representa la parte del histograma correspondiente a siniestros por debajo de 20 millones de coronas danesas.

Número de observaciones	2492
Número de variables	1
Medición mínima	0,3134
Primer cuartil	1,1572
Mediana	1,6339
Tercer cuartil	2,6455
Medición máxima	263,2504
Rango intercuartil	1,4883
Rango de la muestra	262,9370
Media	3,0630
Desviación estándar	7,9767
Coficiente de variación	2,6045
Varianza	63,6278
Asimetría	19,9081
Curtosis	547,6742

Tabla 2.1: Estadísticos descriptivos de los datos daneses

Probabilidad	Cuantil	Probabilidad	Cuantil
5 %	0,9047	70 %	2,3101
10 %	0,9643	75 %	2,6455
15 %	1,0204	80 %	3,1419
20 %	1,0949	85 %	3,8843
25 %	1,1572	90 %	5,0804
30 %	1,2502	95 %	8,4063
35 %	1,3291	96 %	10,5303
40 %	1,4226	97 %	12,8271
45 %	1,5165	98 %	17,1587
50 %	1,6339	99 %	24,6138
55 %	1,7358	99,5 %	32,4313
60 %	1,8816	99,9 %	105,8931
65 %	2,0498	99,99 %	235,6408

Tabla 2.2: Cuantiles empíricos de los datos daneses

En el primero se puede apreciar una característica que aparece frecuentemente en los datos relativos a siniestros de seguros no vida: la presencia de una cola extremadamente pesada. La pérdida media es igual a 3,0630 millones, y más del 95 % de las observaciones se sitúan por debajo de 9 millones. Sin embargo, el máximo de la muestra es 263,3000 millones. Esto es, a pesar de que la gran masa de siniestros son relativamente moderados, de vez en cuando se obtiene una observación ciertamente extrema.

En el histograma parcial se observa la otra característica usual de este tipo de datos: la asimetría positiva. Esto hace deseable que el modelo paramétrico elegido para ajustar a los datos siniestrales presente esta propiedad. Además de esto, es preferible que el modelo tenga una moda estrictamente positiva, en vista de que los datos suelen poseerla también, tal y como se aprecia en la Figura 2.7b.

En la sección 2.12 se muestra el programa en *R* que ajusta el modelo (2.16) de Cooray & Ananda al conjunto de datos de los siniestros derivados de incendios en Dinamarca.

2.12. Ajuste a los datos daneses

Lo primero que se hace es cargar los paquetes necesarios para la ejecución del subsiguiente código: “SMPracticals” [Davison, 2013], “its” [Group & Securities, 2009] y “Bhat” [Luebeck & Meza, 2013].

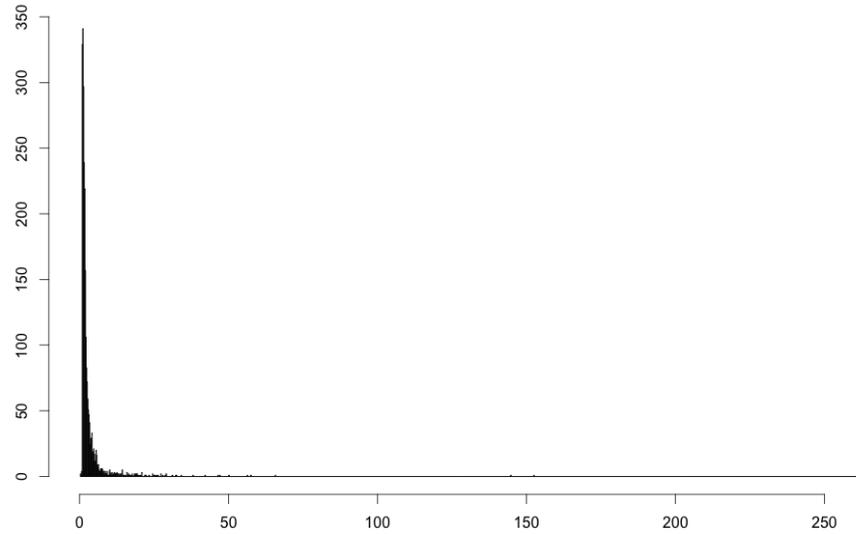
“**SMPracticals**” comprende el conjunto de datos “danish”, que contiene las pérdidas asociadas a los incendios en Dinamarca.

“**its**” contiene la función “core”, usada en este programa simplemente para extraer las cuantías de los siniestros del conjunto de datos “danish”.

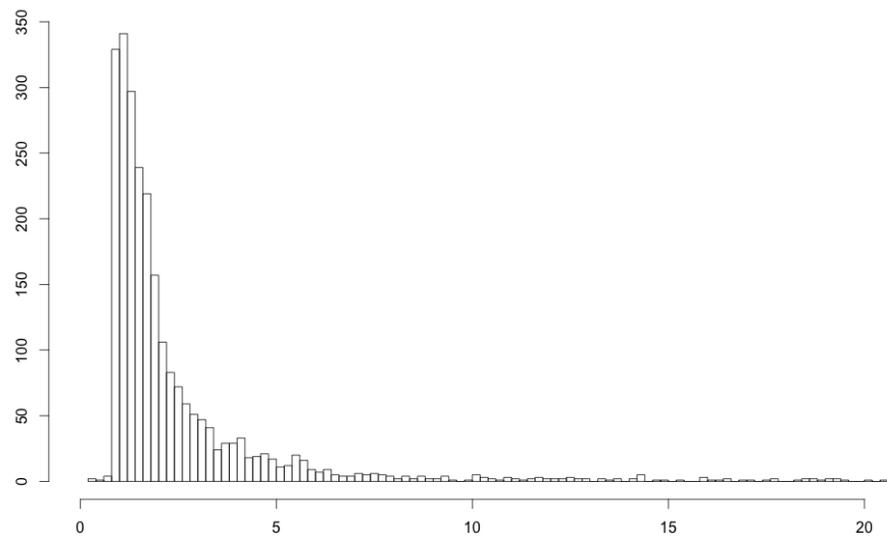
“**Bhat**” contiene la función “dfp”, la cual implementa el algoritmo de optimización de Davidon-Fletcher-Powell. Se usará para maximizar la verosimilitud.

```
library(SMPracticals)
library(its)
library(Bhat)
```

Se extraen las cuantías de los siniestros de incendios daneses:



(a) Histograma completo



(b) Histograma parcial correspondiente a valores inferiores a 20 millones

Figura 2.7: Histogramas de los datos de siniestros derivados de incendios en Dinamarca

```
siniestrosDaneSES=core(danish)
```

Con el objetivo de evitar un modelo sobreajustado, en vez de utilizar toda la muestra disponible para estimar los parámetros del modelo, se ha decidido dedicar un 80 % de la muestra para este propósito de la estimación del modelo, y el restante 20 % de la muestra para validar el ajuste [Arlot & Celisse, 2010].

Se calcula cuál sería el tamaño necesario de esta muestra de entrenamiento para que represente –aproximadamente– un 80 % de la muestra total de 2492 siniestros de incendios daneses. Puesto que el valor $0,8 \cdot 2492$ no es entero, es necesario redondearlo con la función `round`. Dicho valor se almacena en la variable `trainingSampleSize`:

```
trainingSampleSize=round(0.8*length(siniestrosDaneSES))
```

Como es lógico, la muestra del $\sim 80\%$ será escogido al azar dentro de la muestra completa de los datos daneses. Sin embargo, para que los resultados sean reproducibles, se fija una semilla que permitirá llegar a las mismas estimaciones en otras ejecuciones de este, y subsiguientes, programas:

```
set.seed(1234)
```

Con la ayuda de la función `sample`, se elige una muestra aleatoria de tamaño `trainingSampleSize` –variable que recoge el tamaño de la muestra de entrenamiento– dentro del conjunto de $\{1, 2, \dots, 2492\}$. Tomando dentro de la muestra total de 2492 aquellos elementos cuyos índices corresponden a esta muestra aleatoria almacenada en la variable `trainingIndexes`, se tendrá el conjunto de datos de entrenamiento buscado, el cual se almacena en la variable `trainingData`:

```
trainingIndexes=sample(x=length(siniestrosDaneSES),
                      size = trainingSampleSize)
trainingData=siniestrosDaneSES[trainingIndexes]
```

Tal y como se ha adelantado, el restante 20 % de la muestra total se utilizará para validar el modelo estimado. Utilizando los elementos de la muestra global correspondientes a índices que no están contenidos en el conjunto `trainingIndexes`, se obtiene este conjunto complementario que se usará para la verificación de que el modelo ajustado es correcto o no:

```
testData=siniestrosDanese[-trainingIndexes]
```

Se asigna a n la longitud del conjunto de datos de entrenamiento. Esto es, n es el número de observaciones sobre los siniestros utilizados para ajustar el modelo.

```
n=length(trainingData)
```

Se asigna a k el valor constante calculado como la solución de la ecuación $\exp\{-\frac{k^2}{2}\} = \sqrt{2\pi k}$:

```
k=uniroot(function (k) exp(-k^2/2) - sqrt(2*pi)*k,
           interval = c(0,5))$root
```

Se programa la función de verosimilitud del modelo de Cooray & Ananda para la muestra de siniestros de incendios daneses, usando la función de densidad (2.16) obtenida en el desarrollo teórico:

```
lkh=function (x) {
  alpha=x[1]
  theta=x[2]
  lkh=0
  for (i in 1:length(trainingData)) {
    xi=trainingData[i]
    if(xi < theta){
      f=alpha*(theta^alpha)*
        exp(-alpha*alpha*log(xi/theta)*
            log(xi/theta)/(2*k*k))/
        ((1+pnorm(k))*(xi^(alpha+1)))
      lkh=lkh+log(f)
    }else{
      f=alpha*(theta^alpha)/((1+pnorm(k))*(xi^(alpha+1)))
      lkh=lkh+log(f)
    }
  }
  return(-lkh)
}
```

Se declara una lista relativa a los parámetros α y θ a estimar del modelo. Se proporcionan unas estimaciones iniciales de ambos parámetros: $\alpha = 1$ y $\theta = 1000$. Se definen cotas inferiores y superiores, tanto para α , como para θ . La cota inferior para ambos se fija en 0, pues han de ser número positivos. La cota superior para α se coloca en 5, pues con este valor la distribución seguramente tendría una cola mucho más ligera de la que presentan los datos. Para θ , la cota superior se establece en 10000, pues se considera que es un número suficientemente alto para que ya empiece a gobernar la cola de la Pareto.

```
x = list(label=c("alpha", "theta"), est=c(1, 10), low=c(0, 0),
         upp=c(5, 10000))
```

Se aplica el mencionado algoritmo de optimización de Davidon-Fletcher-Powell, para obtener las estimaciones máximo verosímiles. Los resultados del algoritmo se almacenan en el vector `results`:

```
results = dfp(x, f=lkh)
```

Dichos resultados se imprimen automáticamente, y se muestran de forma parcial a continuación:

```
## Optimization Result:
## iter: 9   fmin: 3144.059   nfcn: 82
##
##  label estimate  low      high
## 1 alpha 1.4151789 1.3575914 1.4742205
## 2 theta 1.3850275 1.3553603 1.415344
```

Se aprecia que los valores de los parámetros del modelo (2.16) obtenido maximizando la verosimilitud son

$$\hat{\alpha} = 1,4151789 \text{ y } \hat{\theta} = 1,3850275, \quad (2.23)$$

siendo los intervalos de confianza al 95 % para los parámetros α y θ respectivamente de

$$[1,3575914; 1,4742205] \text{ y } [1,3553603; 1,415344].$$

Esto es, se tiene que

$$\mathbb{P}\{\alpha \in [1,3575914; 1,4742205]\} = \mathbb{P}\{\theta \in [1,3553603; 1,415344]\} = 95\%.$$

Otro valor interesante que aparece en la última salida es la logverosimilitud *negativa*³, que toma el valor de 3144,059 en este caso. Por tanto, el valor de la logverosimilitud alcanzada tras la optimización es

$$-3144,059. \tag{2.24}$$

En la Figura 2.8 se representa el histograma parcial del subconjunto de la muestra dedicado a la validación del modelo, concretamente, el correspondiente a las pérdidas inferiores a 20 millones, junto con la función de densidad del modelo de Cooray & Ananda ajustado, esto es, utilizando los valores (2.23).

En la Figura 2.9 se han trazado la función de distribución empírica de esta submuestra en línea negra sólida junto con la función de distribución del modelo ajustado, utilizando los valores (2.23), en línea roja discontinua. La estimación de la función de distribución empírica se ha realizado utilizando la definición clásica: dada una muestra aleatoria simple (x_1, x_2, \dots, x_n) , la función de distribución empírica se define como

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i \leq t\}},$$

donde $1_{\{x_i \leq t\}}$ es la función indicadora del conjunto $\{x \in \mathbb{R} : x \leq t\}$, esto es,

$$1_{\{x_i \leq t\}} = \begin{cases} 1, & \text{si } x_i \leq t \\ 0, & \text{si } x_i > t \end{cases}.$$

2.13. Prueba χ^2

Desde que Pearson desarrollara en 1900 la prueba de χ^2 [Pearson, 1900], ésta se convirtió en un test de bondad de ajuste ampliamente utilizado, tanto

³ Nótese que en la definición de la función `lkh` se ha programado que devuelva *menos* el sumatorio de los logaritmos de las verosimilitudes individuales.

El procedimiento `dfp` *minimiza* la función que se le pasa como argumento, y puesto que maximizar la verosimilitud es equivalente a minimizar la *menos-verosimilitud*, `lkh` devuelve la verosimilitud cambiada de signo para que pueda ser pasada directamente al procedimiento `dfp` y que éste obtenga los parámetros que maximizan la función de verosimilitud.

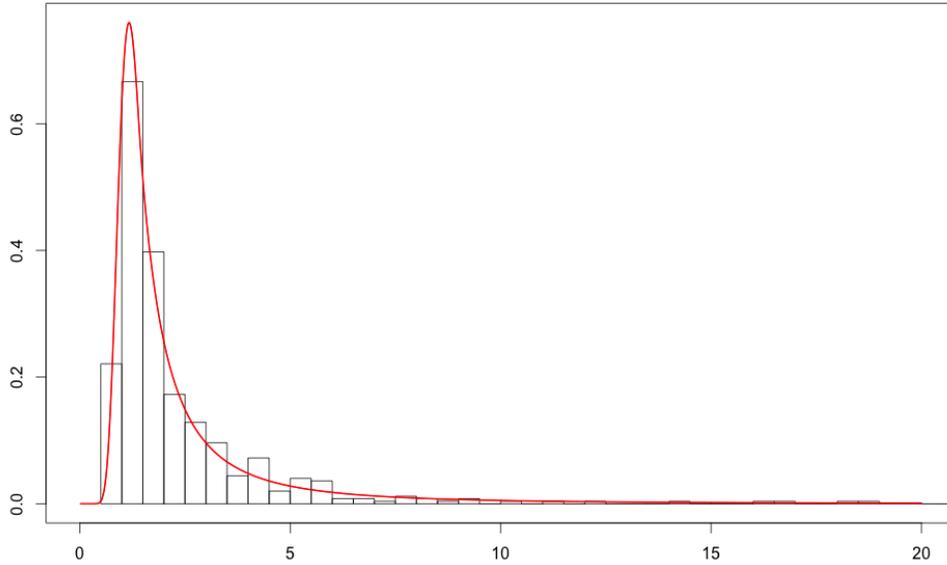


Figura 2.8: Histograma parcial correspondiente a valores inferiores a 20 millones de los incendios daneses con la función de densidad del modelo de Cooray & Ananda ajustado

en artículos de investigación [Cooray & Ananda, 2005; Scollnik, 2007; Vernic & Teodorescu, 2009; Teodorescu, 2010] como en libros de texto: véase, por ejemplo [Klugman, Panjer & Willmot, 2012, p. 333] o [Davison, 2003, p. 133].

Esta prueba se utiliza para comprobar la validez de un modelo que prediga las frecuencias esperadas de los datos en unas categorías previamente especificadas, contrastando estas frecuencias esperadas frente a las frecuencias realmente observadas en los datos. Concretamente, el estadístico del test consiste en la suma de diferencias entre las frecuencias observadas y las esperadas de cada categoría especificada, elevadas al cuadrado –por lo que son siempre positivas– y apropiadamente reescaladas –dividiendo entre el valor de las frecuencias esperadas–.

Formalmente, si hay k categorías en las que se clasifican las observaciones de la muestra, y O_i denota el número de datos de la muestra observados dentro de la categoría i -ésima, mientras que E_i representa el número de datos

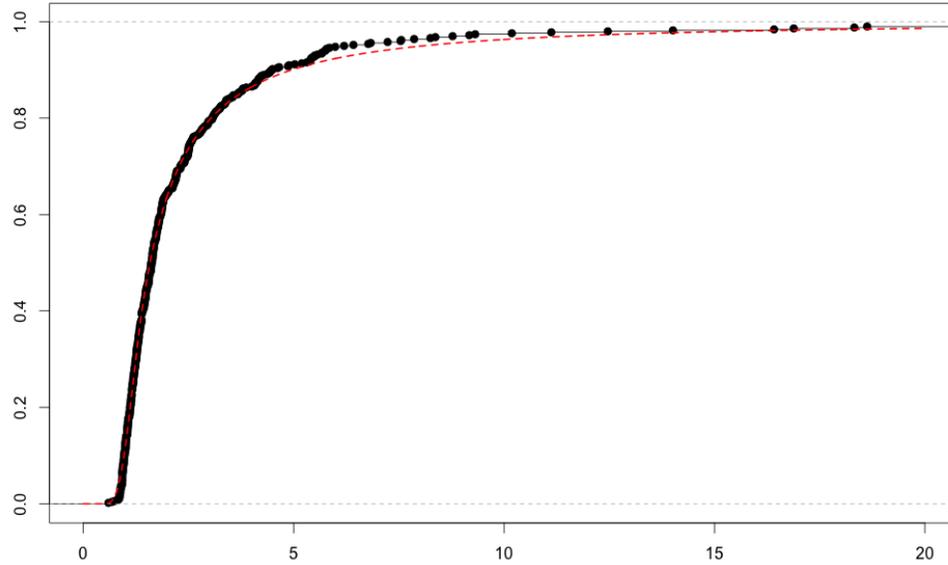


Figura 2.9: Función de distribución empírica (línea negra sólida) junto con la función de distribución del modelo de Cooray & Ananda ajustado (línea roja discontinua) en el tramo de los valores inferiores a 20 millones de los incendios daneses

esperados en esta misma categoría i -ésima bajo la asunción de la validez del modelo, el estadístico de la prueba de χ^2 es

$$\tilde{\chi}^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}. \quad (2.25)$$

Bajo la hipótesis nula de que el modelo es correcto, este estadístico se aproxima asintóticamente a una distribución χ^2 de $k-p-1$ grados de libertad, siendo k el número de categorías especificadas y p el número de parámetros estimados en el modelo ajustado.

Cuanto más se parezcan las frecuencias observadas a las esperadas, menor será el valor del estadístico, debiéndose las diferencias entre los recuentos empíricos y los teóricos al puro azar. Sin embargo, si el valor del estadístico es superior al valor crítico de la distribución χ^2 al nivel de confianza fijado,

habrá evidencia estadística significativa para rechazar la hipótesis nula de que los datos observados obedecen el modelo en cuestión, puesto que será poco probable que las diferencias entre los recuentos empíricos y teóricos hayan ocurrido por casualidad, aumentando la verosimilitud de la hipótesis alternativa que afirma que los datos observados *no* se distribuyen según el modelo postulado.

La aplicación de este test a modelos de variables continuas, como es el caso del modelo de cuantías de siniestros aquí considerado, se realiza previa clasificación de los datos observados en una partición⁴ de un intervalo que contenga el rango de la muestra que se esté usando para testear el modelo ajustado.

Así, puesto que el rango de los datos daneses es $[0,3134; 263,2504]$, para aplicar la prueba χ^2 a la hipótesis de que los datos daneses contenidos en la muestra de test, esto es, los siniestros de la variable `testData`, se distribuyen según el modelo (2.16) ajustado, con los valores de los parámetros dados en (2.23), es necesario clasificarlos en unos intervalos. Por ejemplo, estos datos se pueden clasificar en los 12 intervalos siguientes:

$$\begin{aligned} & (0; 1,25], \quad (1,25; 1,75], \quad (1,75; 2,25], \quad (2,25; 2,75], \quad (2,75; 3,25], \\ & (3,25; 3,75], \quad (3,75; 4,25], \quad (4,25; 4,75], \quad (4,75; 5,25], \quad (5,25; 5,75], \quad (2.26) \\ & \quad (5,75; 6,25], \quad (6,25; \infty), \end{aligned}$$

los cuales son una partición de la semirrecta $(0, \infty)$ que contiene por completo a las 2492 entradas de los datos daneses, y por ende, también contendrá por completo a la muestra para el test.

Afortunadamente, R proporciona el comando `cut` que acepta como argumentos un vector de datos y los límites de la partición en la que se desean clasificar los antedichos datos. Lo primero es almacenar los límites mostrados en (2.26) en un vector:

⁴Una *partición* de un conjunto A es una familia de conjuntos $\{A_i\}_{i \in I}$ de tal forma que

- $A_i \subseteq A, \forall i \in I$ (los elementos A_i de la familia son subconjuntos del conjunto A);
- $A_i \cap A_j = \emptyset$, si $i \neq j$ (los elementos de la familia son disjuntos dos a dos); y
- $A = \bigcup_{i \in I} A_i$ (la familia cubre por completo al conjunto A).

```
limites=seq(from=1.25,to=6.25,by=0.5)
limites=c(0,limites,Inf)
```

Después, se aplica el mencionado comando `cut` para clasificar la muestra para el test en las 12 categorías dadas por (2.26):

```
siniestrosDanesesAgrupados=cut(x = testData,
                               breaks = limites)
```

La variable `siniestrosDanesesAgrupados` almacena 498⁵ entradas, una por cada valor de la muestra para el test de los incendios daneses, indicando a qué categoría pertenece cada observación. Puesto que se está interesado en conocer el recuento de valores de la muestra dentro de cada una de las 12 categorías, se utiliza el comando `table` que realiza precisamente este recuento:

```
freqObservadas=table(siniestrosDanesesAgrupados)
```

Ahora la variable `freqObservadas` almacena las frecuencias observadas O_i para cada una de las 12 categorías.

La otra componente necesaria para seguir con el cálculo del estadístico $\tilde{\chi}^2$ dado por (2.25) son las frecuencias esperadas de cada una de las categorías E_i acordes al modelo ajustado.

Si se denota por C_i a la categoría i -ésima, y por X a la variable aleatoria cuya distribución es el modelo ajustado para explicar la muestra aleatoria simple (x_1, x_2, \dots, x_m) , se tiene que la frecuencia esperada de la categoría C_i es

$$E_i = m \cdot \mathbb{P}[X \in C_i],$$

y cuando C_i consiste en un intervalo de la forma $(a, b]$, como es el caso de las categorías (2.26), esta expresión se puede escribir como

$$= m \cdot [F_X(b) - F_X(a)],$$

⁵La muestra total cuenta con 2492 registros. El tamaño de la muestra para el ajuste ha sido el entero más próximo al producto $0,8 \cdot 2492 = 1993,6$, siendo por tanto, igual a 1994. De aquí se sigue que el tamaño de la muestra para el test sea de $2492 - 1994 = 498$ siniestros.

donde F_X denota a la función de distribución del modelo X .

Por tanto, el cálculo de las frecuencias esperadas E_i pasa por la programación de la función de distribución del modelo de Cooray & Ananda dado por (2.18):

```
cdfCooray<-function(x,alpha,theta){
  if(x<theta){
    pnorm(alpha*log(x/theta)/k+k)/(1+pnorm(k))
  }else{
    1-((theta/x)^alpha)/(1+pnorm(k))
  }
}

cdfCooray<-Vectorize(cdfCooray)
```

El comando `Vectorize` sirve simplemente para convertir la función de distribución `cdfCooray` en una función vectorial que admita por parámetro un vector de valores y devuelva otro vector cuyas componentes son los valores de la función de distribución aplicada a cada una de los valores del vector de entrada.

Con esto, se calculan las probabilidades teóricas de que una observación resulte dentro de cada una de los 12 intervalos como la diferencia de la función de distribución evaluada en los extremos superior e inferior del intervalo.

```
probsTeoricas=numeric(length = length(limites)-1)
for (i in 1:length(probsTeoricas)) {
  probsTeoricas[i]=cdfCooray(limites[i+1],alpha,theta)-
  cdfCooray(limites[i],alpha,theta)
}
```

Calculadas estas probabilidades, multiplicando por m , se consiguen obtener las frecuencias esperadas E_i .

Obsérvese que en este caso $m = \text{length}(\text{testData})$.

```
freqEsperadas=length(testData)*probsTeoricas
```

Conocidas las frecuencias observadas y esperadas, se procede a calcular el estadístico $\tilde{\chi}^2$ según la fórmula (2.25):

```
estadistico=sum(((freqObservadas-freqEsperadas)^2)/
                (freqEsperadas))
```

Finalmente, se calcula la probabilidad de observar un valor del estadístico igual o mayor al obtenido, utilizando la aproximación asintótica a la distribución χ^2 ya comentada. Los grados de libertad son $k - p - 1$, donde $k = \text{length}(\text{freqObservadas})$ es el número de categorías definidas y $p = 2$ es el número de parámetros estimados en el modelo de Cooray & Ananda ajustado.

Dicho valor es popularmente conocido como *p-valor*.

```
pchisq(estadistico,df=length(freqObservadas)-2-1,
        lower.tail=FALSE)

## [1] 0.1074995
```

El p-valor obtenido es igual a

$$p_{\chi^2} = 0,1074995, \quad (2.27)$$

lo que no permite rechazar la hipótesis nula de la validez del modelo (2.16) para explicar los datos daneses a cualquiera de los niveles de confianza usuales, incluso utilizando un nivel tan alto como 0,1.

Más concretamente, el resultado obtenido significa que en el $\approx 10,7\%$ de los casos en los que los datos se distribuyen según el modelo de Cooray & Ananda, se obtienen diferencias entre las frecuencias observadas y esperadas iguales, o *superiores*, a las obtenidas en el caso de los datos daneses y el modelo ajustado con los valores (2.23), donde la *magnitud* de las diferencias se mide de acuerdo al estadístico $\tilde{\chi}^2$ para las categorías (2.26).

Este resultado va en línea con el ajuste satisfactorio que se ha observado gráficamente en las Figuras 2.8 y 2.9.

Ejecutando el siguiente comando, se obtienen las frecuencias esperadas para los intervalos (2.26) según el modelo ajustado:

```
print(freqEsperadas)
```

Dichas frecuencias se muestran en la Table 2.3. En vista de que una de las frecuencias esperadas es inferior a 5, es aconsejable realizar la corrección

(0; 1,25]	92,999610	(1,25; 1,75]	158,198498
(1,75; 2,25]	78,837653	(2,25; 2,75]	43,812375
(2,75; 3,25]	27,357501	(3,25; 3,75]	18,460922
(3,75; 4,25]	13,168204	(4,25; 4,75]	9,792771
(4,75; 5,25]	7,523079	(5,25; 5,75]	5,931937
(5,75; 6,25]	4,778249	(6,25; ∞)	37,139201

Tabla 2.3: Frecuencias esperadas para los intervalos (2.26) según el modelo de Cooray & Ananda

de Yates [Yates, 1934] al efectuar la prueba χ^2 , modificando el estadístico $\tilde{\chi}^2$ del siguiente modo:

$$\tilde{\chi}^2 = \sum_{i=1}^k \frac{(|O_i - E_i| - 0,5)^2}{E_i}.$$

La implementación en R de esta modificación es inmediata, siendo la siguiente:

```
estadisticoCorregido=sum(((abs(freqObservadas-freqEsperadas)
-0.5)^2)/freqEsperadas)
```

Calculando el p-valor de nuevo, tras la corrección, se obtiene el siguiente resultado:

```
pchisq(estadisticoCorregido,df=length(freqObservadas)-2-1,
lower.tail=FALSE)
## [1] 0.2328648
```

Tal y como se puede observar, tras la corrección de Yates el ajuste se ha vuelto aún más significativo, obteniéndose ahora un p-valor de

$$p_{\chi^2} = 0,2328648.$$

2.14. Prueba G

Otro test de bondad de ajuste, menos popular que la prueba χ^2 , pero también utilizado en la literatura científica, tanto en artículos [Hoey, 2012],

como en libros de texto [McDonald, 2014, pp. 53-58] es el test G .

El estadístico del test en cuestión es

$$G = 2 \sum_{i=1}^k O_i \ln \left(\frac{O_i}{E_i} \right),$$

donde se utiliza la misma nomenclatura introducida al hablar de la prueba χ^2 :

- k es el número de categorías en los que se clasifican los datos de la muestra;
- O_i es la frecuencia observada en la categoría i -ésima; y
- E_i es la frecuencia esperada en la categoría i -ésima según el modelo ajustado.

Bajo la hipótesis nula de la validez del modelo, el estadístico G sigue la misma distribución que el estadístico $\tilde{\chi}^2$, esto es, una distribución χ^2 con $k - p - 1$ grados de libertad, siendo p el número de parámetros estimados en el modelo ajustado a la muestra.

Como ya se tienen calculadas las frecuencias observadas y esperadas del test χ^2 anterior, es inmediato calcular el estadístico G :

```
estadisticoG=2*sum(freqObservadas*
                    log(freqObservadas/freqEsperadas))
```

El cálculo del p-valor se realiza de un modo similar al caso de la prueba χ^2 :

```
pchisq(estadisticoG,df=length(freqObservadas)-2-1,
        lower.tail = FALSE)

## [1] 0.1456099
```

Similarmente al test anterior, el p-valor de

$$p_G = 0,1456099 \tag{2.28}$$

no permite rechazar la hipótesis nula de la validez del modelo.

2.15. Gráfico QQ

El gráfico QQ⁶ es un método gráfico para comparar dos distribuciones de probabilidad. Cada punto del gráfico QQ se construye tomando un $p \in (0, 1)$ y calculando los cuantiles p -ésimos de ambas distribuciones. El cuantil de una de las distribuciones se asigna a la abscisa, y el otro a la ordenada, obteniendo así las dos coordenadas de un punto del gráfico QQ.

Así, si F denota a la función de distribución de la primera distribución de probabilidad, y G a la de la segunda, y ambas son invertibles, el gráfico QQ de estas dos distribuciones es la curva paramétrica

$$\left\{ \left(F^{-1}(p), G^{-1}(p) \right) : p \in (0, 1) \right\}. \quad (2.29)$$

Si la ley de ambas de distribuciones es la misma, esta curva paramétrica estará muy próxima a la línea $y = x$, y si las dos distribuciones están linealmente relacionadas, la curva (2.29) será aproximadamente una línea recta, pero no necesariamente la recta $y = x$.

El gráfico QQ puede ser usado también como un método para evaluar la bondad de ajuste de un modelo teórico a unos datos empíricos. En este caso, los cuantiles correspondientes al modelo se calculan usando la función cuantil teórica, y los correspondientes a la muestra han de ser estimados, usando algún método como el explicado en 2.10.

En este caso, la curva continua (2.29) se ve sustituida por una colección de puntos finita correspondientes a los cuantiles calculados para un conjunto discreto de probabilidades $\{p_i\}_{i=1}^N$ contenido en el intervalo $(0, 1)$. Así, se puede tomar como el conjunto $\{p_i\}_{i=1}^N = \{0,01k : 1 \leq k \leq 99\}$, y estimar los cuantiles con estas probabilidades, obteniendo los *percentiles*.

Cuanto más cerca se localicen los puntos correspondientes a estos cuantiles discretos de la recta $y = x$, mejor ajuste habrá.

Si estos cuantiles discretos aparecen arqueados en el gráfico QQ, esto es una señal de un ajuste pobre, indicando que una de las distribuciones es más asimétrica o que posee colas más pesadas que la otra.

Estos gráficos QQ son frecuentemente preferidos para valorar la bondad de ajuste sobre otras técnicas analíticas, como por ejemplo las pruebas χ^2 y G realizada previamente en las secciones 2.13 y 2.14. En [NIST/SEMATECH, Croarkin, Tobias & Zey, 2001] se puede leer:

⁶*Q* de *quantile*: cuantil, en inglés.

«The q-q plot can provide more insight into the nature of the difference than analytical methods such as the chi-square and Kolmogorov-Smirnov 2-sample tests.»

En la Figura 2.10 se representa el gráfico QQ de los datos daneses frente al modelo (2.16) de Cooray & Ananda ajustado. Los cuantiles teóricos fueron calculados de acuerdo a la función cuantil (2.22) obtenida en la sección 2.9 y se representan en el eje horizontal; mientras que los empíricos se obtuvieron usando la definición 7 de [Hyndman & Fan, 1996] explicada en 2.10 y aparecen en el eje vertical del gráfico.

Casi todos los puntos se localizan bastante cercanos a la línea $y = x$, aunque hay un punto que se queda muy alejado de dicha recta, en la esquina superior derecha. Esto indica que uno de los valores presentes en la muestra del test tiene un valor extremo dentro del modelo ajustado.

Teniendo en cuenta la naturaleza de los datos, un comportamiento así es de esperar, y puede ser interesante el estudio de esa observación extrema dentro de la muestra global, y la consideración de incluirla dentro de la muestra de entrenamiento dado su posiblemente alta importancia dentro de la muestra.

Con respecto a los demás valores de la muestra del test, se puede afirmar que se aprecia un ajuste satisfactorio.

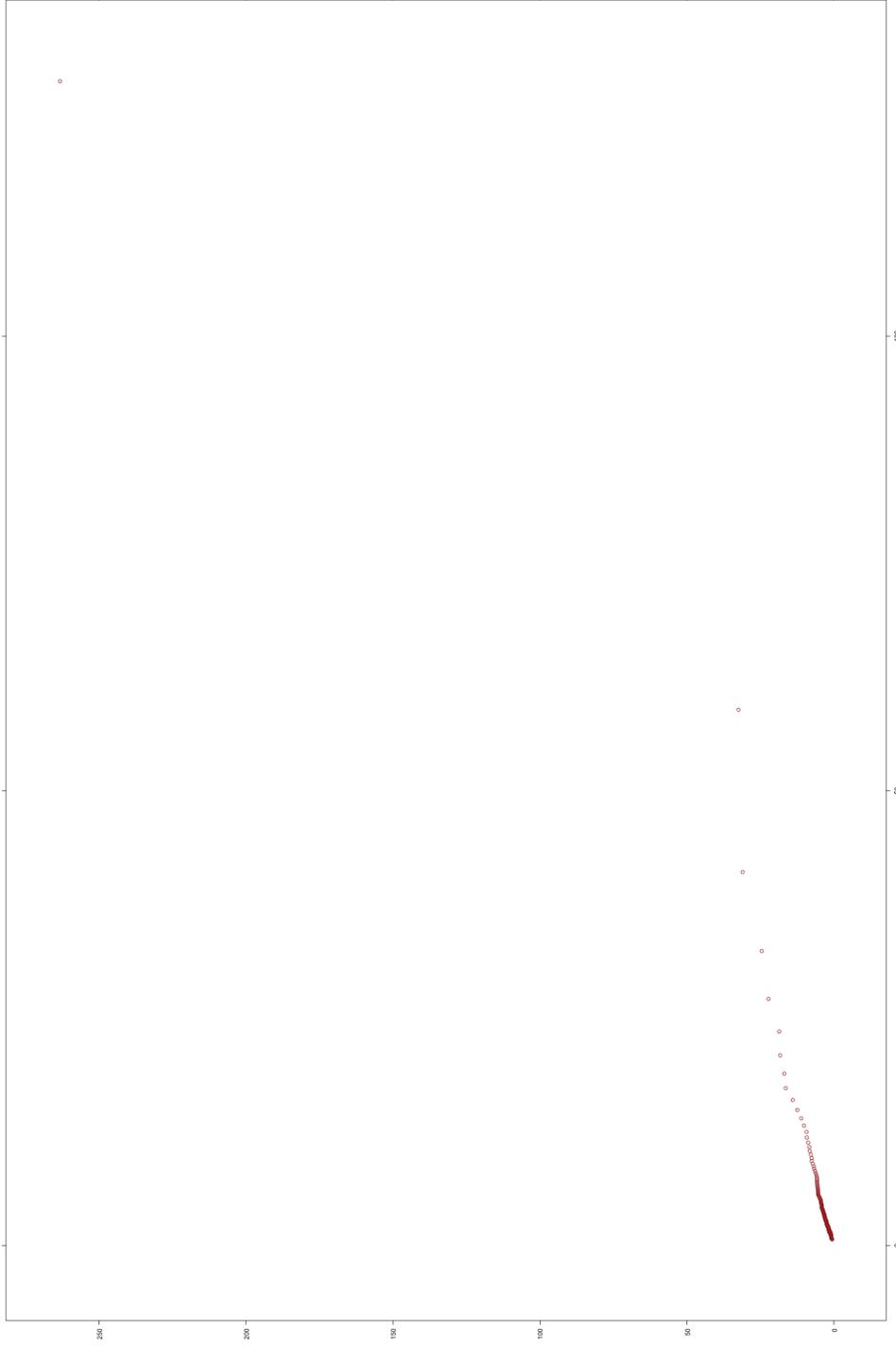


Figura 2.10: Gráfico QQ de los datos daneses frente al modelo de Cooray & Ananda ajustado

Capítulo 3

Modelo basado en las distribuciones lognormal y Pareto con pesos variables

La Figura 2.8 donde se muestra el histograma de los datos daneses frente a la función de densidad del modelo de Cooray & Ananda ajustado y la Figura 2.9 en la cual se representan las funciones de distribución empírica de los datos daneses y del modelo, junto con los p-valores (2.27) y (2.28) de las pruebas χ^2 y G indican un ajuste satisfactorio del modelo (2.16) a los datos de los siniestros derivados de incendios daneses.

A pesar del correcto ajuste a los datos daneses, el modelo de Cooray & Ananda presenta una limitación ya adelantada en el Capítulo 2, concretamente, en la ecuación (2.20): según este modelo el porcentaje de la muestra explicada por la componente lognormal de la cabeza de la distribución es siempre $\approx 39,215\%$ y la proporción de los datos explicados por la componente Pareto de la cola es $\approx 60,785\%$.

Esta propiedad parece demasiado restrictiva de cara a ajustar el modelo a datos en los que el comportamiento de la cola de la muestra empiece a asemejarse a una Pareto en un cuantil distinto al $\approx 39,215\%$.

La causa de que se dé esta restricción en el modelo de Cooray & Ananda es haber multiplicado por la misma constante normalizadora c las dos componentes, tanto la lognormal como la Pareto, en la definición del modelo (2.1).

3.1. Desarrollo teórico del modelo

El anterior inconveniente del modelo (2.1) se puede evitar si a cada componente se le permite tener un peso distinto, en vez de estar multiplicadas las dos por la misma constante:

$$f(x) = \begin{cases} c_1 f_1(x), & 0 < x \leq \theta \\ c_2 f_2(x), & \theta < x < \infty \end{cases}, \quad (3.1)$$

donde c_1 y c_2 son dos constantes positivas en general distintas, y las funciones f_1 y f_2 son las mismas funciones de densidad (2.3) y (2.4) del Capítulo 2.

Para que esta función sea una función de densidad válida es necesario imponer la condición de que la integral de (3.1) sobre todo su dominio valga uno:

$$1 = \int_0^{\infty} f(x) dx = \int_0^{\theta} c_1 f_1(x) dx + \int_{\theta}^{\infty} c_2 f_2(x) dx = c_1 \int_0^{\theta} f_1(x) dx + c_2, \quad (3.2)$$

donde la última igualdad es consecuencia de que

$$\int_{\theta}^{\infty} f_2(x) dx = \int_{\theta}^{\infty} \frac{\alpha \theta^{\alpha}}{x^{\alpha+1}} dx = \alpha \theta^{\alpha} \left[\frac{-1}{\alpha x^{\alpha}} \right]_{x=\theta}^{x=\infty} = 1.$$

Despejando de (3.2), se tiene

$$c_1 = \frac{1 - c_2}{\int_0^{\theta} f_1(x) dx} = \frac{1 - c_2}{\mathbb{P}[X \leq \theta]}, \quad (3.3)$$

donde X es una variable aleatoria que sigue una distribución lognormal de parámetros μ y σ . Por tanto, la primera componente de (3.1) se puede escribir como

$$c_1 f_1(x) = (1 - c_2) \left(\frac{1}{\mathbb{P}[X \leq \theta]} f_1(x) \right).$$

Es decir, el modelo (3.1) consiste en una mezcla de dos funciones de densidad: la de una lognormal *condicionada a que tome valores inferiores a θ* con un peso $1 - c_2$, y la de una Pareto con el peso restante c_2 .

Llamando $\tilde{f}_1(x)$ a la función de densidad de una variable aleatoria lognormal de parámetros μ y σ condicionada al intervalo $(0, \theta)$, donde θ es

un parámetro positivo, esto es, $\tilde{f}_1(x)$ denota a la función de densidad de $X|X < \theta$, de tal suerte que vale

$$\tilde{f}_1(x) = \frac{1}{\mathbb{P}[X \leq \theta]} f_1(x), \quad (3.4)$$

y haciendo el cambio de variable $r = 1 - c_2$, de tal modo que $c_2 = 1 - r$, el modelo (3.1) es la mencionada mixtura de las densidades lognormal condicionada y Pareto:

$$f(x) = \begin{cases} r\tilde{f}_1(x), & 0 < x \leq \theta \\ (1-r)f_2(x), & \theta < x < \infty \end{cases}. \quad (3.5)$$

Esta es la forma en la que aparece parametrizado este modelo en [Scollnik, 2007].

Nótese que puesto que c_1 y c_2 son valores positivos, y $\int_0^\theta f_1(x)dx > 0$, de la ecuación (3.2) se deduce que $0 \leq c_1, c_2 \leq 1$. Por tanto, también el peso r ha de verificar

$$0 \leq r \leq 1. \quad (3.6)$$

3.1.1. Condición de continuidad

Imponiendo la condición de continuidad (2.5) al modelo (3.5) se llega a

$$r\tilde{f}_1(\theta) = (1-r)f_2(\theta),$$

lo cual es equivalente a

$$r \frac{1}{\mathbb{P}[X \leq \theta]} f_1(\theta) = f_2(\theta) - r f_2(\theta). \quad (3.7)$$

Puesto que X es una variable aleatoria que sigue una distribución lognormal de parámetros μ y σ , se tiene

$$\begin{aligned} \mathbb{P}[X \leq \theta] &= \mathbb{P}[\ln(X) \leq \ln(\theta)] = \mathbb{P}\left[\frac{\ln(X) - \mu}{\sigma} \leq \frac{\ln(\theta) - \mu}{\sigma}\right] \\ &= \mathbb{P}\left[Z \leq \frac{\ln(\theta) - \mu}{\sigma}\right] = \Phi\left(\frac{\ln(\theta) - \mu}{\sigma}\right), \end{aligned} \quad (3.8)$$

donde Z denota a una variable aleatoria que sigue una distribución normal estándar, y $\Phi(\cdot)$ es su función de distribución.

Llamando z al cociente $\frac{\ln(\theta)-\mu}{\sigma}$, esto último permite escribir la condición (3.7) como

$$r \left[\frac{1}{\Phi(z)} f_1(\theta) + f_2(\theta) \right] = f_2(\theta).$$

Despejando,

$$r = \frac{f_2(\theta)}{\frac{1}{\Phi(z)} f_1(\theta) + f_2(\theta)} = \frac{\Phi(z) f_2(\theta)}{f_1(\theta) + \Phi(z) f_2(\theta)} = \frac{\Phi(z) \frac{f_2(\theta)}{f_1(\theta)}}{\Phi(z) \frac{f_2(\theta)}{f_1(\theta)} + 1}. \quad (3.9)$$

Utilizando las definiciones (2.3) y (2.4), se tiene que

$$\frac{f_2(\theta)}{f_1(\theta)} = \frac{\frac{\alpha \theta^\alpha}{\theta^{\alpha+1}}}{\frac{1}{\sqrt{2\pi}\sigma\theta} \exp\left\{\frac{-z^2}{2}\right\}} = \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{z^2}{2}\right\}.$$

Utilizando esto en (3.9), se obtiene la restricción necesaria para dotar al modelo (3.5) de continuidad, la cual expresa el parámetro de peso r en función de los otros parámetros:

$$\begin{aligned} r &= \frac{\Phi(z) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{z^2}{2}\right\}}{\Phi(z) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{z^2}{2}\right\} + 1} \\ &= \frac{\Phi\left(\frac{\ln(\theta)-\mu}{\sigma}\right) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{1}{2}\left(\frac{\ln(\theta)-\mu}{\sigma}\right)^2\right\}}{\Phi\left(\frac{\ln(\theta)-\mu}{\sigma}\right) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{1}{2}\left(\frac{\ln(\theta)-\mu}{\sigma}\right)^2\right\} + 1}. \end{aligned} \quad (3.10)$$

Con esta restricción, el número de parámetros del modelo de Scollnik se reduce de los inicialmente cinco (el parámetro de peso r y dos parámetros por cada una de las dos componentes f_1 y f_2) a cuatro.

3.1.2. Estudio del parámetro de peso r

Llamando $A(\alpha)$ a la expresión

$$\Phi\left(\frac{\ln(\theta)-\mu}{\sigma}\right) \sqrt{2\pi} \cdot \alpha \cdot \sigma \exp\left\{\frac{1}{2}\left(\frac{\ln(\theta)-\mu}{\sigma}\right)^2\right\} \quad (3.11)$$

que aparece tanto en el numerador, como en el denominador de (3.10), para unos valores fijos del parámetro de corte $\theta > 0$ y de los parámetros de la lognormal $\mu \in \mathbb{R}$ y $\sigma > 0$, se tiene que

- $A(\alpha)$ es una función lineal en α ;
- $A(\alpha)$ es una función creciente en α ;
- $A(\alpha) > 0$, puesto que tanto α como el resto de los factores que aparecen en (3.11) son estrictamente mayores que 0: α y σ por ser parámetros de distribuciones estrictamente positivos y $\Phi(\cdot)$ y $\exp\{\cdot\}$ por ser funciones cuya imagen está contenida en la semirrecta positiva $(0, \infty)$;
- $\lim_{\alpha \rightarrow 0} A(\alpha) = 0$, en vista de que el resto de factores del producto de (3.11) son constantes una vez fijados los valores de los parámetros θ , μ y σ ;
- $\lim_{\alpha \rightarrow \infty} A(\alpha) = \infty$, consecuencia de la misma razón del punto anterior.

Puesto que

$$r = \frac{A(\alpha)}{A(\alpha) + 1},$$

como consecuencia de las propiedades enunciadas sobre $A(\alpha)$, se sigue que $0 \leq r \leq 1$, en concordancia con (3.6).

De hecho, viendo ahora el parámetro de peso r escrito como función del parámetro α ,

$$r(\alpha) = \frac{A(\alpha)}{A(\alpha) + 1},$$

y usando las dos últimas propiedades de la función $A(\alpha)$ sobre su comportamiento cuando α se aproxima a los extremos de su dominio de definición $(0, \infty)$, se puede ver que

$$\lim_{\alpha \rightarrow 0} r(\alpha) = \lim_{\alpha \rightarrow 0} \frac{A(\alpha)}{A(\alpha) + 1} = 0$$

y

$$\lim_{\alpha \rightarrow \infty} r(\alpha) = \lim_{\alpha \rightarrow \infty} \frac{A(\alpha)}{A(\alpha) + 1} = 1.$$

Esto es, el parámetro de peso r varía ahora en todo el intervalo $(0, 1)$. Nótese que los extremos del intervalo son abiertos: para ningún valor válido de α se tiene que $r = 0$ o $r = 1$. De todos modos, esto no es un inconveniente del modelo, puesto que los casos donde $r \in \{0, 1\}$ se corresponden con los modelos degenerados en los que toda la masa probabilística se concentra en la componente lognormal o Pareto.

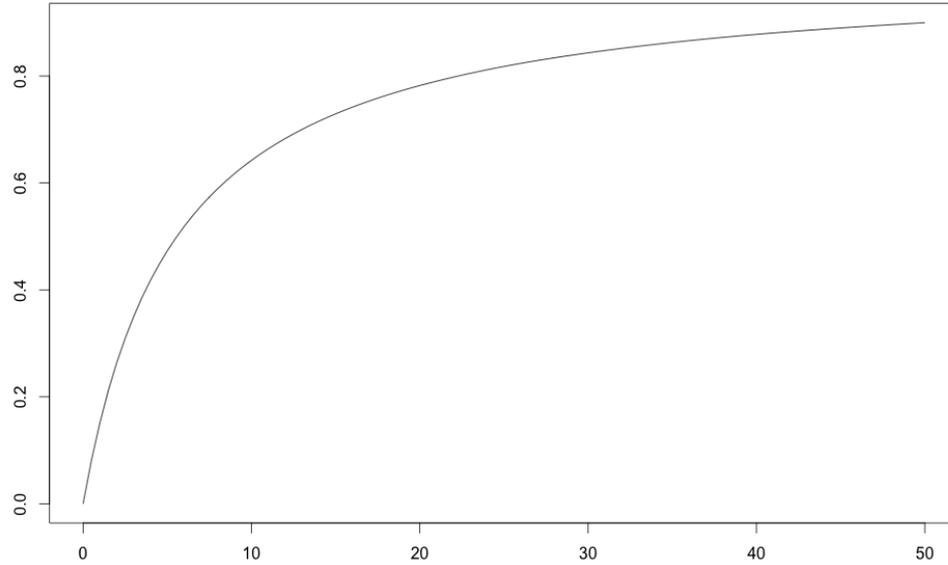


Figura 3.1: Valores del parámetro de peso r como función del parámetro de forma de la componente Pareto α

Escogido cualquier valor de r_0 en el intervalo $(0, 1)$ existirá un valor α_0 del parámetro α tal que se alcance el deseado valor para r , esto es, tal que se verifique

$$r_0 = \frac{A(\alpha_0)}{A(\alpha_0) + 1}.$$

Esto permite asociar un peso positivo¹ arbitrario a cualquiera de las dos componentes, lognormal y Pareto, del modelo (3.5).

En la Figura 3.1 se ha representado el valor del parámetro de peso r en función del valor del parámetro de forma de la componente Pareto del modelo α . Este último varía entre $\alpha = 0$ y $\alpha = 50$, para unos valores fijados $\theta = 100$, $\mu = 10$ y $\sigma = 1$.

¹La positividad del peso no es una restricción: una probabilidad negativa, derivada de un peso menor que cero, es difícilmente interpretable, mientras que un peso nulo resulta en un modelo degenerado, en vez de una mixtura de dos densidades diferentes, como ya se ha hecho notar.

Nótese como r crece con α desde valores casi nulos (cuando α tiende a 0) hasta superar el valor 0,8 (cuando α se aproxima 50). El valor de r se puede aproximar arbitrariamente cerca a 1, aunque esto requiere tomar valores de α demasiado elevados. En cualquier caso, conforme aumenta el valor del parámetro de peso r , más protagonismo toma la componente lognormal dentro del modelo (3.5), y a partir de un cierto valor de este parámetro es preferible prescindir de la complejidad del modelo de Scollnik en favor de un modelo basado simplemente en la distribución lognormal.

Obsérvese que la probabilidad acumulada por el modelo (3.5) hasta $x = \theta$ es

$$\begin{aligned} \int_0^\theta f(x)dx &= \int_0^\theta r \tilde{f}_1(x)dx = r \int_0^\theta \frac{1}{\mathbb{P}[X \leq \theta]} f_1(x)dx \\ &= \frac{r}{\mathbb{P}[X \leq \theta]} \mathbb{P}[X \leq \theta] = r, \end{aligned} \quad (3.12)$$

donde como antes X es una variable aleatoria que sigue una distribución lognormal de parámetros μ y σ .

Es decir, en este modelo, la probabilidad acumulada hasta $x = \theta$ es arbitraria dentro del intervalo $(0, 1)$, y ya no es necesariamente

$$\frac{\Phi(k)}{1 + \Phi(k)} \approx 0,39214992251571$$

como en el modelo de Cooray & Ananda.

3.1.3. Condición de diferenciabilidad

De forma similar a (2.6), imponiendo al modelo (3.5) la condición de diferenciabilidad se llega a

$$r \tilde{f}'_1(\theta) = (1 - r) f'_2(\theta)$$

lo cual es equivalente a

$$\frac{r}{\Phi(z)} f'_1(\theta) = (1 - r) f'_2(\theta), \quad (3.13)$$

donde $z = \frac{\ln(\theta) - \mu}{\sigma}$ igual que antes.

Utilizando las derivadas de las funciones f_1 y f_2 , (2.10) y (2.11), ya calculadas en Capítulo 2, la condición (3.13) se escribe como

$$\frac{r}{\Phi(z)} \left[-\frac{\exp\left\{-\frac{z^2}{2}\right\} [\ln(\theta) - \mu + \sigma^2]}{\sqrt{2\pi}\sigma^3\theta^2} \right] = (1-r) \left[-\frac{(\alpha+1)\alpha\theta^\alpha}{\theta^{\alpha+2}} \right],$$

de donde simplificando y reordenando se obtiene

$$\frac{r}{(1-r)\Phi(z)} = \frac{\sqrt{2\pi}\sigma^3(\alpha+1)\alpha \exp\left\{\frac{z^2}{2}\right\}}{\ln(\theta) - \mu + \sigma^2}. \quad (3.14)$$

Si $r = \frac{A(\alpha)}{A(\alpha)+1}$, entonces $1-r = \frac{1}{A(\alpha)+1}$, y por tanto,

$$\frac{r}{(1-r)\Phi(z)} = \frac{\frac{A(\alpha)}{A(\alpha)+1}}{\frac{1}{A(\alpha)+1}\Phi(z)} = \frac{A(\alpha)}{\Phi(z)} = \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{z^2}{2}\right\} \quad (3.15)$$

donde la última igualdad es consecuencia de (3.11).

Esto permite reescribir (3.14) como

$$\sqrt{2\pi}\alpha\sigma \exp\left\{\frac{z^2}{2}\right\} = \frac{\sqrt{2\pi}\sigma^3(\alpha+1)\alpha \exp\left\{\frac{z^2}{2}\right\}}{\ln(\theta) - \mu + \sigma^2},$$

donde tras simplificar queda

$$1 = \frac{\sigma^2(\alpha+1)}{\ln(\theta) - \mu + \sigma^2},$$

o lo que es lo mismo,

$$\ln(\theta) - \mu + \sigma^2 = \sigma^2(\alpha+1) = \sigma^2\alpha + \sigma^2,$$

donde tras simplificar el término σ^2 , queda

$$\ln(\theta) - \mu = \alpha\sigma^2. \quad (3.16)$$

Esto se puede escribir como

$$\frac{\ln(\theta) - \mu}{\sigma} = \alpha\sigma, \quad (3.17)$$

lo cual es idéntico a la condición (2.12) obtenida en el Capítulo 2 para el modelo de Cooray & Ananda.

Es curioso que para el modelo de Scollnik se obtenga la misma restricción necesaria para obtener suavidad en el punto de corte θ que para el modelo de Cooray & Ananda, a pesar de que este modelo es más general que el estudiado en el Capítulo 2.

Nótese que no era evidente que se fuera a obtener la misma condición, en vista de la presencia del parámetro de peso adicional r en el modelo (3.5), el cual no aparecía en (2.16).

Utilizando (3.17) en (3.10), la condición sobre el parámetro de peso r para la continuidad se simplifica a

$$r = \frac{\Phi(\alpha\sigma) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{1}{2}(\alpha\sigma)^2\right\}}{\Phi(\alpha\sigma) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{1}{2}(\alpha\sigma)^2\right\} + 1}. \quad (3.18)$$

Obsérvese que a pesar de que ahora α aparezca dentro de las funciones $\Phi(\cdot)$ y $\exp(\cdot)$, la identidad (3.17) permite deshacer el cambio simplificador que llevó a la expresión (3.18) para volver a repetir el razonamiento con la función $A(\alpha)$ anterior y garantizar las propiedades relativas a que el parámetro de peso r tome cualquier valor en el intervalo $(0, \infty)$ deducidas para el modelo (3.5) con la condición de continuidad, pero no la de diferenciabilidad, como la posibilidad de elegir la probabilidad relativa a cada una de las dos componentes, lognormal y Pareto, del modelo de Scollnik.

Impuestas al modelo de Scollnik tanto la condición de continuidad, como la de diferenciabilidad, el número de parámetros se reduce aún más hasta un total de tres.

3.1.4. Restricción al modelo de Cooray & Ananda

Si se toman los valores de los parámetros α y σ tales que $\alpha\sigma = k$, donde k es la solución de la ecuación (2.13) del Capítulo 2

$$\exp\left\{-\frac{1}{2}k^2\right\} = \sqrt{2\pi}k,$$

el modelo (3.5) de Scollnik se reduce al modelo (2.16) de Cooray & Ananda:

- El valor de r se simplifica a

$$r = \frac{\Phi(k) \sqrt{2\pi}k \exp\left\{\frac{k^2}{2}\right\}}{\Phi(k) \sqrt{2\pi}k \exp\left\{\frac{k^2}{2}\right\} + 1} = \frac{\Phi(k)}{1 + \Phi(k)},$$

donde la última igualdad es consecuencia de la propia definición de k como la solución de la ecuación (2.13).

Usando el resultado de la probabilidad acumulada por el modelo de Scollnik hasta $x = \theta$ calculado en (3.12), se obtiene que dicha probabilidad es

$$\frac{\Phi(k)}{1 + \Phi(k)} \approx 0,39214992251571,$$

igual que en el modelo de Cooray & Ananda, tal y como y se vio en (2.20).

- El peso por el que va multiplicada la primera componente \tilde{f}_1 del modelo (3.5) es $r = \frac{\Phi(k)}{1 + \Phi(k)}$, y en vista de la definición (3.4) de \tilde{f}_1 , se tiene que el peso de la componente f_1 en el modelo de Scollnik es

$$c_1 = r \frac{1}{\mathbb{P}[X \leq \theta]} = \frac{\Phi(k)}{1 + \Phi(k)} \frac{1}{\mathbb{P}[X \leq \theta]} \stackrel{*}{=} \frac{\Phi(k)}{1 + \Phi(k)} \frac{1}{\Phi(k)} = \frac{1}{1 + \Phi(k)},$$

donde como antes X es una variable aleatoria que sigue una distribución lognormal de parámetros μ y σ , y por tanto, para deducir la igualdad * basta hacer

$$\begin{aligned} \mathbb{P}[X \leq \theta] &= \mathbb{P}[\ln(X) \leq \ln(\theta)] = \mathbb{P}\left[\frac{\ln(X) - \mu}{\sigma} \leq \frac{\ln(\theta) - \mu}{\sigma}\right] \\ &= \mathbb{P}\left[Z \leq \frac{\ln(\theta) - \mu}{\sigma}\right] = \Phi\left(\frac{\ln(\theta) - \mu}{\sigma}\right) \\ &= \Phi(\alpha\sigma) = \Phi(k), \end{aligned} \quad (3.19)$$

donde Z representa una variable aleatoria normal estándar.

Así, el peso c_1 del modelo de Scollnik coincide con el peso $c = \frac{1}{1 + \Phi(k)}$ del modelo de Cooray & Ananda, calculado en el Capítulo 2 cuando $\alpha\sigma = k$.

- La diferencia entre los modelos de Cooray & Ananda y Scollnik es que el primero aplica el mismo peso c a las dos componentes lognormal y Pareto, y el segundo permite a ambas componentes tomar pesos c_1 y c_2 en general distintos.

Como se ha visto en (3.19), si $\alpha\sigma = k$ resulta que $\mathbb{P}[X \leq \theta] = \Phi(k)$, por tanto, la relación (3.3) puede ser escrita como

$$c_1 = \frac{1 - c_2}{\Phi(k)},$$

y si se hace $c_1 = c_2$, y se llama d a estos pesos idénticos de tal suerte que $d = c_1 = c_2$, se tiene

$$d = \frac{1 - d}{\Phi(k)},$$

de donde despejando d se llega a

$$d = \frac{1}{1 + \Phi(k)},$$

que es exactamente el mismo resultado obtenido en (2.15) con respecto al peso del modelo de Cooray & Ananda.

3.1.5. Condición de diferenciabilidad de segundo orden

Resulta tentador intentar reducir aún más el número de parámetros hasta llegar a dos grados de libertad, imponiendo una condición de diferenciabilidad de segundo orden de tal suerte que se verifique

$$\lim_{x \rightarrow \theta^-} f''(x) = \lim_{x \rightarrow \theta^+} f''(x). \quad (3.20)$$

En vista de la definición (3.5) del modelo de Scollnik y de la continuidad de sus dos componentes se tiene que

$$\begin{aligned} \lim_{x \rightarrow \theta^-} f''(x) &= \lim_{x \rightarrow \theta^-} \frac{d^2}{dx^2} r \tilde{f}_1(x) = \lim_{x \rightarrow \theta^-} \frac{d^2}{dx^2} \frac{r}{\mathbb{P}[X \leq \theta]} f_1(x) \\ &= \lim_{x \rightarrow \theta^-} \frac{d^2}{dx^2} \frac{r}{\Phi(z)} f_1(x) = \lim_{x \rightarrow \theta^-} \frac{r}{\Phi(z)} \frac{d^2}{dx^2} f_1(x) = \frac{r}{\Phi(z)} f_1''(\theta), \end{aligned}$$

donde $z = \frac{\ln(\theta) - \mu}{\sigma}$, y la primera igualdad de la segunda línea es consecuencia de que

$$\begin{aligned} \mathbb{P}[X \leq \theta] &= \mathbb{P}[\ln(X) \leq \ln(\theta)] = \mathbb{P}\left[\frac{\ln(X) - \mu}{\sigma} \leq \frac{\ln(\theta) - \mu}{\sigma}\right] \\ &= \mathbb{P}\left[Z \leq \frac{\ln(\theta) - \mu}{\sigma}\right] = \Phi\left(\frac{\ln(\theta) - \mu}{\sigma}\right) = \Phi(z). \end{aligned}$$

Análogamente, volviendo a usar la definición (3.5) y aplicando la continuidad de las componentes en la última igualdad, se obtiene que

$$\lim_{x \rightarrow \theta^+} f''(x) = \lim_{x \rightarrow \theta^+} \frac{d^2}{dx^2} (1 - r) f_2(x) = \lim_{x \rightarrow \theta^+} (1 - r) \frac{d^2}{dx^2} f_2(x) = (1 - r) f_2''(\theta).$$

Con esto, (3.20) se escribe como

$$\frac{r}{\Phi(z)} f_1''(\theta) = (1-r) f_2''(\theta),$$

donde reordenando se obtiene

$$\frac{r}{(1-r)\Phi(z)} f_1''(\theta) = f_2''(\theta),$$

y usando (3.15),

$$\sqrt{2\pi}\alpha\sigma \exp\left\{\frac{z^2}{2}\right\} f_1''(\theta) = f_2''(\theta). \quad (3.21)$$

Ahora, derivando f_1' que aparece en (2.10) y simplificando, se obtiene

$$f_1''(x) = \frac{\exp\left\{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2\right\}}{\sqrt{2\pi}\sigma^3 x^3} \left[\left(\frac{\ln(x)-\mu}{\sigma^2} + 2 \right) (\ln(x) - \mu + \sigma^2) - 1 \right].$$

Por otro lado, derivando f_2' que aparece en (2.11), se llega a

$$f_2''(x) = \frac{(\alpha+2)(\alpha+1)\alpha\theta^\alpha}{x^{\alpha+3}}.$$

Evaluando estas derivadas en $x = \theta$ y simplificando, la condición (3.21) se transforma en

$$\frac{1}{\sigma^2} \left[\left(\frac{\ln(\theta)-\mu}{\sigma^2} + 2 \right) (\ln(\theta) - \mu + \sigma^2) - 1 \right] = (\alpha+2)(\alpha+1).$$

Desarrollando,

$$\left[\left(\frac{\ln(\theta)-\mu}{\sigma^2} + 2 \right) \left(\frac{\ln(\theta)-\mu}{\sigma^2} + 1 \right) - \frac{1}{\sigma^2} \right] = (\alpha+2)(\alpha+1),$$

y usando (3.17),

$$(\alpha+2)(\alpha+1) - \frac{1}{\sigma^2} = (\alpha+2)(\alpha+1),$$

de donde simplificando se obtiene la condición necesaria para que se cumpla la igualdad de las segundas derivadas en el punto $x = \theta$ establecida en (3.20):

$$-\frac{1}{\sigma^2} = 0.$$

Obviamente, esta condición es absurda, y por tanto, se deduce que es imposible obtener condiciones de diferenciabilidad de segundo orden en el modelo (3.5).

3.2. Variación de la densidad con los parámetros

El comportamiento de las funciones de densidad es más complejo en el modelo de Scollnik que en el de Cooray & Ananda.

3.2.1. Cambios en la densidad con el parámetro de forma α

Recuérdese que en el caso del modelo del Capítulo 2, a medida que aumentaba el parámetro de forma de la componente Pareto α , mientras se mantenía constante el parámetro de corte θ , la moda del modelo compuesto se desplazaba a la derecha, el valor de la función de densidad evaluada en esta moda también crecía con α y la función de densidad compuesta se hacía más leptocúrtica, volviéndose más aguda en torno a esta moda. Este comportamiento se mostró en el Capítulo 2 en la Figura 2.2.

Usando el mismo rango de valores para el parámetro de forma de la componente Pareto, concretamente, $\alpha = \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}$ y 3 y el mismo valor para el parámetro de corte, concretamente, $\theta = 100$ que para generar la Figura 2.2 en el modelo de Scollnik (3.5), calculando el resto de parámetros como

$$\sigma = \frac{k}{\alpha} \quad (3.22)$$

$$r = \frac{\Phi(\alpha\sigma) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{1}{2}(\alpha\sigma)^2\right\}}{\Phi(\alpha\sigma) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{1}{2}(\alpha\sigma)^2\right\} + 1} \quad (3.23)$$

$$\mu = \ln(\theta) - \alpha\sigma^2 \quad (3.24)$$

donde k es, como antes, la solución a la ecuación (2.13), se obtendría exactamente la misma figura que la Figura 2.2, pues como ya se ha mostrado, el modelo de Scollnik con la restricción $\alpha\sigma = k$, usada para el cálculo del parámetro σ en (3.22), se reduce al modelo de Cooray & Ananda.

Ahora, mientras que las condiciones (3.23) y (3.24) son necesarias para dotar al modelo de Scollnik de continuidad y diferenciabilidad en $x = \theta$, la condición (3.22) es totalmente prescindible, pudiendo el parámetro de escala de la componente lognormal σ tomar cualquier valor mayor que cero.

Así, se puede calcular este parámetro como

$$\sigma = 0,5 \frac{k}{\alpha}, \quad (3.25)$$

obteniendo una menor dispersión en la componente lognormal del modelo de Scollnik que la que se obtendría en el caso del modelo de Cooray & Ananda, en vista de que el parámetro de escala se redujo en un 50 %.

Nótese que el expresar σ como función del cociente k/α es tan sólo a propósito de comparar más adelante las funciones de densidad del modelo de Scollnik con σ calculadas como en (3.25) con las funciones de densidad del modelo de Cooray & Ananda: recuérdese que en el modelo de Scollnik al parámetro σ se le permite tomar cualquier valor arbitrario dentro del intervalo $(0, \infty)$.

En la Figura 3.2 se representan funciones de densidad del modelo de Scollnik para el parámetro de forma de la componente Pareto $\alpha = \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}$ y 3, y el parámetro de corte $\theta = 100$, calculándose el parámetro σ de dos formas distintas.

En la Figura 3.2a este parámetro de escala de la componente lognormal se ha calculado como $\sigma = \frac{k}{\alpha}$, obteniéndose, tal y como ya se adelantaba, las mismas funciones de densidad que en la Figura 2.2.

En la Figura 3.2b el parámetro σ se ha calculado como $0,5 \frac{k}{\alpha}$, buscando infradispersión de la componente lognormal en comparación con el modelo de Cooray & Ananda. Obsérvese que en este caso la probabilidad de la moda aumenta y las funciones de densidad se hacen más leptocúrticas con respecto al caso de $\sigma = \frac{k}{\alpha}$.

De hecho, si se observan las gráficas más detenidamente se aprecia como parece que en la Figura 3.2b la componente Pareto tiene más peso que en la Figura 3.2a. Usando la expresión (3.18)

$$r = \frac{\Phi(\alpha\sigma) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{1}{2}(\alpha\sigma)^2\right\}}{\Phi(\alpha\sigma) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{1}{2}(\alpha\sigma)^2\right\} + 1}$$

se puede calcular el peso r asociado a las componentes lognormales² de las Figuras 3.2a y 3.2b: en el primer caso, cuando $\sigma = \frac{k}{\alpha}$ se tiene que $r \approx 0,3921499$

² Nótese que puesto que dentro de cada una de las subfiguras de la Figura 3.2 el producto $\alpha\sigma$ es constante, también lo es r , calculado según (3.18). Por tanto, el peso de *todas* las componentes lognormales dentro de cada subfigura es el mismo.

y en el segundo, cuando $\sigma = 0,5\frac{k}{\alpha}$, el valor de r es $\approx 0,2140742$. Es decir, tal y como se apreció visualmente, al parecer las componentes lognormales menos dispersas hacen que el peso de la componente lognormal se reduzca dentro del modelo global.

Recíprocamente, la Figura 3.3 es análoga a la Figura 3.2, donde el único cambio es que en la Figura 3.3b σ se calcula como $\sigma = 1,25\frac{k}{\alpha}$, en vez de $\sigma = 0,5\frac{k}{\alpha}$, buscando sobredispersión esta vez. La Figura 3.3a es la misma que la Figura 3.2a, la cual a su vez es la misma que la Figura 2.2, y se vuelve a mostrar para facilitar el estudio de cómo afecta la variación del parámetro de escala de la componente lognormal σ .

En la Figura 3.3 la componente lognormal gana peso dentro del modelo a medida que se aumenta su parámetro de escala: el valor del parámetro r es $\approx 0,4688374$ cuando $\sigma = 1,25\frac{k}{\alpha}$, frente al $r \approx 0,3921499$ del caso $\sigma = \frac{k}{\alpha}$.

En vista de las Figuras 3.2 y 3.3 se podría pensar que el comportamiento de las funciones de densidad deducido de la Figura 2.2 para el modelo de Cooray & Ananda se vuelve a dar en el modelo de Scollnik, y a medida que aumenta el parámetro α la moda se desplaza a la derecha, aumentando el valor de la función de densidad evaluada en esta moda y la función de densidad se vuelve más leptocúrtica.

Sin embargo, la Figura 3.4 muestra que esto no es así. Si σ se calcula como $2\frac{k}{\alpha}$, el comportamiento visto hasta ahora cambia: aunque la moda sigue desplazándose a la derecha a medida que aumenta α , el valor de la función de densidad más alto corresponde al valor de α más bajo considerado. Igualmente, a este valor del parámetro de forma de la componente α más bajo le corresponde la función de densidad más leptocúrtica.

La característica que sigue comportándose igual que en los casos anteriores es la variación del parámetro de peso r . Para $\sigma = 2\frac{k}{\alpha}$ el peso de las componentes lognormales es $\approx 0,655169$, el más alto de los calculados hasta ahora.

3.2.2. Cambios en la densidad con el parámetro de corte θ

Análogamente, se ha estudiado también como afecta la variación del parámetro θ a las funciones de densidad del modelo de Scollnik, para distintos valores de σ , al igual que se ha hecho con el parámetro α anteriormente.

Igual que en caso del modelo de Cooray & Ananda, se ha fijado el pará-

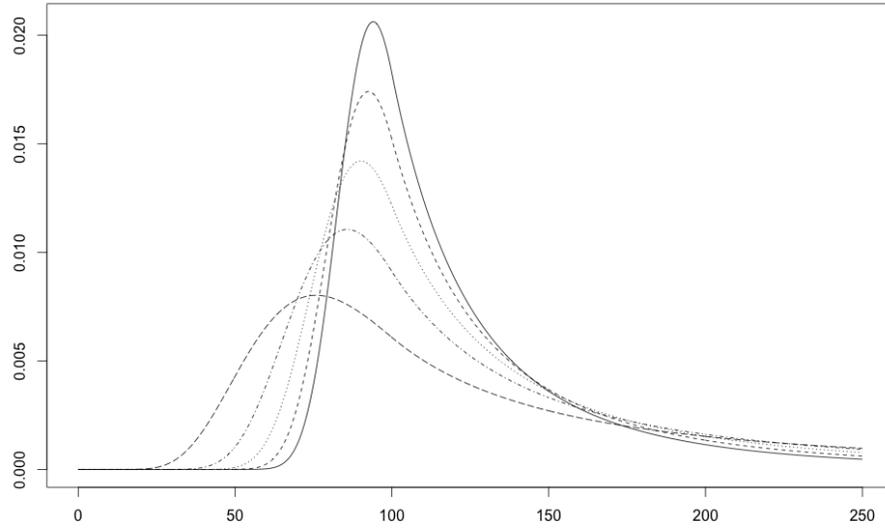
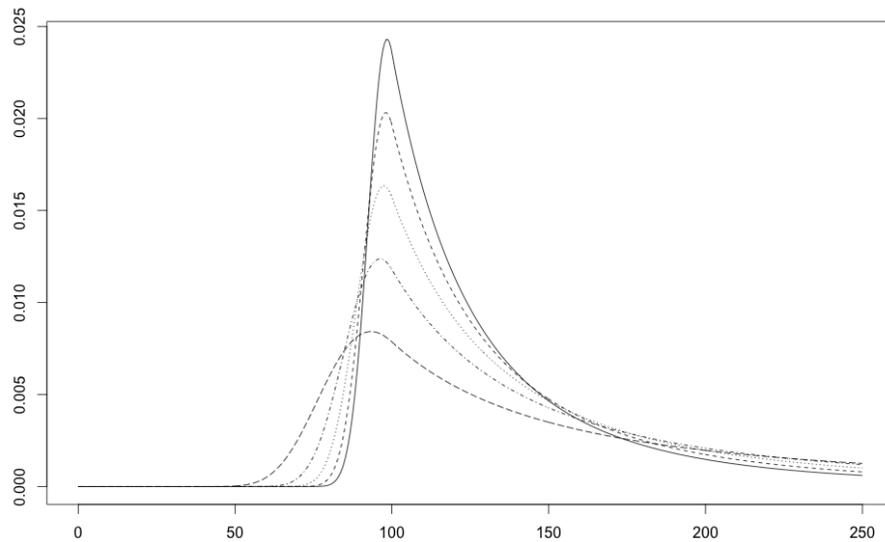
(a) $\sigma = \frac{k}{\alpha}$ (b) $\sigma = 0,5 \frac{k}{\alpha}$

Figura 3.2: Densidades del modelo de Scollnik para dos formas de calcular σ , incluyendo un σ *infradisperso* con respecto a k/α , y para un rango de valores de α

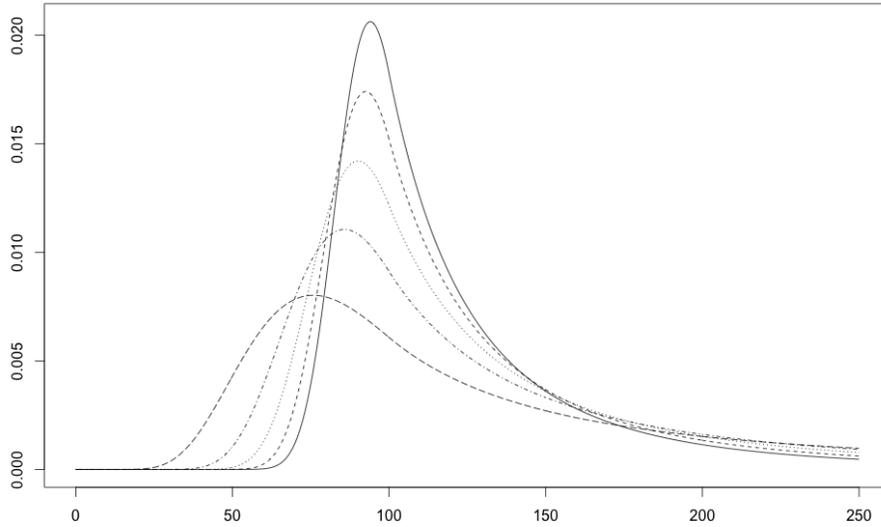
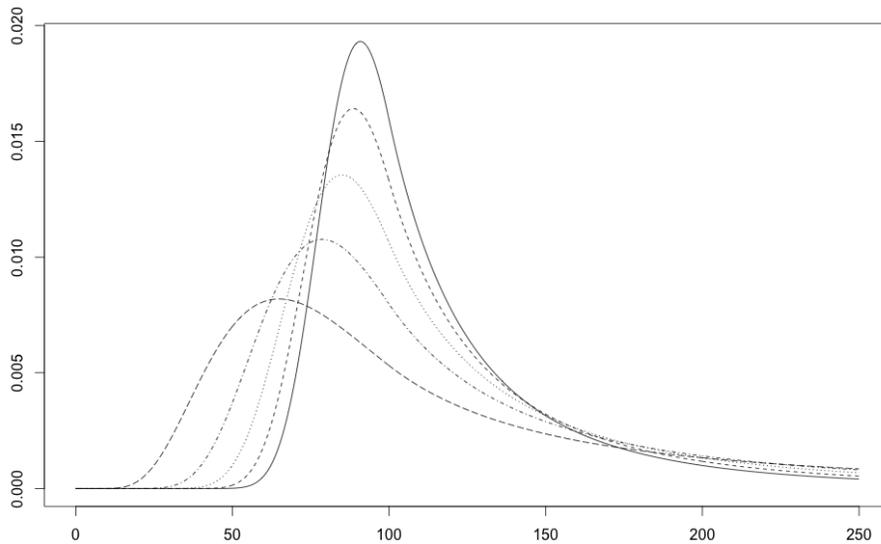
(a) $\sigma = \frac{k}{\alpha}$ (b) $\sigma = 1,25 \frac{k}{\alpha}$

Figura 3.3: Densidades del modelo de Scollnik para dos formas de calcular σ , incluyendo un σ *sobredisperso* con respecto a k/α , y para un rango de valores de α

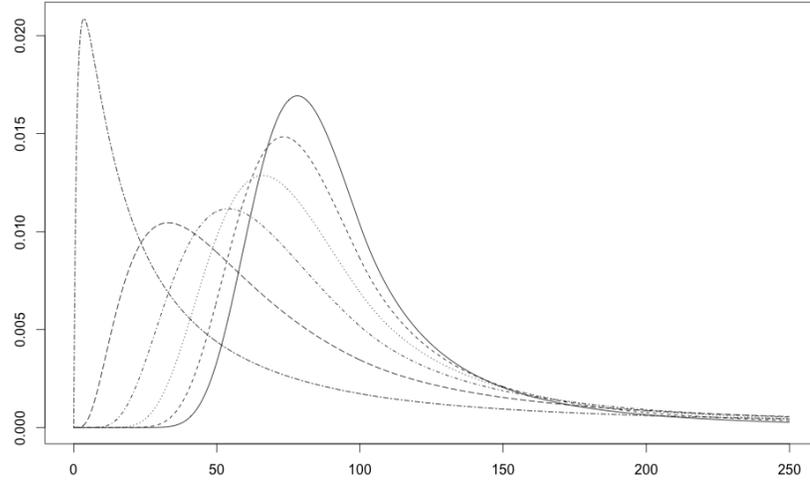


Figura 3.4: Densidades del modelo de Scollnik con $\sigma = 2\frac{k}{\alpha}$ para un rango de valores de α

metro de escala de la componente Pareto en $\alpha = 1$, y el parámetro de corte tomó los valores: $\theta = 50, 100, 150, 200, 250$ y 300 .

En la Figura 3.5a se muestran las densidades del modelo (3.5) para este rango de valores de θ con el parámetro de escala de la componente lognormal calculado como $\sigma = 0,15\frac{k}{\alpha}$, esto es, buscando infradispersión en comparación con el modelo Cooray & Ananda. En la Figura 3.5b se muestran las densidades del caso sobredisperso: calculando σ como $2\frac{k}{\alpha}$. Como se puede apreciar, en ambos casos se obtiene el mismo comportamiento que en la Figura 2.3 correspondiente al caso particular de $\sigma = \frac{k}{\alpha}$ del modelo de Cooray & Ananda: a medida que aumenta θ la moda se desplaza a la derecha, el valor de la función de densidad evaluada en esta moda decrece a medida que crece el parámetro de corte θ y la función de densidad se vuelva cada vez más platocúrtica. Incluso con valores de σ tan bajos como $0,15\frac{k}{\alpha}$ y tan altos como $2\frac{k}{\alpha}$ el modelo de Scollnik presenta el mismo comportamiento que el modelo de Cooray & Ananda.

Puesto que se obtiene exactamente el mismo resultado con respecto a la variación del parámetro de θ para ambos modelos para todos los valores de σ considerados, no se adjuntan las densidades del modelo (3.5) donde σ es

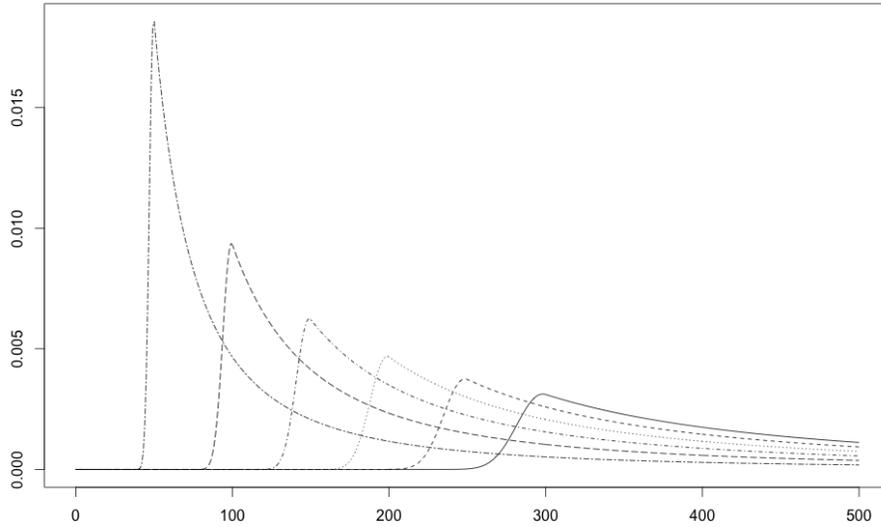
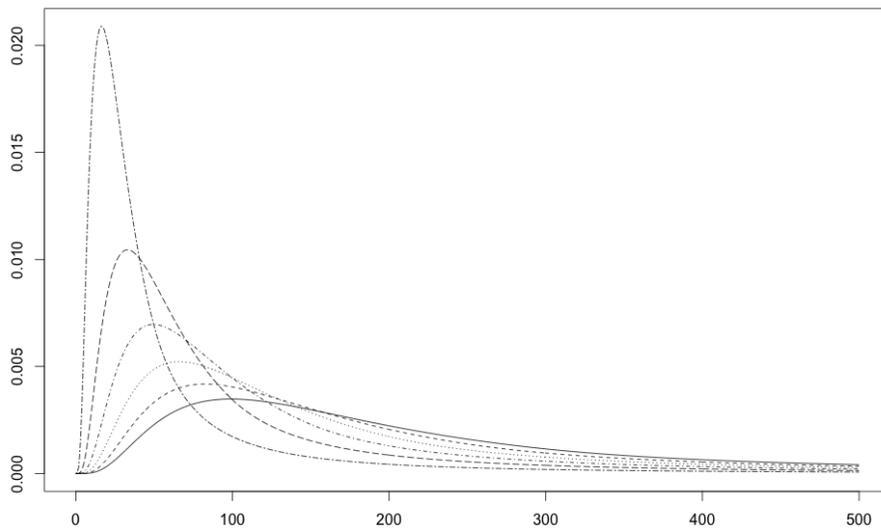
(a) $\sigma = 0,15 \frac{k}{\alpha}$ (b) $\sigma = 2 \frac{k}{\alpha}$

Figura 3.5: Densidades del modelo de Scollnik para dos formas de calcular σ , una que proporciona un σ *infradiserso* y otra que proporciona un σ *sobredisperso*, con respecto a k/α , y para un rango de valores de θ

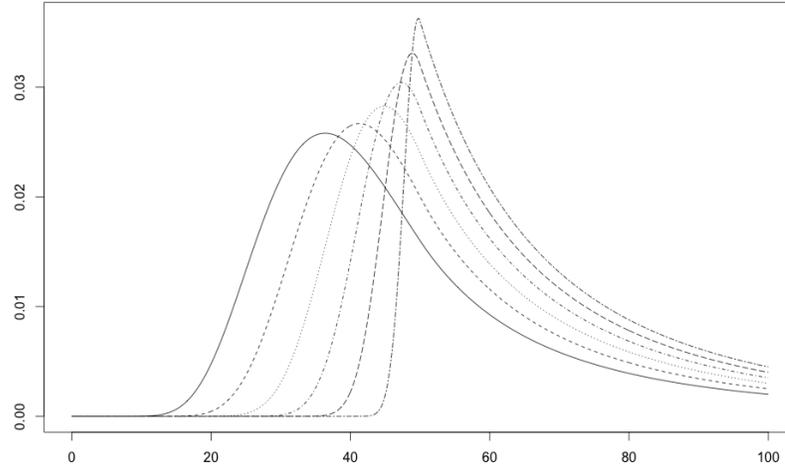


Figura 3.6: Densidades del modelo de Scollnik para un rango de valores de r con $\alpha = 2$

calculado como $\frac{k}{\alpha}$, pero se hace notar que se obtuvo exactamente la misma gráfica que la mostrada en Figura 2.3.

Con respecto a los pesos de las componentes lognormales r de los modelos de la Figura 3.5, iguales dentro de cada caso, según lo explicado en el pie de página 2 de la página 52, éste fue de $\approx 0,06821582$ para el caso infradiserso donde $\sigma = 0,15\frac{k}{\alpha}$ y $\approx 0,655169$ para el caso sobredisperso donde $\sigma = 2\frac{k}{\alpha}$.

Nótese que este peso tan bajo del primer caso está en consonancia con la Figura 3.5a, en la cual las funciones de densidad recuerdan mucho a las funciones de densidad de las distribuciones de Pareto puras.

El peso obtenido para las densidades de la Figura 3.5b es el mismo que para las densidades de Figura 3.4, lo cual es lógico, de nuevo en vista de lo comentado en el pie de página 2 de la página 52: ambas figuras presentan densidades donde se verifica $\sigma = 2\frac{k}{\alpha}$, o lo que es lo mismo, $\alpha\sigma = 2k$, y como se ha visto en (3.18), el peso de las componentes del modelo (3.5) sólo depende del producto $\alpha\sigma$, y no de los valores concretos de cada uno de los parámetros.

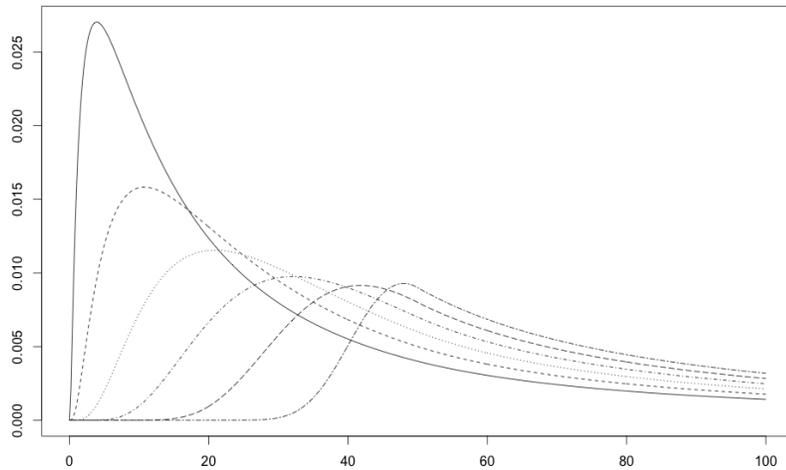


Figura 3.7: Densidades del modelo de Scollnik para un rango de valores de r con $\alpha = 0,5$

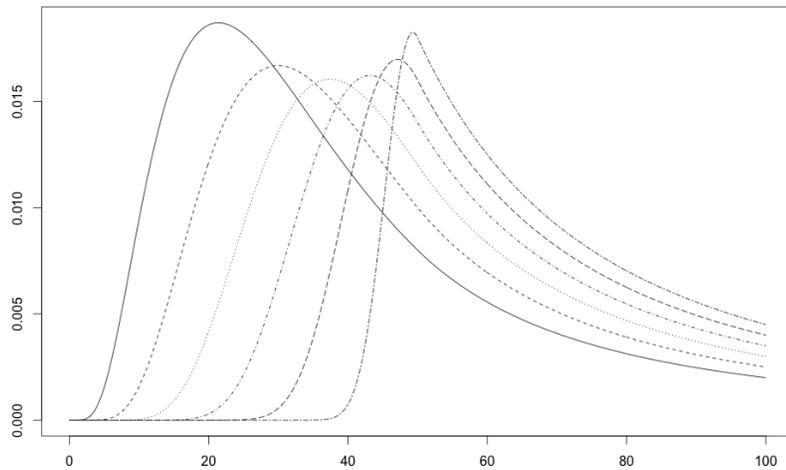


Figura 3.8: Densidades del modelo de Scollnik para un rango de valores de r con $\alpha = 1$

3.2.3. Cambios en la densidad con el parámetro de peso r

El comportamiento de las funciones de densidad de (3.5) en función de r tampoco es sencillo.

Tomando constantes los valores del parámetro de forma de la componente Pareto $\alpha = 2$ y del parámetro de corte $\theta = 50$, y calculando las densidades para un rango del parámetro de peso de la componente lognormal $r = 0,1; 0,2; 0,3; 0,4; 0,5$ y $0,6$ se obtiene la Figura 3.6. Se aprecia que a medida que aumenta r disminuye el valor de la moda, el valor de la función de densidad evaluada en esta moda también decrece y la función de densidad en sí se vuelve más platocúrtica.

Sin embargo, con el mismo valor del parámetro de corte $\theta = 50$, y para los mismos valores del parámetro de peso de la componente lognormal $r = 0,1; 0,2; 0,3; 0,4; 0,5$ y $0,6$, pero cambiando el valor del parámetro de forma de la componente Pareto a $\alpha = 0,5$ se obtienen las densidades representadas en la Figura 3.7: conforme aumenta r la moda vuelve a disminuir, pero esta vez la función de densidad evaluada en la moda aumenta con r , al contrario de lo que sucedía en la Figura 3.6 con $\alpha = 2$. Del mismo modo, cambia el carácter de la curtosis: ahora a medida que aumenta r la función de densidad se vuelve cada vez más leptocúrtica, al contrario de lo que ocurría en la Figura 3.6 con $\alpha = 2$, en la que las funciones se volvían cada vez platocúrticas.

De hecho, si el valor del parámetro de forma de la componente Pareto se fija en $\alpha = 1$, manteniendo $\theta = 50$ y haciendo que r tome los valores $0,1; 0,2; 0,3; 0,4; 0,5$ y $0,6$, igual que en los casos anteriores, se obtiene el comportamiento mostrado en la Figura 3.8: a medida que aumenta r la moda decrece, pero el valor de la función de densidad evaluada en esta moda primero decrece a medida que aumenta r , para luego empezar a aumentar, llegando a superar el valor de la densidad evaluada en la moda para el primer valor del parámetro de peso más bajo $r = 0,1$.

En los tres casos, el cálculo del parámetro de σ se llevó a cabo de forma numérica resolviendo la ecuación (3.18)

$$r = \frac{\Phi(\alpha\sigma) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{1}{2}(\alpha\sigma)^2\right\}}{\Phi(\alpha\sigma) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{1}{2}(\alpha\sigma)^2\right\} + 1}$$

para los valores de α y r correspondientes.

El cálculo del parámetro μ se hizo de acuerdo a la restricción (3.16), de

donde se puede despejar el valor del parámetro de localización de la componente lognormal μ en función del parámetro de corte θ , el parámetro de forma de la componente Pareto α y el propio parámetro de escala de la componente lognormal σ :

$$\mu = \log(\theta) - \alpha\sigma^2.$$

3.3. Función de distribución

Integrando la función de densidad (3.5) es inmediato obtener la función de distribución del modelo de Scollnik.

Así, sea $x \leq \theta$. Entonces, utilizando la primera componente de esta densidad (3.5), se tiene que

$$\begin{aligned} F(x) &= \int_0^x f(\xi) d\xi = \int_0^x f(\xi) d\xi = \int_0^x r \tilde{f}_1(\xi) d\xi = \int_0^x r \frac{1}{\mathbb{P}[X \leq \theta]} f_1(\xi) d\xi \\ &= \frac{r}{\mathbb{P}[X \leq \theta]} \int_0^x f_1(\xi) d\xi = \frac{r}{\mathbb{P}[X \leq \theta]} \mathbb{P}[X \leq x], \end{aligned}$$

y utilizando el razonamiento de (3.8), esto se puede escribir como

$$F(x) = r \frac{\Phi((\ln(x) - \mu)/\sigma)}{\Phi((\ln(\theta) - \mu)/\sigma)}. \quad (3.26)$$

Sea ahora $x > \theta$. Entonces,

$$F(x) = \int_0^x f(\xi) d\xi = \int_0^\theta f(\xi) d\xi + \int_\theta^x f(\xi) d\xi = F(\theta) + \int_\theta^x (1-r) f_2(\xi) d\xi,$$

donde utilizando lo obtenido en (3.26),

$$\begin{aligned} &= r + (1-r) \int_\theta^x f_2(\xi) d\xi \\ &= r + (1-r) \int_\theta^x \frac{\alpha\theta^\alpha}{\xi^{\alpha+1}} d\xi \\ &= r + (1-r)\alpha\theta^\alpha \int_\theta^x \xi^{-(\alpha+1)} d\xi. \end{aligned} \quad (3.27)$$

Como $\alpha > 0$, $\alpha + 1 \neq 0$, y por tanto, la integral anterior es

$$\begin{aligned} &= r + (1-r)\alpha\theta^\alpha \left[-\frac{\xi^{-\alpha}}{\alpha} \right]_{\xi=\theta}^{\xi=x} = r - (1-r)\theta^\alpha \left[\xi^{-\alpha} \right]_{\xi=\theta}^{\xi=x} \\ &= r - (1-r)\theta^\alpha \left(\frac{1}{x^\alpha} - \frac{1}{\theta^\alpha} \right) = r - (1-r) \left(\left(\frac{\theta}{x} \right)^\alpha - 1 \right) \\ &= r + (1-r) \left(1 - \left(\frac{\theta}{x} \right)^\alpha \right). \end{aligned}$$

Nótese que

$$1 - \left(\frac{\theta}{x} \right)^\alpha, \quad x > \theta$$

es la función de distribución de una Pareto de parámetro de escala θ y de forma α , lo que está en consonancia con (3.27), puesto que f_2 es la densidad de una Pareto de parámetros θ y α .

Por tanto, la función de distribución del modelo de Scollnik es

$$F(x) = \begin{cases} r \frac{\Phi((\ln(x) - \mu)/\sigma)}{\Phi((\ln(\theta) - \mu)/\sigma)}, & 0 < x \leq \theta \\ r + (1-r) \left(1 - \left(\frac{\theta}{x} \right)^\alpha \right), & \theta < x < \infty \end{cases}. \quad (3.28)$$

3.4. Función cuantil

Igual que se hizo en la sección 2.9 para el modelo de Cooray & Ananda, se puede calcular la función cuantil para el modelo de Scollnik, usando la expresión (3.28) obtenida en 3.3.

Lo primero es observar que la función de distribución del modelo de Scollnik (3.28) es también monótona y estrictamente creciente, igual que la función de distribución de Cooray & Ananda, y por tanto, se puede volver a aplicar el razonamiento de la sección 2.9 para afirmar que la función de distribución del modelo (3.5) resulta invertible, y así, dado un $p \in (0, 1)$, la función cuantil aplicada a este p es simplemente

$$Q(p) = F^{-1}(p).$$

Siguiendo el esquema utilizado en la sección 2.9, se calcula la probabilidad acumulada hasta el umbral θ :

$$F(\theta) = r \frac{\Phi((\ln(\theta) - \mu)/\sigma)}{\Phi((\ln(\theta) - \mu)/\sigma)} = r.$$

Por tanto, la expresión de la función cuantil dependerá de si p es mayor o menor que el parámetro de peso r .

Dado un p , se tiene que la función cuantil $Q(p)$ es la solución en x a la ecuación $F(x) = p$, y si esta probabilidad verifica $p \leq r$, se tiene que la función cuantil es la solución en x a la ecuación

$$r \frac{\Phi((\ln(x) - \mu)/\sigma)}{\Phi((\ln(\theta) - \mu)/\sigma)} = p.$$

Como es bien sabido, la función de distribución de la normal estándar $\Phi(\cdot)$ es perfectamente invertible³, y por tanto es válido escribir

$$\frac{\ln(x) - \mu}{\theta} = \Phi^{-1} \left(\frac{p}{r} \Phi \left(\frac{\ln(\theta) - \mu}{\theta} \right) \right),$$

de donde se despeja el cuantil x :

$$x = \exp \left\{ \mu + \sigma \Phi^{-1} \left(\frac{p}{r} \Phi \left(\frac{\ln(\theta) - \mu}{\theta} \right) \right) \right\}.$$

Por otro lado, si $p > r$ es necesario utilizar la segunda componente de 3.28, resultando el cuantil la solución en x a la ecuación

$$r + (1 - r) \left(1 - \left(\frac{\theta}{x} \right)^\alpha \right) = p,$$

de donde se despeja que

$$x = \frac{\theta}{\left(1 - \frac{p-r}{1-r} \right)^{1/\alpha}}.$$

En definitiva, la función cuantil $Q(\cdot)$ del modelo de Scollnik se escribe como

$$Q(p) = \begin{cases} \exp \left\{ \mu + \sigma \Phi^{-1} \left(\frac{p}{r} \Phi \left(\frac{\ln(\theta) - \mu}{\theta} \right) \right) \right\}, & 0 < p \leq r \\ \frac{\theta}{\left(1 - \frac{p-r}{1-r} \right)^{1/\alpha}}, & r < p < 1 \end{cases} \quad (3.29)$$

³Aunque es invertible solamente de forma numérica.

3.5. Ajuste a los datos daneses

El ajuste del modelo de Scollnik a los datos daneses, igual que en el caso del modelo de Cooray & Ananda, se realizó con un programa en *R*. El esqueleto del programa es el mismo que el mostrado en 2.12, siendo el cambio esencial la definición de la función de verosimilitud, cuyo código se muestra a continuación:

```
lkh=function (x) {
  alpha=x[1]
  theta=x[2]
  sigma=x[3]
  mu=log(theta)-alpha*sigma^2
  exponencial=exp(0.5*alpha*alpha*sigma*sigma)
  numerador=sqrt(2*pi)*alpha*sigma*pnorm(alpha*sigma)*
    exponencial
  r=numerador/(numerador+1)
  lkh=0
  for (i in 1:length(trainingData)) {
    xi=trainingData[i]
    if(xi <= theta){
      f=r*dlnorm(xi,meanlog = mu,sdlog = sigma)/
        plnorm(theta,meanlog = mu,sdlog = sigma)
      lkh=lkh+log(f)
    }else{
      f=(1-r)*dpareto(xi-theta,shape = alpha,scale = theta)
      lkh=lkh+log(f)
    }
  }
  return(-lkh)
}
```

La función `lkh` acepta como parámetro un vector `x` que ha de tener tres componentes: `alpha`, `theta` y `sigma`, representando, como es natural, a los parámetros α , θ y σ del modelo (3.5).

Recuérdese que el modelo (3.5) de Scollnik consistía en mezclar con un peso variable r dos densidades: por un lado la de una lognormal con dos parámetros μ y σ , y por otro lado, la densidad de una Pareto, con otros dos

parámetros θ y α . Por tanto, inicialmente el modelo (3.5) dispone de cinco parámetros, pero al exigirle continuidad y diferenciabilidad, resultan las restricciones (3.10) y (3.17) que reducen el número de parámetros efectivos a tres, de tal suerte que α , θ y σ son parámetros libres de variar y tomar cualquier valor dentro de sus dominios de definición, concretamente, la semi-recta positiva $(0, \infty)$, mientras que los parámetros μ y r han de tomar un valor concreto que se expresa en función de los anteriores tres parámetros libres:

$$\mu = \ln(\theta) - \alpha\sigma^2 \quad (3.30)$$

$$r = \frac{\Phi(\alpha\sigma) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{1}{2}(\alpha\sigma)^2\right\}}{\Phi(\alpha\sigma) \sqrt{2\pi}\alpha\sigma \exp\left\{\frac{1}{2}(\alpha\sigma)^2\right\} + 1}, \quad (3.31)$$

donde para la expresión de r se utilizó la restricción (3.17) necesaria para la diferenciabilidad, obteniendo una expresión más simple que la que aparece en (3.10), tal y como se comentó al desarrollar (3.18).

Así, con las variables `alpha`, `theta` y `sigma` representando α , θ y σ se calculan las variables `mu` y `r`, que naturalmente representan a los parámetros μ y r usando las ecuaciones (3.30) y (3.31), obteniendo todo lo necesario para calcular la función de verosimilitud para los datos daneses usando la densidad (3.5).

Nótese que para calcular la verosimilitud correspondiente a los valores que están por encima del umbral θ se ha utilizado la función `dpareto` del paquete `actuar` [Dutang, Goulet & Pigeon, 2008] en vez de programar la función f_2 a mano.

En realidad, esto no sería complicado y bastaría hacer lo siguiente:

```
f2 <- function(x,alpha,theta) {
  return(alpha*theta^alpha/(x^(alpha+1)))
}
```

Así, se usaría `f2` en la definición de `lkh`. Sin embargo, se ha preferido usar la función incorporada en `actuar` explícitamente para darle crédito a este útil paquete que implementa muchos procedimientos útiles para la Ciencia Actuarial, y que sin duda ha sido de ayuda a lo largo de la realización de este trabajo.

Nótese que la densidad de la Pareto implementada en el paquete `actuar` es

$$\tilde{f}_2(x) = \frac{\alpha\theta^\alpha}{(x + \theta)^{\alpha+1}}$$

para $x > 0$, $\alpha > 0$ y $\theta > 0$, lo cual es ligeramente distinto de la parametrización de la distribución de Pareto usada en f_2 , y por tanto el argumento que se pasa a la función `dpareto` es `(xi-theta)`, en vez de `xi`.

Para poder utilizar la función `dpareto` basta cargar el paquete `actuar` antes de llamar a la función ejecutando la siguiente línea de código:

```
library(actuar)
```

En vista del parámetro σ añadido a la función de verosimilitud se ha de modificar también el vector `x` con los parámetros que se pasa a la función `dfp` que utiliza el algoritmo de optimización de Davidon-Fletcher-Powell para maximizar la verosimilitud.

Se añade la variable `sigma`, cuya estimación inicial se fija en `sigma = 2` y cuyo rango dentro del cual puede variar para maximizar la función de verosimilitud se establece en `(0, 100)`:

```
x <- list(label=c("alpha", "theta", "sigma"), est=c(1, 10, 2),
          low=c(0, 0, 0), upp=c(5, 10000, 100))
```

A continuación se muestra parcialmente la salida que se obtuvo tras ejecutar la función `dfp`:

```
## Optimization Result:
## iter: 13  fmin: 3133.858  nfcn: 188
##
##  label estimate  low      high
## 1 alpha 1.3059099 1.2382295 1.3759366
## 2 theta 1.199442  1.1313288 1.2716555
## 3 sigma 0.19727009 0.17171346 0.22662176
```

En este caso, los valores de los parámetros del modelo (3.5) tras maximizar la verosimilitud son

$$\begin{aligned}\hat{\alpha} &= 1,3059099 \\ \hat{\theta} &= 1,199442 \text{ ,} \\ \hat{\sigma} &= 0,19727009\end{aligned}\tag{3.32}$$

siendo el intervalo de confianza al 95 % para el parámetro α

$$[1,2382295; 1,3759366],$$

para el parámetro θ

$$[1,1313288; 1,2716555],$$

y para el parámetro σ

$$[0,17171346; 0,22662176].$$

Esto es, se tiene que

$$\begin{aligned} \mathbb{P}\{\alpha \in [1,2382295; 1,3759366]\} &= \mathbb{P}\{\theta \in [1,1313288; 1,2716555]\} \\ &= \mathbb{P}\{\sigma \in [0,17171346; 0,22662176]\} = 95 \%. \end{aligned}$$

Otro valor interesante que aparece en la última salida es la logverosimilitud *negativa* (véase el pie de página 3 de la página 26), que toma el valor de 3133,858. Por tanto, el valor de la logverosimilitud alcanzada tras la optimización en este caso es

$$-3133,858. \tag{3.33}$$

Puesto que la estructura del resto del programa que ajusta el modelo de Scollnik a los datos daneses es muy similar a la del programa que ajusta a estos mismos datos el modelo de Cooray & Ananda, mostrado completamente en 2.12, el correspondiente a este capítulo no se muestra en su totalidad. El lector interesado puede realizar las modificaciones necesarias expuestas en esta sección al código correspondiente al ajuste del modelo de Cooray & Ananda mostrado en 2.12.

En la Figura 3.9 se representa el histograma parcial de los datos sobre los siniestros derivados de grandes incendios en Dinamarca de la muestra del test, concretamente, el correspondiente a las pérdidas inferiores a 20 millones, junto con la función de densidad del modelo de Scollnik ajustado, esto es, utilizando los valores (3.32).

En la Figura 3.10 se han trazado la función de distribución empírica de este subconjunto reservado para la validación en línea negra sólida junto con la función de distribución del modelo ajustado, utilizando los valores (3.32), en línea roja discontinua.

De las Figuras 3.9 y 3.10 se deduce un ajuste satisfactorio, al menos de forma gráfica.

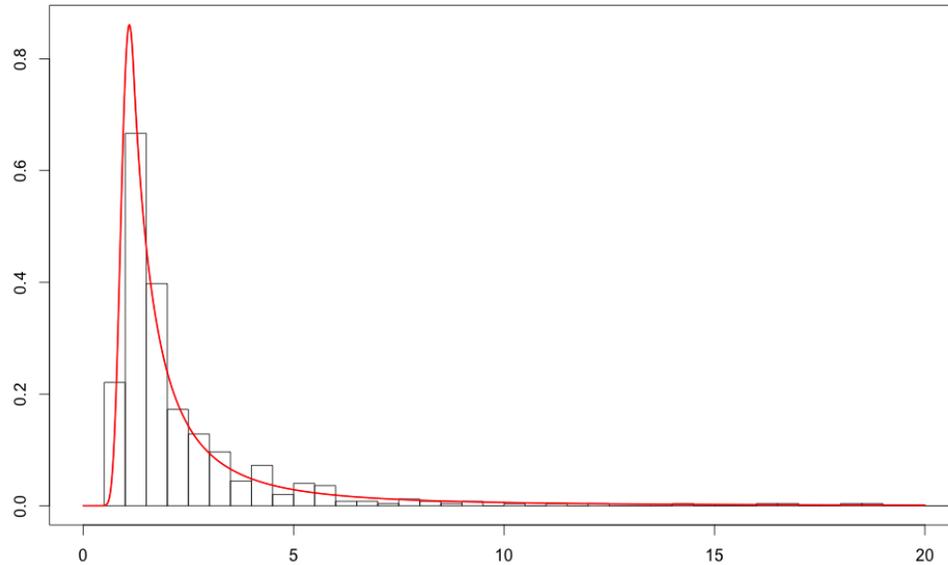


Figura 3.9: Histograma parcial correspondiente a valores inferiores a 20 millones de los incendios daneses con la función de densidad del modelo de Scollnik ajustado

3.6. Pruebas χ^2 y G

De igual modo que para el ajuste del modelo de Cooray & Ananda se llevaron a cabo las pruebas analíticas χ^2 y G en las secciones 2.13 y 2.14, éstas pruebas se realizaron también para el ajuste del modelo de Scollnik.

Igual que ocurría con el programa del ajuste del modelo en sí, la estructura del programa que realiza las pruebas χ^2 y G al ajuste del modelo (3.5) es similar a la del programa que ejecuta estas pruebas al modelo ajustado en el Capítulo 2.

La única diferencia importante deriva de que el modelo de Scollnik posee un parámetro más que el modelo de Cooray & Ananda, y por tanto, los grados de libertad de la distribución χ^2 utilizada en los test disminuyen en uno⁴.

⁴Recuérdese que bajo la hipótesis nula de que el modelo es correcto, los estadísticos de

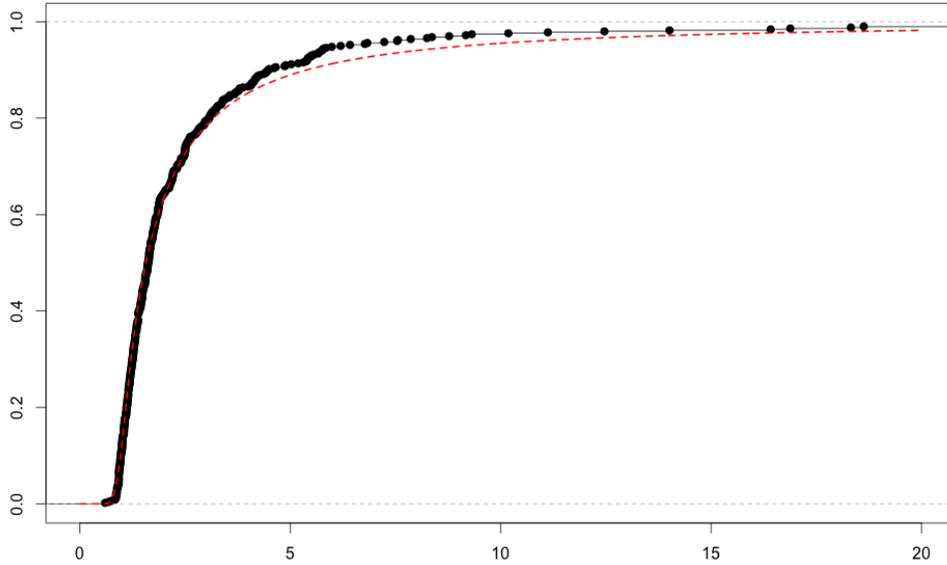


Figura 3.10: Función de distribución empírica (línea negra sólida) junto con la función de distribución del modelo de Scollnik ajustado (línea roja discontinua) en el tramo de los valores inferiores a 20 millones de los incendios daneses

El resultado de llevar a cabo la prueba χ^2 al ajuste del modelo de Scollnik considerando para la prueba las mismas categorías (2.26) usadas en la sección 2.13 para comprobar la validez del modelo de Cooray & Ananda se muestra a continuación. Nótese que los grados de libertad df toman ahora el valor

$$\text{length}(\text{frecuenciasObservadas})-3-1$$

en vez del valor

$$\text{length}(\text{frecuenciasObservadas})-2-1$$

los tests χ^2 y G se aproximan asintóticamente a una distribución χ^2 de $k - p - 1$ grados de libertad, siendo k el número de categorías especificadas y p el número de parámetros estimados en el modelo ajustado.

Si el número de parámetros p aumenta, los grados de libertad de la distribución χ^2 disminuyen en una unidad por cada parámetro adicional.

por el comentario del pie de página 4 de la página 69.

```
pchisq(estadistico,df=length(frecuenciasObservadas)-3-1,
       lower.tail = FALSE)

## [1] 0.009866088
```

El p-valor es 0,009866088, lo cual haría rechazar la hipótesis nula de la validez del modelo incluso con el nivel de significación 0,01.

Si se practica la corrección de Yates explicada en la sección 2.13, la salida obtenida es la siguiente:

```
pchisq(estadisticoCorregido,
       df=length(frecuenciasObservadas)-3-1,
       lower.tail = FALSE)

## [1] 0.02799944
```

El modelo se vuelve significativo al nivel de significación 0,01, pero sigue sin serlo al nivel 0,05: el p-valor es bajo.

A continuación, se muestra la salida resultante de realizar la prueba G al ajuste del modelo (3.5).

```
pchisq(estadisticoG,df=length(frecuenciasObservadas)-3-1,
       lower.tail = FALSE)

## [1] 0.01033832
```

El p-valor obtenido es significativo al nivel 0,01, pero sigue siendo bajo, sin ser significativo al nivel 0,05, siendo su valor concretamente de 0,01033832.

Puede parecer que el modelo de Scollnik ofrece un ajuste peor que el modelo de Cooray & Ananda, pero esto es imposible debido a que, como se ha visto en la sección 3.1.4, el modelo de Cooray & Ananda es un caso particular del modelo de Scollnik, y en caso de ser mejor el modelo más sencillo –con pesos fijos e introducido en el Capítulo 2–, el algoritmo de la maximización de la verosimilitud llevaría a su elección.

Se concluye que la significación de los modelos continuos calculada con las pruebas χ^2 y G depende fuertemente de la elección inicial de la partición de los datos, y pudiéndose obtener resultados completamente incongruentes.

Es precisamente por este tipo de resultados por los que puede ser más aconsejable la evaluación de la bondad de ajuste mediante gráficos de la función de densidad frente al histograma, como el de las Figuras 2.8 y 3.9, comparación de las funciones de distribución empírica y teórica, como la de las Figuras 2.9 y 3.10, y sobre todo, los gráficos QQ, como los de las Figuras 2.10 y 3.11.

3.7. Gráfico QQ

De forma similar a lo realizado en la sección 2.15 para el modelo de Cooray & Ananda, en esta sección se evalúa la bondad de ajuste del modelo de Scollnik a los datos daneses mediante la técnica de los gráficos QQ.

En la Figura 3.11 se presenta el gráfico QQ de los datos daneses frente al modelo (3.5) de Scollnik ajustado. De un modo similar al anterior, los cuantiles teóricos del eje de las abscisas fueron calculados de acuerdo a la función cuantil (3.29) obtenida en la sección 3.4, y los empíricos del eje de las ordenadas usando la definición 7 de [Hyndman & Fan, 1996] explicada en 2.10.

El gráfico QQ obtenido en este capítulo es muy similar al obtenido en el Capítulo 2 para el modelo de Cooray & Ananda, y valen los mismos comentarios hechos en la sección 2.15, por lo que en esta ocasión, es aconsejable llevar otras pruebas para la validación de este modelo, como los presentados en las secciones 3.8 y 3.9.

3.8. Ratio de verosimilitudes

Como una generalización del lema de Neyman-Pearson, Wilks desarrolla en uno de sus artículos [Wilks, 1938] el llamado *test del ratio de verosimilitudes*.

Dados dos modelos anidados, esto es, dos modelos en los que uno de ellos es un caso particular del otro modelo más general, y llamando L_0 a la verosimilitud obtenida con el primer modelo y L_a a la verosimilitud proporcionada por el modelo más general, se tiene que el cociente

$$\frac{L_0}{L_a}$$

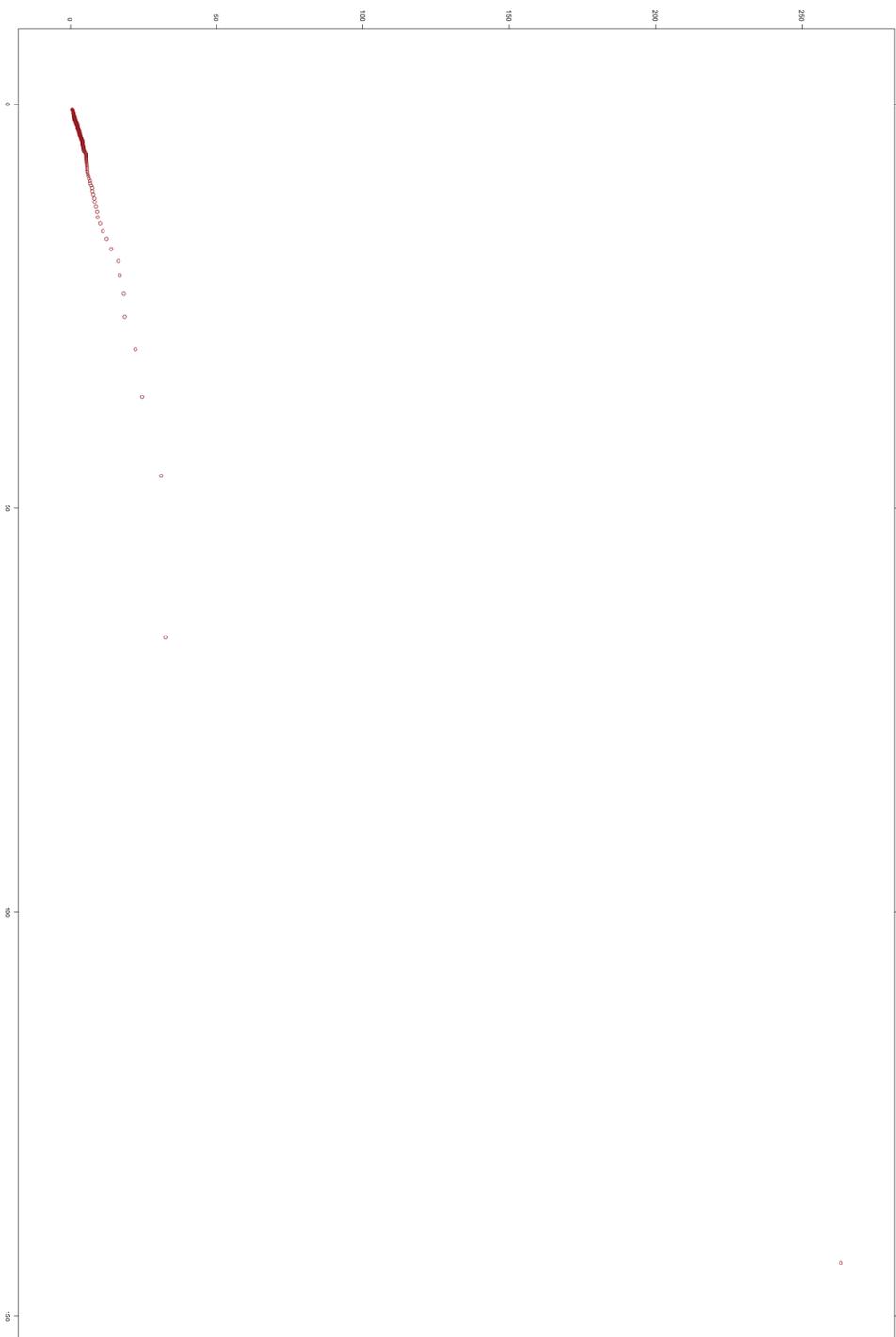


Figura 3.11: Gráfico QQ de los datos daneses frente al modelo de Scollnik ajustado

es un valor entre 0 y 1, a consecuencia de que las verosimilitudes son siempre cuantías no negativas y el modelo más general siempre proporcionará un ajuste al menos tan bueno como el modelo particular, y por tanto, siempre se verifica que $L_0 \leq L_a$.

Wilks demuestra en su artículo que el estadístico

$$D = -2 \ln \left(\frac{L_0}{L_a} \right) = 2 \ln \left(\frac{L_a}{L_0} \right) = 2 (\ln(L_a) - \ln(L_0)) \quad (3.34)$$

sigue asintóticamente⁵ una distribución χ^2 con el número de grados de libertad dado por la diferencia entre la cantidad de parámetros del modelo general y el particular, bajo la hipótesis nula de validez del modelo particular.

Recuérdese que el modelo de Cooray & Ananda es un caso particular del modelo de Scollnik, y por tanto, es posible aplicar la prueba de Wilks para averiguar si hay evidencia estadística suficiente para desechar el modelo particular de Cooray & Ananda en favor del modelo más general de Scollnik.

En (2.24) de la sección 2.12 se obtuvo que la logverosimilitud obtenida con el modelo de Cooray & Ananda es

$$L_0 = -3144,059,$$

mientras que la logverosimilitud proporcionada por el modelo de Scollnik fue mostrada en (3.33) de la sección 3.5, siendo su valor de

$$L_a = -3133,858.$$

Así, es inmediato calcular el estadístico D de (3.34) para este caso:

$$D = 2(-3133,858 + 3144,059) = 20,402. \quad (3.35)$$

Puesto que la diferencia entre el número de parámetros de los modelos de Cooray & Ananda y Scollnik es 1, con la instrucción siguiente se obtiene el p-valor para la cantidad (4.21) del estadístico D .

```
pchisq(20.402,df=1,lower.tail = FALSE)
```

```
## [1] 6.276417e-06
```

Tal y como se puede apreciar, se tiene evidencia estadística suficiente a cualquier nivel de significación usual para afirmar que los datos daneses de la muestra para el test siguen el modelo de Scollnik en vez del modelo de Cooray & Ananda.

⁵La diferencia entre la distribución exacta de (3.34) y la χ^2 es del orden de $\mathcal{O}(\frac{1}{\sqrt{n}})$, siendo n el número de observaciones.

3.9. Criterio de información de Akaike

En 1974 Akaike publicó un artículo [Akaike, 1974] en el que desarrollaba una medida de bondad de ajuste para modelos estadísticos. Basado en la teoría de la información, el criterio de Akaike proporciona una estimación de la cantidad de la información perdida al representar un conjunto de datos con un modelo. Se calcula a partir de la verosimilitud obtenida con el modelo ajustado y el número de parámetros de éste, y trata de cuantificar el nivel de compromiso entre la calidad del ajuste y la complejidad del modelo. Su fórmula exacta es

$$\text{AIC} = 2k - 2\ln(L), \quad (3.36)$$

donde k es el número de parámetros del modelo, y L el valor de la función de verosimilitud alcanzado en el óptimo.

Son preferibles los modelos donde el AIC es más bajo, porque su valor decrece conforme aumenta la verosimilitud, y por tanto, la bondad del ajuste, y aumenta a medida que se incorporan más parámetros en el modelo, y por tanto, su complejidad.

Nótese que el multiplicar por 2 los sumandos k y $-\ln(L)$ en (3.36) se hace para asimilar los valores del criterio de Akaike a otras medidas de bondad de ajuste como el ratio de verosimilitudes discutido en la sección 3.8 y hacerlos más comparables: de cara a usar solamente el AIC para seleccionar entre modelos alternativos es equivalente elegir el modelo que minimice el valor (3.36) a elegir aquel modelo que haga mínima la diferencia $k - \ln(L)$, pues es simplemente una cuestión de multiplicación por una constante.

La ventaja del criterio de Akaike es que, a diferencia del ratio de verosimilitudes, permite comparar modelos no anidados. El principal inconveniente es que *no* es un test estadístico en el sentido de que no se conoce la distribución de (3.36), ni siquiera asintóticamente. Por tanto, el AIC se utiliza para elegir el modelo más adecuado entre los posibles, pero si ninguno de ellos proporciona un ajuste satisfactorio, el criterio de Akaike no alertará sobre esta circunstancia, por ejemplo, mediante un p-valor no significativo.

Usando la logverosimilitud (2.24) del modelo de Cooray & Ananda calculada en 2.12, y teniendo en cuenta que este modelo tiene dos parámetros, es posible calcular el criterio de información de Akaike (3.36) para este ajuste:

$$\text{AIC}_C = 2 \cdot 2 - 2(-3144,059) = 6292,118.$$

Análogamente, usando la logverosimilitud (3.33) del modelo de Scollnik calculada en 3.5, y teniendo en cuenta que este modelo tiene 3 parámetros,

el criterio AIC para este ajuste queda del siguiente modo:

$$AIC_S = 2 \cdot 3 - 2(-3133,858) = 6273,716.$$

Puesto que $AIC_S < AIC_C$, se deduce que el criterio de información de Akaike también aconseja elegir el modelo de Scollnik frente al modelo de Cooray & Ananda.

Capítulo 4

Modelos basados en las distribuciones lognormal y Pareto generalizada

En los Capítulos 2 y 3 se obtuvieron unos ajustes satisfactorios para los datos daneses. La distribución de Pareto definida en (2.4) explica los valores de la cola pesada de un modo adecuado, pero en realidad no se ha visto aún una justificación teórica para el uso de esta distribución para modelar las colas de las distribuciones de siniestros. En este capítulo se va a tratar de encontrar esta justificación, y de construir un modelo aún más general que en el Capítulo 3.

4.1. Distribución de Pareto generalizada

Para un $\xi \neq 0$, se define la función de distribución de la Pareto generalizada (conocida también como *GPD*, por sus siglas en inglés: *Generalized Pareto Distribution*) como

$$G_{\xi,\beta}(x) = 1 - \left(1 + \frac{\xi}{\beta}x\right)^{-1/\xi}, \quad (4.1)$$

donde $\beta > 0$ es el parámetro de escala, y el valor de ξ controla la forma de la distribución.

Esta parametrización es la usual, y es la que aparece en [McNeil, 1997].

Por continuidad, para $\xi = 0$, se define

$$G_{\xi=0,\beta}(x) = 1 - \exp\left(-\frac{x}{\beta}\right).$$

Nótese que este último caso se corresponde simplemente con la distribución exponencial.

El dominio de definición para la distribución de Pareto generalizada cuando $\xi \geq 0$ es $x \geq 0$, y cuando $\xi < 0$, el dominio se convierte en el intervalo acotado $[0, -\beta/\xi]$.

Esta familia de distribuciones de Pareto generalizadas se puede extender añadiendo un parámetro de localización μ , y definiendo la función de distribución que dependa de tres parámetros: los ξ y β anteriores, de forma y de escala respectivamente, y el nuevo parámetro de localización μ del siguiente modo:

$$G_{\xi,\mu,\beta}(x) = G_{\xi,\beta}(x - \mu). \quad (4.2)$$

En este caso, el dominio de definición para los casos en los que $\xi \geq 0$ se convierte en $x \geq \mu$, y en el caso de que $\xi < 0$, el intervalo acotado del dominio de definición es $[\mu, \mu - \beta/\xi]$.

El principal atractivo de esta distribución es el que se enuncia en el teorema de Pickands–Balkema–de Haan [Balkema & De Haan, 1974; Pickands, 1975]. A grandes rasgos, establece que la distribución de los excesos por encima de un cierto umbral u tiende a la distribución de Pareto generalizada para una gran clase de distribuciones de probabilidad, a medida que aumenta u .

Más en concreto, sea F una función de distribución y sea x_0 el extremo derecho, finito o infinito, de la distribución de probabilidad F . Esto es,

$$x_0 = \sup \{x \in \mathbb{R} : F(x) < 1\} \leq \infty.$$

Se define la función de distribución de los excesos por encima de un umbral u como

$$F_u(x) = \mathbb{P}[X - u \leq x | X > u],$$

y por la definición de la probabilidad condicionada,

$$\begin{aligned} &= \frac{\mathbb{P}[u < X \leq x + u]}{\mathbb{P}[X > u]} \\ &= \frac{F(x + u) - F(u)}{1 - F(u)}, \end{aligned}$$

donde $0 \leq x < x_0 - u$.

El teorema de Pickands–Balkema–de Haan establece que bajo unas condiciones razonables sobre la posibilidad de normalizar las realizaciones de la distribución de probabilidad subyacente, la función de distribución de los excesos $F_u(\cdot)$ por encima de un umbral u tiende a la distribución de Pareto generalizada (4.1) a medida que se aumenta este umbral u de tal suerte que

$$\lim_{u \rightarrow x_0} \sup_{0 \leq x < x_0 - u} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0, \quad (4.3)$$

para unos ciertos valores de ξ y $\beta(u)$ a estimar. Nótese que se escribe $\beta(u)$ en vez de simplemente β para hacer explícito que el parámetro de escala de la distribución de Pareto generalizada límite es dependiente del umbral u escogido.

Por tanto, el teorema de Pickands–Balkema–de Haan sugiere que para unos umbrales suficientemente altos es adecuado ajustar una distribución de Pareto a los excesos en vista de la identidad (4.3): de hecho, lo que se deduce de este teorema es que si se verifican las hipótesis del teorema y el umbral es suficientemente alto la distribución (4.1) es *la única* distribución posible.

Así, parece apropiado construir un modelo en el que la cola de la distribución a partir del parámetro de corte $x = \theta$ siga una distribución Pareto generalizada dada por (4.2) con $\mu = \theta$. Se está usando la parametrización general con el parámetro de escala en vista de que el modelo compuesto se aplicará a datos de siniestros *ground-up*, esto es, a datos de siniestros brutos y no a excesos por encima del umbral $u = \theta$.

Haciendo el cambio de variable $\alpha = 1/\xi$, la función de distribución que gobernará la cola de este modelo se puede escribir como

$$G(x) = 1 - \left(1 + \frac{x - \theta}{\alpha\beta}\right)^{-\alpha}, \quad (4.4)$$

donde $\theta > 0$, $\alpha > 0$ y $\beta > 0$ y el dominio de definición de la distribución es $x \geq \theta$. Nótese que decir que $\alpha > 0$ es equivalente a decir que $\xi > 0$. El caso en el que $\xi = 0$ sería el modelo con una cola exponencial, la cual no es pesada y no respondería al objetivo de encontrar un modelo que sea capaz de explicar los siniestros más graves del sector asegurador. El caso de $\xi < 0$ se corresponde con una cola acotada superiormente, y aparece cuando la distribución subyacente también lo está, como la distribución uniforme o la distribución Beta. Este tipo de distribuciones no parecen útiles para explicar

la cola de los siniestros observados, y por tanto se asume que la distribución de Pareto generalizada tiene su parámetro de forma ξ estrictamente positivo.

Derivando en (4.4) se obtiene la densidad de la cola del modelo a construir:

$$g(x) = \frac{\alpha(\alpha\beta)^\alpha}{(\alpha\beta - \theta + x)^{\alpha+1}},$$

para $x > \theta$. Definiendo $\lambda = \alpha\beta - \theta$, lo anterior se puede escribir como

$$g(x) = \frac{\alpha(\lambda + \theta)^\alpha}{(\lambda + x)^{\alpha+1}}, \quad (4.5)$$

también para $x > \theta$.

La parametrización (4.5) permite apreciar claramente que la distribución de Pareto generalizada engloba a la distribución de Pareto usual (2.4): cuando α es positivo, y el parámetro de escala β se toma como $\theta\xi = \theta/\alpha$, resulta que $\lambda = 0$, y (4.5) se transforma en la densidad (2.4) de la Pareto usual.

Nótese que cuando α es estrictamente positivo¹, el producto $\alpha\beta$ es también estrictamente mayor que cero, y por tanto, en vista de la definición del parámetro λ como $\alpha\beta - \theta$, éste ha de verificar que $\lambda > -\theta$.

4.2. Modelo compuesto

De manera análoga a lo que se hizo en el Capítulo 2 en (3.5), se define el siguiente modelo compuesto:

$$f(x) = \begin{cases} r\tilde{f}_1(x), & 0 < x \leq \theta \\ (1-r)g(x), & \theta < x < \infty \end{cases}, \quad (4.6)$$

donde igual que antes r es un parámetro de peso que varía entre 0 y 1, y \tilde{f}_1 es la densidad de una distribución lognormal de parámetros μ y σ truncada al intervalo $(0, \theta)$ tal y como aparece en (3.4), y g es la densidad de una Pareto generalizada parametrizada con los parámetros λ , θ y α como en (4.5).

¹Esto ocurre si, y solamente si, la cola es más pesada que la cola de una distribución exponencial.

4.2.1. Condición de continuidad

En este caso, la condición de continuidad en $x = \theta$

$$r\tilde{f}_1(\theta) = (1 - r)g(\theta)$$

deriva en la siguiente expresión para el parámetro de peso r :

$$r = \frac{\sqrt{2\pi}\alpha\theta\sigma\Phi\left(\frac{\ln(\theta)-\mu}{\sigma}\right)\exp\left\{\frac{1}{2}\left(\frac{\ln(\theta)-\mu}{\sigma}\right)^2\right\}}{\sqrt{2\pi}\alpha\theta\sigma\Phi\left(\frac{\ln(\theta)-\mu}{\sigma}\right)\exp\left\{\frac{1}{2}\left(\frac{\ln(\theta)-\mu}{\sigma}\right)^2\right\} + \lambda + \theta}. \quad (4.7)$$

Obsérvese que este parámetro de peso, igual que en el caso del modelo (3.5), sigue variando entre 0 y 1, dependiendo su valor concreto de los valores que tomen los parámetros θ , μ , σ , α y λ .

4.2.2. Condición de diferenciabilidad

Análogamente, la condición de diferenciabilidad

$$\tilde{f}'_1(\theta) = (1 - r)g'(\theta)$$

se transforma en la siguiente restricción sobre los parámetros:

$$\frac{\ln(\theta) - \mu}{\sigma} = \left(\frac{\alpha\theta - \lambda}{\lambda + \theta}\right)\sigma,$$

de donde se puede despejar μ en función de los demás parámetros:

$$\mu = \ln(\theta) - \left(\frac{\alpha\theta - \lambda}{\lambda + \theta}\right)\sigma^2. \quad (4.8)$$

De ese modo, el modelo (4.6) bajo las restricciones de continuidad y diferenciabilidad ve reducidos sus seis parámetros iniciales a cuatro parámetros efectivos: θ , σ , α y λ , siendo los otros dos parámetros, r y μ , expresados en función de estos cuatro, tal y como se muestra en (4.7) y (4.8).

4.2.3. Condición de diferenciabilidad de segundo orden

A diferencia del modelo del Capítulo 3, el modelo (4.6) sí puede ser sometido a una condición de diferenciabilidad de segundo orden $\tilde{f}_1''(\theta) = (1-r)g''(\theta)$, dando lugar a la restricción

$$\frac{\ln(\theta) + \sigma^2 - \mu - 1}{\theta^2 \sigma^2} = \frac{\alpha + 1}{(\lambda + \theta)^2}, \quad (4.9)$$

de donde se puede despejar el parámetro α del siguiente modo:

$$\alpha = \frac{\ln(\theta) + \sigma^2 - \mu - 1}{\theta^2 \sigma^2} (\lambda + \theta)^2 - 1. \quad (4.10)$$

Bajo la restricción adicional (4.10), el modelo pasa a tener tres parámetros efectivos (θ , σ y λ), siendo los otros tres parámetros (r , μ y α) expresados en función de éstos (véase las ecuaciones (4.7), (4.8) y (4.10)).

En este caso, para expresar el parámetro α en función de los tres parámetros θ , σ y λ efectivos solamente, es necesario despejar μ de (4.9), lo que da lugar a la ecuación

$$\mu = \ln(\theta) + \sigma^2 - 1 - \frac{\alpha + 1}{(\lambda + \theta)^2} \theta^2 \sigma^2,$$

e igualando ésta a (4.8) se obtiene la identidad

$$\sigma^2 - 1 - \frac{\alpha + 1}{(\lambda + \theta)^2} \theta^2 \sigma^2 = - \left(\frac{\alpha \theta - \lambda}{\lambda + \theta} \right) \sigma^2,$$

de la cual, con un poco de álgebra se despeja que

$$\alpha = \frac{1 - \frac{1}{\sigma^2} - \frac{\lambda}{\lambda + \theta} - \left(\frac{\theta}{\lambda + \theta} \right)^2}{\left(\frac{\theta}{\lambda + \theta} \right)^2 - \frac{\theta}{\lambda + \theta}}. \quad (4.11)$$

Con esto, se tiene que una vez decididos los valores de los parámetros efectivos θ , σ y λ , (4.11) permite calcular el valor de α en función de éstos para el caso del modelo sujeto a la condición de diferenciabilidad de segundo orden; conocido el valor de α , (4.8) facilita el cálculo del parámetro de localización de la componente lognormal μ ; y este a su vez, junto con los valores de los demás parámetros permite calcular el valor del parámetro de peso r según la ecuación (4.7).

4.3. Función de distribución

Puesto que las restricciones sobre los parámetros para garantizar continuidad, diferenciabilidad y diferenciabilidad de segundo orden no afectan a la definición del modelo (4.6), se puede calcular la función de distribución del modelo de Scollnik con cola GPD independientemente de si éste se ve restringido a condiciones de continuidad y diferenciabilidad de primer orden solamente, o también a la condición (4.9) de diferenciabilidad de segundo orden.

En vista de que la primera componente de la cabeza definida hasta $x = \theta$ del modelo (4.6) es *exactamente* igual a la del modelo estudiado en el Capítulo 3, la función de distribución de los modelos presentados en este capítulo para valores inferiores a θ vale *exactamente* lo mismo que la función de distribución calculada en 3.3. Así, si $x \leq \theta$, la función de distribución del modelo (4.6) vale

$$F(x) = r \frac{\Phi((\ln(x) - \mu) / \sigma)}{\Phi((\ln(\theta) - \mu) / \sigma)}, \quad x \leq \theta. \quad (4.12)$$

Ahora, si $x > \theta$,

$$\begin{aligned} F(x) &= \int_0^x f(\xi) d\xi = \int_0^{\theta} f(\xi) d\xi + \int_{\theta}^x f(\xi) d\xi = F(\theta) + \int_{\theta}^x (1-r)g(\xi) d\xi \\ &= r + (1-r) \int_{\theta}^x g(\xi) d\xi, \end{aligned}$$

donde g es la función de densidad dada por (4.5).

Por definición, la integral

$$\int_{\theta}^x g(\xi) d\xi$$

vale $G(x)$, donde G es la función de distribución dada por (4.4).

Por otro lado, puesto que $\lambda = \alpha\beta - \theta$, es inmediato despejar β en función de los demás parámetros, obteniendo

$$\beta = \frac{\lambda + \theta}{\alpha}.$$

Sustituyendo esto en (4.4) se llega a que

$$G(x) = 1 - \left(1 + \frac{x - \theta}{\lambda + \theta}\right)^{-\alpha},$$

y por tanto, para $x \geq \theta$ la función de distribución del modelo (4.6) vale

$$F(x) = r + (1 - r) \left[1 - \left(1 + \frac{x - \theta}{\lambda + \theta} \right)^{-\alpha} \right], \quad x \geq \theta. \quad (4.13)$$

Resumiendo (4.12) y (4.13) en una ecuación, queda la siguiente expresión para la función de distribución para el modelo de Scollnik con cola GPD:

$$F(x) = \begin{cases} r \frac{\Phi((\ln(x) - \mu)/\sigma)}{\Phi((\ln(\theta) - \mu)/\sigma)}, & 0 < x \leq \theta \\ r + (1 - r) \left[1 - \left(1 + \frac{x - \theta}{\lambda + \theta} \right)^{-\alpha} \right], & \theta < x < \infty \end{cases}.$$

4.4. Función cuantil

En vista de lo expuesto en la sección 4.3 sobre la igualdad de las cabezas de los modelos de Scollnik con cola Pareto o cola GPD, se tiene que para $x \leq \theta$ las dos funciones de distribución son exactamente iguales. Puesto que las funciones de distribución son funciones no decrecientes, esto implica que también las funciones cuantil son exactamente iguales para $p \leq F(\theta) = r$. Por tanto, se tiene que para el modelo (4.6) y para $p \leq r$, la función cuantil vale

$$Q(p) = \exp \left\{ \mu + \sigma \Phi^{-1} \left(\frac{p}{r} \Phi \left(\frac{\ln(\theta) - \mu}{\sigma} \right) \right) \right\}. \quad (4.14)$$

Si $p > r$, es necesario utilizar la componente GPD del modelo, siendo la función cuantil $Q(p)$ la solución en x a la ecuación

$$p = r + (1 - r) \left[1 - \left(1 + \frac{x - \theta}{\lambda + \theta} \right)^{-\alpha} \right],$$

de donde se despeja que

$$Q(p) = \theta + (\lambda + \theta) \left[1 + \left(1 - \frac{p - r}{1 - r} \right)^{-1/\alpha} \right]. \quad (4.15)$$

Agrupando las expresiones (4.14) y (4.15) en una sola, se tiene la función cuantil para el modelo (4.6):

$$Q(p) = \begin{cases} \exp \left\{ \mu + \sigma \Phi^{-1} \left(\frac{p}{r} \Phi \left(\frac{\ln(\theta) - \mu}{\sigma} \right) \right) \right\}, & 0 < p \leq r \\ \theta + (\lambda + \theta) \left[1 + \left(1 - \frac{p - r}{1 - r} \right)^{-1/\alpha} \right], & r < p < 1 \end{cases}. \quad (4.16)$$

4.5. Ajuste a los datos daneses

En esta sección se llevan a cabo dos ajustes a los datos daneses. El primero consiste en ajustar a estos datos de entrenamiento ya utilizados en capítulos previos el modelo (4.6) sujeto a las condiciones de continuidad y diferenciabilidad de primer orden, pero sin utilizar la condición (4.9) de diferenciabilidad de segundo orden, y un segundo ajuste basado en el modelo (4.6), pero esta vez haciendo uso de todas las restricciones: las de continuidad y diferenciabilidad de primer orden, y también de la condición de diferenciabilidad de segundo orden (4.9).

4.5.1. Ajuste del modelo con restricciones de continuidad y diferenciabilidad de primer orden

Al igual que se hizo en la sección análoga a esta en el Capítulo 3 al estudiar el ajuste del modelo de Scollnik con la cola Pareto ordinaria, solamente se muestra el código *R* de la función de verosimilitud, puesto que el resto de la estructura del programa es semejante a la expuesta en detalle en el Capítulo 2, pudiendo el lector interesado replicar el programa del ajuste del modelo de Scollnik con la cola GPD de este capítulo utilizando la siguiente función de verosimilitud:

```
lkh=function (x) {
  alpha=x[1]
  theta=x[2]
  sigma=x[3]
  lambda=x[4]
  mu=log(theta)-((alpha*theta-lambda)/(lambda+theta))*sigma^2
  z=(log(theta)-mu)/sigma
  factor=sqrt(2*pi)*alpha*theta*sigma*pnorm(z)*exp(0.5*z*z)
  r=factor/(factor+lambda+theta)
  lkh=0
  for (i in 1:length(trainingData)) {
    xi=trainingData[i]
    if(xi <= theta){
      f=r*dlnorm(xi,meanlog = mu,sdlog = sigma)/
        plnorm(theta,meanlog = mu,sdlog = sigma)
      lkh=lkh+log(f)
    }
  }
}
```

```

}else{
  gxi=(alpha*(lambda+theta)^alpha)/((lambda+xi)^(alpha+1))
  f=(1-r)*gxi
  lkh=lkh+log(f)
}
}
return(-lkh)
}

```

La función `lkh` de este modelo acepta como parámetro un vector \mathbf{x} que ha de tener cuatro componentes: `alpha`, `theta`, `sigma` y `lambda`, representando, como es natural, a los parámetros α , θ , σ y λ del modelo (4.6).

Similarmente al modelo (3.5), el modelo (4.6) consiste en mezclar con un peso variable r dos densidades: por un lado la de una lognormal con dos parámetros μ y σ , y por otro lado, la densidad de una Pareto generalizada, con otros tres parámetros θ , α y λ . Por tanto, inicialmente el modelo (4.6) dispone de seis parámetros, pero al exigirle continuidad y diferenciabilidad, resultan las restricciones (4.7) y (4.8) que reducen el número de parámetros efectivos a cuatro, de tal suerte que α , θ , σ y λ son parámetros libres de variar y tomar cualquier valor dentro de sus dominios de definición, concretamente, la semirrecta positiva $(0, \infty)$ en el caso de α , θ , σ y la semirrecta $(-\theta, \infty)$ en el caso de λ , mientras que los parámetros μ y r han de tomar un valor concreto que se expresa en función de los anteriores tres parámetros libres utilizando las mencionadas restricciones (4.7) y (4.8):

$$\mu = \ln(\theta) - \left(\frac{\alpha\theta - \lambda}{\lambda + \theta} \right) \sigma^2$$

$$r = \frac{\sqrt{2\pi}\alpha\theta\sigma\Phi\left(\frac{\ln(\theta)-\mu}{\sigma}\right)\exp\left\{\frac{1}{2}\left(\frac{\ln(\theta)-\mu}{\sigma}\right)^2\right\}}{\sqrt{2\pi}\alpha\theta\sigma\Phi\left(\frac{\ln(\theta)-\mu}{\sigma}\right)\exp\left\{\frac{1}{2}\left(\frac{\ln(\theta)-\mu}{\sigma}\right)^2\right\} + \lambda + \theta}.$$

Así, con las variables `alpha`, `theta`, `sigma` y `lambda` representando α , θ , σ y λ se calculan las variables `mu` y `r`, que naturalmente representan a los parámetros μ y r usando las ecuaciones anteriores, obteniendo todo lo necesario para calcular la función de verosimilitud para los datos daneses con la ayuda de la densidad (4.6).

Puesto que se añade un nuevo parámetro λ a la función de verosimilitud se ha de modificar también el vector \mathbf{x} con los parámetros que se pasa a la

función `dfp` que utiliza el algoritmo de optimización de Davidon-Fletcher-Powell para maximizar la verosimilitud.

Por ello, se agrega al vector `x` la variable `lambda`, cuya estimación inicial se fija en `lambda=10` y cuyo rango dentro del cual puede variar para maximizar la función de verosimilitud se establece en $(0, 1000)$:

```
x <- list(label=c("alpha", "theta", "sigma", "lambda"),
          est=c(1, 10, 2, 10),
          low=c(0, 0, 0, 0), upp=c(5, 10000, 100, 1000))
```

La salida parcial obtenida tras ejecutar la función `dfp` es la siguiente:

```
## Optimization Result:
## iter: 44   fmin: 3130.023   nfcn: 584
##
##   label  estimate   low      high
## 1 alpha  1.5180021  1.3405478  1.7079831
## 2 theta  1.1429054  1.0810346  1.2083168
## 3 sigma  0.18486076  0.16173718  0.21128332
## 4 lambda 0.33207661  0.15042104  0.73309199
```

Los valores de los parámetros del modelo (4.6) tras maximizar la verosimilitud aplicada a los datos daneses son

$$\begin{aligned}\hat{\alpha} &= 1,5180021 \\ \hat{\theta} &= 1,1429054 \\ \hat{\sigma} &= 0,18486076 \\ \hat{\lambda} &= 0,33207661\end{aligned}\tag{4.17}$$

siendo el intervalo de confianza al 95 % para el parámetro α

$$[1,3405478; 1,7079831],$$

para el parámetro θ

$$[1,0810346; 1,2083168],$$

para el parámetro σ

$$[0,16173718; 0,21128332],$$

y para el parámetro λ

$$[0,15042104; 0,73309199].$$

Esto es, se tiene que

$$\begin{aligned} \mathbb{P}\{\alpha \in [1,3405478; 1,7079831]\} &= \mathbb{P}\{\theta \in [1,0810346; 1,2083168]\} \\ &= \mathbb{P}\{\sigma \in [0,16173718; 0,21128332]\} \\ &= \mathbb{P}\{\lambda \in [0,15042104; 0,73309199]\} = 95\%. \end{aligned}$$

Otro valor interesante que aparece en la última salida es la logverosimilitud *negativa* (véase el pie de página 3 de la página 26), que toma el valor de 3130,023. Por tanto, el valor de la logverosimilitud alcanzada tras la optimización en este caso es

$$-3130,023. \tag{4.18}$$

Es decir, se ha vuelto a conseguir a aumentar la verosimilitud al pasar del modelo con la cola Pareto ordinaria al modelo con la cola Pareto generalizada².

En la Figura 4.1 se representa el histograma parcial de los datos sobre los siniestros derivados de grandes incendios en Dinamarca de la muestra del test, concretamente, el correspondiente a las pérdidas inferiores a 20 millones, junto con la función de densidad del modelo de Scollnik con cola GPD ajustado, esto es, utilizando los valores (4.17).

En la Figura 4.2 se han trazado la función de distribución empírica de esta submuestra dedicada a la validación en línea negra sólida junto con la función de distribución del modelo ajustado, utilizando los valores (4.17), en línea roja discontinua.

Las Figuras 4.1 y 4.2, similarmente a las obtenidas en los Capítulos 2 y 3, parecen indicar un ajuste apropiado.

4.5.2. Ajuste del modelo con restricciones de continuidad y diferenciabilidad de segundo orden

Si el modelo (4.6) se considera también con la condición de diferenciabilidad de segundo orden, esto es, para que verifique

$$\lim_{x \rightarrow \theta^-} f''(x) = \lim_{x \rightarrow \theta^+} f''(x),$$

²Recuérdese que el valor de la función de verosimilitud maximizada para el modelo con la cola Pareto ordinaria era de $-3133,858$.

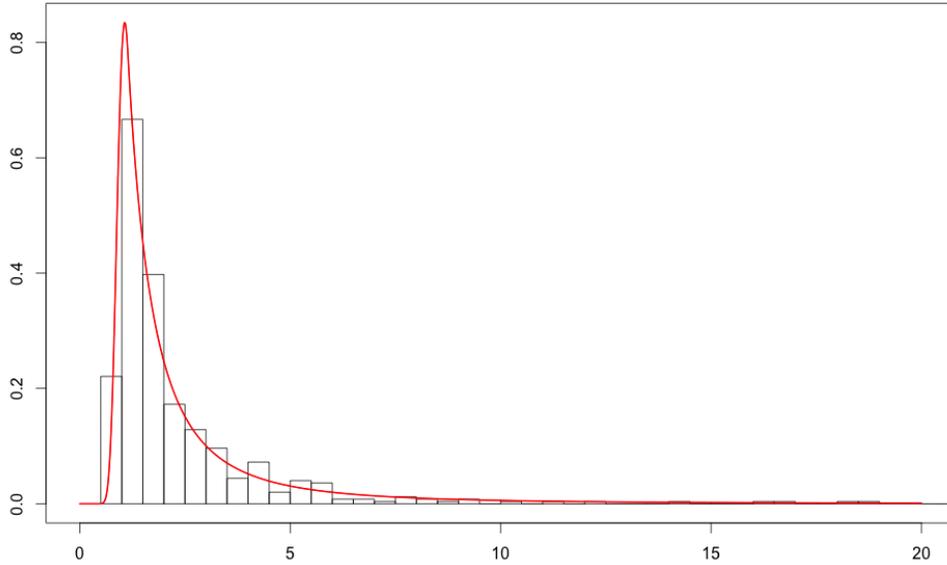


Figura 4.1: Histograma parcial correspondiente a valores inferiores a 20 millones de los incendios daneses con la función de densidad del modelo de Scollnik con cola GPD ajustado

además de la condición de continuidad y diferenciabilidad de primer orden, esto da lugar a la restricción (4.9). Tal y como se vio en la sección 4.2.3, esto da lugar a la siguiente fórmula para el parámetro de forma α como función de los parámetros θ , σ y λ , mostrada originalmente en la ecuación (4.11):

$$\alpha = \frac{1 - \frac{1}{\sigma^2} - \frac{\lambda}{\lambda + \theta} - \left(\frac{\theta}{\lambda + \theta}\right)^2}{\left(\frac{\theta}{\lambda + \theta}\right)^2 - \frac{\theta}{\lambda + \theta}}.$$

Con este valor calculado, la restricción de continuidad de primer orden da la siguiente regla para calcular el parámetro μ en función de los parámetros ya calculados:

$$\mu = \ln(\theta) - \left(\frac{\alpha\theta - \lambda}{\lambda + \theta}\right) \sigma^2.$$

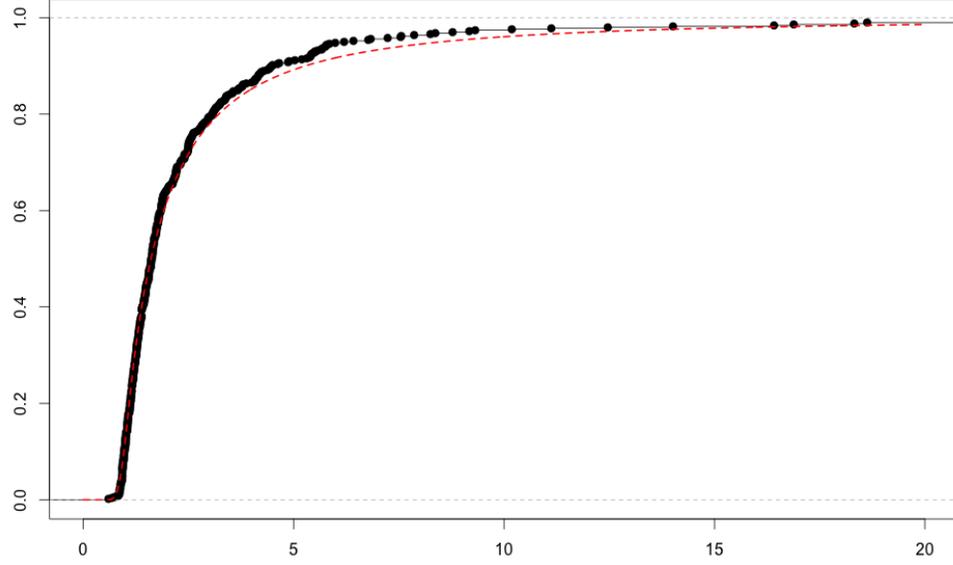


Figura 4.2: Función de distribución empírica (línea negra sólida) junto con la función de distribución del modelo de Scollnik con cola GPD ajustado (línea roja discontinua) en el tramo de los valores inferiores a 20 millones de los incendios daneses

Por último, se calcula el parámetro de peso r necesario para que haya continuidad:

$$r = \frac{\sqrt{2\pi}\alpha\theta\sigma\Phi\left(\frac{\ln(\theta)-\mu}{\sigma}\right)\exp\left\{\frac{1}{2}\left(\frac{\ln(\theta)-\mu}{\sigma}\right)^2\right\}}{\sqrt{2\pi}\alpha\theta\sigma\Phi\left(\frac{\ln(\theta)-\mu}{\sigma}\right)\exp\left\{\frac{1}{2}\left(\frac{\ln(\theta)-\mu}{\sigma}\right)^2\right\} + \lambda + \theta}.$$

Con estas tres restricciones de suavidad, el modelo (4.6) ve reducidos sus seis parámetros iniciales a tan sólo tres: los parámetros θ , σ y λ son parámetros libres de variar y tomar cualquier valor dentro de sus dominios de definición, concretamente, la semirrecta positiva $(0, \infty)$ en el caso de θ y σ y la semirrecta $(-\theta, \infty)$ en el caso de λ , mientras que los parámetros α , μ y r han de tomar un valor concreto que se expresa en función de los anteriores tres parámetros libres utilizando las mencionadas restricciones (4.11), (4.7) y (4.8).

A continuación se muestra el código de la función *R* `lkh2`, que contiene implementada la función de verosimilitud del modelo (4.6) sujeto a la restricción de diferenciabilidad de segundo orden.

```
lkh2=function (x) {
  theta=x[1]
  sigma=x[2]
  lambda=x[3]
  factorLambda=lambda/(lambda+theta)
  factorTheta=theta/(lambda+theta)
  numeradorAlpha=1-(1/sigma^2)-factorLambda-factorTheta^2
  denominadorAlpha=factorTheta^2-factorTheta
  alpha=numeradorAlpha/denominadorAlpha
  mu=log(theta)-((alpha*theta-lambda)/(lambda+theta))*sigma^2
  z=(log(theta)-mu)/sigma
  factor=sqrt(2*pi)*alpha*theta*sigma*pnorm(z)*exp(0.5*z*z)
  r=factor/(factor+lambda+theta)
  lkh=0
  for (i in 1:length(trainingData)) {
    xi=trainingData[i]
    if(xi <= theta){
      f=r*dlnorm(xi,meanlog = mu,sdlog = sigma)/
        plnorm(theta,meanlog = mu,sdlog = sigma)
      lkh=lkh+log(f)
    }else{
      gxi=(alpha*(lambda+theta)^alpha)/((lambda+xi)^(alpha+1))
      f=(1-r)*gxi
      lkh=lkh+log(f)
    }
  }
  return(-lkh)
}
```

La función `lkh2` de este modelo acepta como parámetro un vector `x` que ha de tener tres componentes: `theta`, `sigma` y `lambda`, representando, como es natural, a los parámetros θ , σ y λ del modelo (4.6).

Así, con las variables `theta`, `sigma` y `lambda` representando θ , σ y λ se calculan las variables `alpha`, `mu` y `r`, que naturalmente representan a los

parámetros α , μ y r usando el razonamiento y las ecuaciones anteriormente expuestas en esta sección, obteniendo todo lo necesario para calcular la función de verosimilitud para los datos daneses con la ayuda de la densidad (4.6).

En vista de que ahora el parámetro α no es libre de variar, sino que depende de los valores de los parámetros θ , σ y λ , el vector \mathbf{x} queda modificado, eliminándose de su definición la variable `alpha` con sus respectivas cotas y estimación inicial. Esta nueva construcción se almacena en el vector `x2`:

```
x2 <- list(label=c("theta","sigma","lambda"),est=c(1,1,1),
           low=c(0.1,0.1,0.1),upp=c(10,10,10))
```

En este caso, la salida parcial obtenida tras ejecutar la función `dfp` es la siguiente:

```
## Optimization Result:
## iter: 15  fmin: 3583.105  nfcn: 290
##
##  label  estimate  low      high
## 1 theta  8.3304823  4.6884869  9.6601801
## 2 sigma  0.74385801 0.72114425 0.76734262
## 3 lambda 8.3304902  7.8953638  8.6907647
```

Así, se deduce que los valores de los parámetros del modelo (4.6) tras maximizar la verosimilitud aplicada a los datos daneses son

$$\begin{aligned}\hat{\theta} &= 8,3304823 \\ \hat{\sigma} &= 0,74385801 \\ \hat{\lambda} &= 8,3304902\end{aligned}\tag{4.19}$$

siendo el intervalo de confianza al 95 % para el parámetro θ

$$[4,6884869; 9,6601801],$$

para el parámetro σ

$$[0,72114425; 0,76734262],$$

y para el parámetro λ

$$[7,8953638; 8,6907647].$$

Esto es, se tiene que

$$\begin{aligned}\mathbb{P}\{\theta \in [4,6884869; 9,6601801]\} &= \mathbb{P}\{\sigma \in [0,72114425; 0,76734262]\} \\ &= \mathbb{P}\{\lambda \in [7,8953638; 8,6907647]\} = 95\%.\end{aligned}$$

De nuevo, otro valor interesante que aparece en la última salida es la logverosimilitud *negativa* (véase el pie de página 3 de la página 26), que toma el valor de 3583,105. Por tanto, el valor de la logverosimilitud alcanzada tras la optimización en este caso es

$$-3583,105. \tag{4.20}$$

Por tanto, con este modelo la verosimilitud *decae* bruscamente en comparación con el resto de los modelos considerados en este trabajo. Recuérdese que el valor de la función de logverosimilitud maximizada para el modelo del Capítulo 2 era $-3144,059$, para el modelo del Capítulo 3 de $-3133,858$ y para el modelo (4.6) con cola GPD pero sin condición de diferenciabilidad de segundo orden es $-3130,023$.

Esto es, la suavidad de segundo orden en el modelo (4.6) lo lleva a tener la mínima verosimilitud de entre todos los modelos estudiados.

En la Figura 4.3 se representa el histograma parcial de los datos sobre los siniestros derivados de grandes incendios en Dinamarca de la muestra del test, concretamente, el correspondiente a las pérdidas inferiores a 20 millones, junto con la función de densidad del modelo de Scollnik ajustado con cola GPD y restricción de diferenciabilidad de segundo orden, esto es, utilizando los valores (4.19).

En la Figura 4.4 se han trazado la función de distribución empírica de esta muestra dedicada a la validación en línea negra sólida junto con la función de distribución del modelo ajustado, utilizando los valores (4.19), en línea roja discontinua.

Las Figuras 4.3 y 4.4, indican algo ya adelantado por el valor de la función de verosimilitud del modelo (4.6) con la condición de diferenciabilidad de segundo orden en comparación con las funciones de verosimilitud maximizadas de los demás modelos: el ajuste del modelo con condición de suavidad de segundo orden parece ofrecer el peor ajuste entre los considerados.

El modelo subestima la probabilidad en todo el entorno de la moda empírica –véase la Figura 4.3– asignando probabilidades que parecen valer la mitad de los valores observados. Por otro lado, alejándose de la moda, el

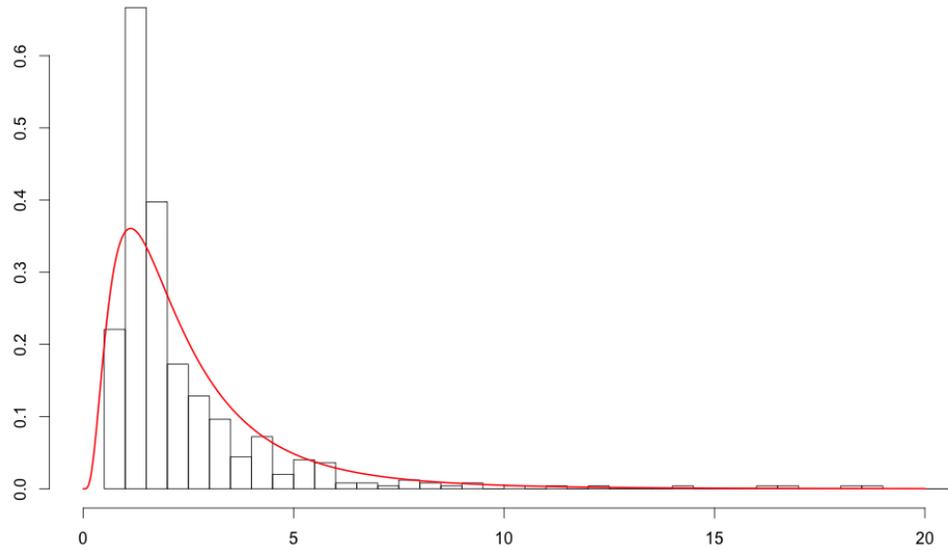


Figura 4.3: Histograma parcial correspondiente a valores inferiores a 20 millones de los incendios daneses con la función de densidad del modelo de Scollnik ajustado con cola GPD y restricción de diferenciabilidad de segundo orden

modelo sobreestima las probabilidades observadas, independientemente de si son valores inferiores o superiores a la moda.

Este comportamiento se puede apreciar también en la Figura 4.4 en la que se aprecia como el modelo primero sobreestima la probabilidad acumulada, luego la sobreestima –aproximadamente hasta el valor de la variable aleatoria $x = 5$ –, para luego volver a sobreestimarla, provocando que no se capture correctamente todo el riesgo subyacente en la cola de los datos empíricos.

4.6. Gráficos QQ

En las secciones 2.13 y 2.14 se realizaron pruebas χ^2 y G al ajuste del modelo de Cooray & Ananda que indicaban un ajuste estadísticamente significativo.

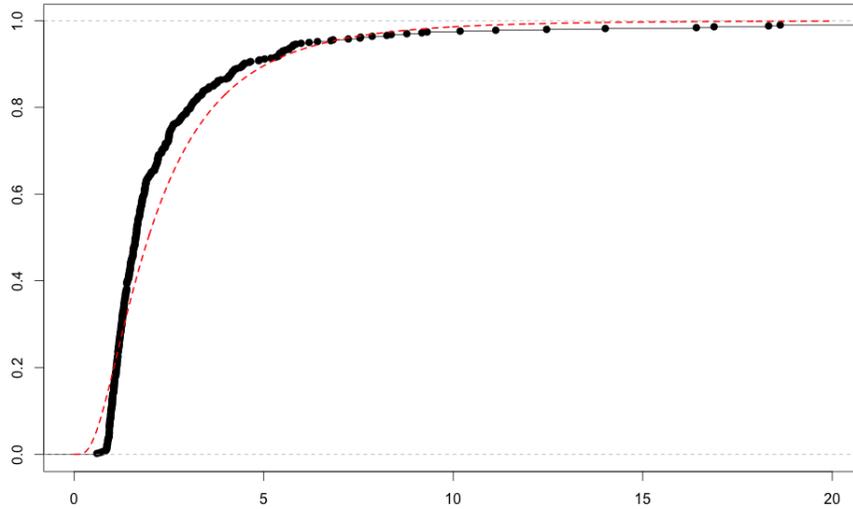


Figura 4.4: Función de distribución empírica (línea negra sólida) junto con la función de distribución del modelo de Scollnik ajustado con cola GPD y restricción de diferenciabilidad de segundo orden (línea roja discontinua) en el tramo de los valores inferiores a 20 millones de los incendios daneses

Sin embargo, en la sección 3.6 al evaluar la significatividad del modelo de Scollnik se descubrió la alta sensibilidad de estas dos pruebas $-\chi^2$ y $G-$ a la elección inicial de la partición del intervalo que contiene a los datos.

En vista de esta sensibilidad, se prescinde de estas pruebas para evaluar el ajuste de los modelos (4.6) con cola GPD. En esta sección se presentan los gráficos QQ, introducidos anteriormente en la sección 2.15, para los ajustes que dieron lugar a los valores de los parámetros estimados (4.17) y (4.19), dejando las pruebas analíticas para la sección 4.7 posterior.

4.6.1. Modelo con restricciones de continuidad y diferenciabilidad de primer orden

En la Figura 4.5 se presenta el gráfico QQ de los datos daneses frente al modelo (4.6) de Scollnik con cola GPD y restricción de diferenciabilidad de primer orden ajustado. Los cuantiles teóricos fueron calculados de acuerdo a

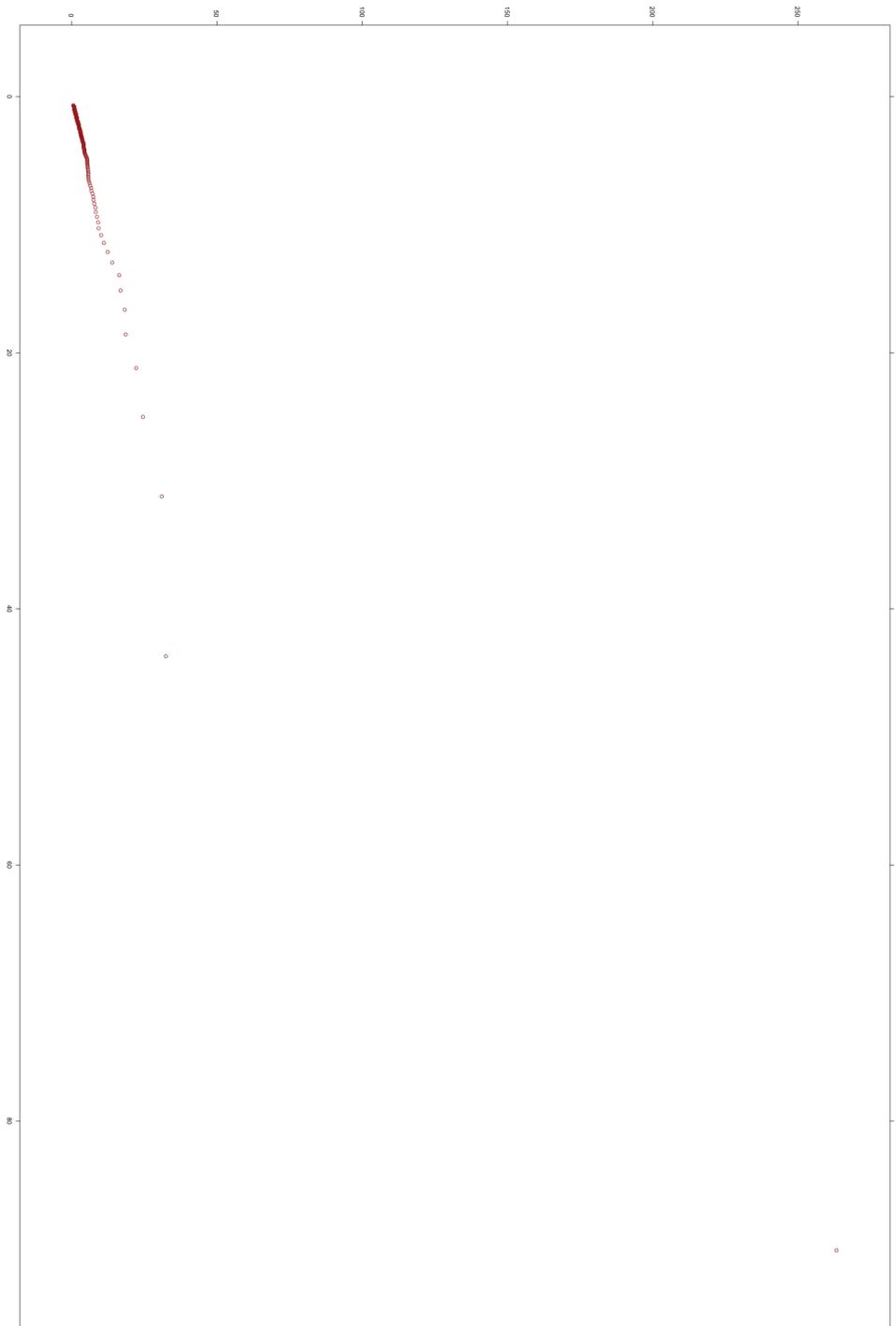


Figura 4.5: Gráfico QQ de los datos daneses frente al modelo de Scollnik ajustado con cola GPD y restricción de diferenciabilidad de primer orden

la función cuantil (4.16) obtenida en la sección 4.4, y los empíricos usando la definición 7 de [Hyndman & Fan, 1996] explicada en 2.10.

Igual que en los casos de los gráficos QQ de las Figuras 2.10 y 3.11, se puede apreciar que el ajuste es satisfactorio bajo este criterio gráfico: la curva paramétrica de los cuantiles teóricos frente a los empíricos se distribuye en –aproximadamente– una línea recta, hasta llegar al valor extremo de la esquina del gráfico.

4.6.2. Modelo con restricciones de continuidad y diferenciabilidad de segundo orden

La Figura 4.6 consiste en el gráfico QQ de los datos daneses frente al modelo (4.6) de Scollnik con cola GPD y restricción de diferenciabilidad de segundo orden ajustado. Análogamente al caso anterior, los cuantiles teóricos fueron calculados de acuerdo a la función cuantil (4.16) obtenida en la sección 4.4, y los empíricos usando la definición 7 de [Hyndman & Fan, 1996] explicada en 2.10.

En este caso, el gráfico QQ indica lo sugerido anteriormente con el valor de la función de verosimilitud maximizada y las Figuras 4.3 y 4.4: el ajuste del modelo con la condición de suavidad de segundo orden ofrece un ajuste muy pobre, tal y como indica la curvatura del gráfico en los valores más altos.

4.7. Ratios de verosimilitud y AIC

4.7.1. Ratios de verosimilitud

Ratio de verosimilitud para la significatividad de la cola GPD

En la sección 4.1 se justificó el nombre de Pareto *generalizada* dado a la distribución (4.1) en el sentido de que se mostró que si el parámetro λ tomaba el valor 0 en la densidad (4.5) de esta distribución de Pareto generalizada, esto daba lugar a la densidad (2.4) de la distribución de Pareto ordinaria.

Puesto que las cabezas de las distribuciones (3.5) y (4.6) son iguales –una distribución lognormal– se tiene que el modelo de Scollnik con cola GPD y restricción de diferenciabilidad de primer orden engloba el modelo de Scollnik con cola Pareto ordinaria, pudiendo aplicarse en este caso la teoría expuesta en la sección 3.8 sobre el ratio de verosimilitudes.

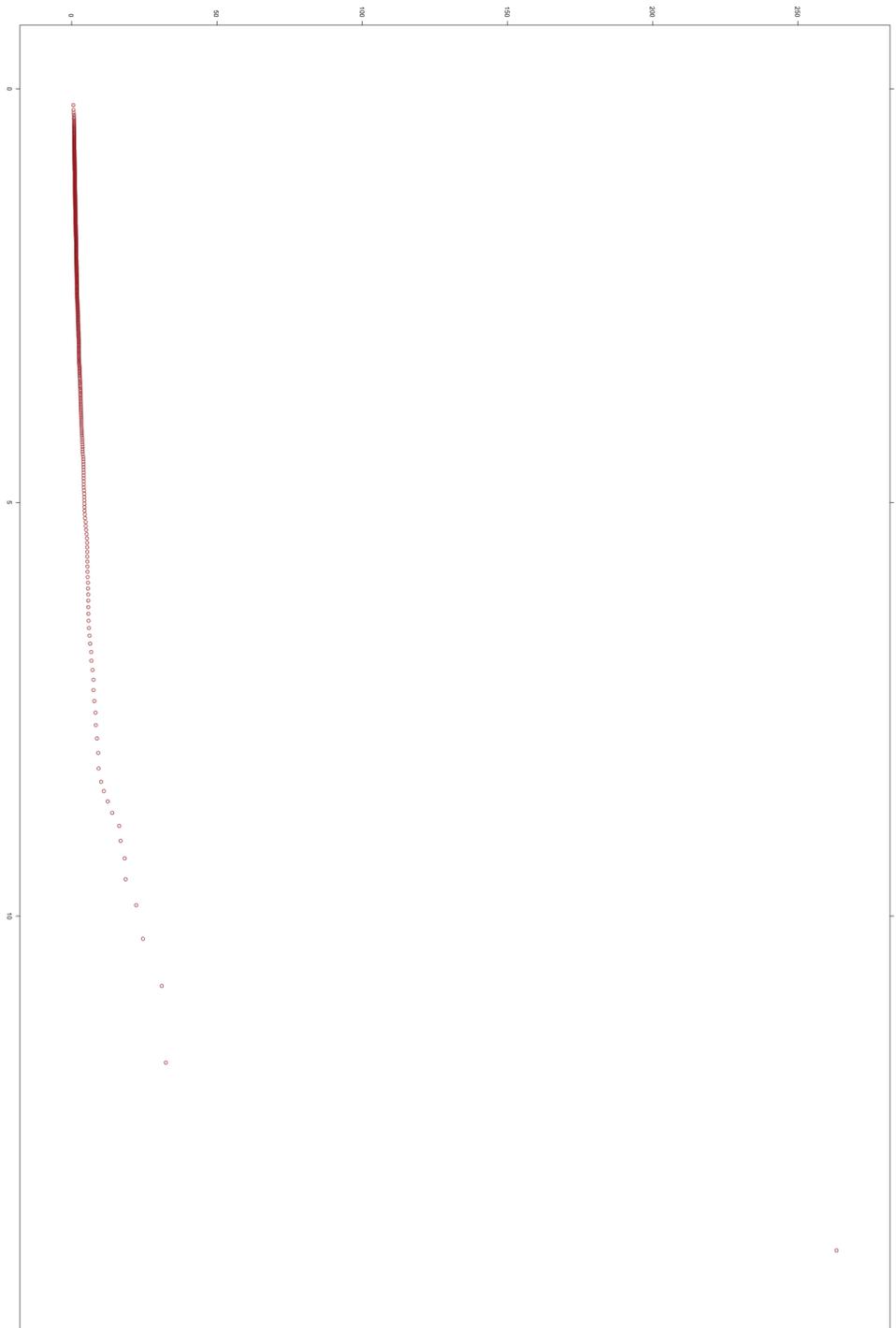


Figura 4.6: Gráfico QQ de los datos daneses frente al modelo de Scollnik ajustado con cola GPD y restricción de diferenciability de segundo orden

En (3.33) de la sección 3.5 se obtuvo que la logverosimilitud obtenida con el modelo de Scollnik con cola Pareto ordinaria es

$$L_0 = -3133,858,$$

mientras que la logverosimilitud proporcionada por el modelo de Scollnik con cola GPD y restricción de diferenciabilidad de primer orden fue mostrada en (4.18) de la sección 4.5.1, siendo su valor de

$$L_a = -3130,023.$$

Así, es inmediato calcular el estadístico D de para este caso:

$$D = 2(-3130,023 + 3133,858) = 7,67. \quad (4.21)$$

Puesto que la diferencia entre el número de parámetros de los modelos de Scollnik con cola GPD y cola Pareto ordinaria es 1, con la instrucción siguiente se obtiene el p-valor para la cantidad (4.21) del estadístico D .

```
pchisq(7.67,df=1,lower.tail = FALSE)
## [1] 0.005614646
```

Tal y como se puede apreciar, se tiene evidencia estadística suficiente a cualquier nivel de significación usual para preferir el modelo de Scollnik con cola GPD frente al modelo del mismo autor con cola Pareto ordinaria.

Ratio de verosimilitud para la significatividad del modelo con restricción de diferenciabilidad de segundo orden

Similarmente, el modelo con cola GPD y restricción de continuidad de segundo orden es un caso particular del modelo con cola GPD y restricción de orden uno: se trata pues de dos modelos anidados, y se puede volver a usar la teoría sobre el ratio de verosimilitud para testear la significatividad del modelo.

Como se acaba de recordar, la ecuación (4.18) proporciona la logverosimilitud alcanzada por el modelo de Scollnik con cola GPD y restricción de diferenciabilidad de primer orden:

$$L_a = -3130,023.$$

Para la logverosimilitud del modelo con cola GPD y restricción de continuidad de segundo orden hay que volver a la ecuación (4.20) de la sección 4.5.2, obteniendo

$$L_0 = -3583,105.$$

Por tanto, el estadístico D ahora es:

$$D = 2(-3130,023 + 3583,105) = 906,164. \quad (4.22)$$

Puesto que la diferencia entre el número de parámetros de los modelos considerados en esta prueba del ratio de verosimilitud es 1^3 , con la instrucción siguiente se obtiene el p-valor para la cantidad (4.22) del estadístico D .

```
pchisq(906.164,df=1,lower.tail = FALSE)
## [1] 4.485866e-199
```

Como cabría esperar después de los análisis gráficos del modelo con restricción de diferenciabilidad de segundo orden 4.3, 4.4 y 4.6, y directamente del bajo valor de la logverosimilitud alcanzada con este modelo, el test de Wilks hace rechazar la validez del modelo con la condición de suavidad de segundo orden en favor del modelo con cola GPD solamente con restricciones de primer orden.

4.7.2. AIC

De manera similar a lo realizado en la sección 3.9, es posible calcular los valores de la medida AIC para los modelos con cola GPD, tanto con la restricción de diferenciabilidad de primer orden como con la restricción de segundo orden.

Concretamente, en vista de que logverosimilitud del modelo sin suavidad de segundo orden es $-3130,023$ y que este modelo tiene 4 parámetros, el valor del criterio para este ajuste queda del siguiente modo:

$$AIC_{\text{GPD I}} = 2 \cdot 4 - 2(-3130,023) = 6268,046.$$

³La diferencia entre ambos modelos es que en el modelo con la restricción de suavidad de segundo orden el valor del parámetro α está fijado en función de los valores de los demás parámetros tal y como se puede apreciar en la ecuación (4.11), reduciéndose por tanto en 1 el número de parámetros efectivos con respecto al modelo sin la restricción de diferenciabilidad de segundo orden.

Análogamente, el valor de esta medida para el modelo con la restricción de suavidad de segundo orden, que tiene 3 parámetros y alcanza un valor de la logverosimilitud maximizada de $-3583,105$, queda como

$$AIC_{\text{GPD II}} = 2 \cdot 3 - 2(-3583,105) = 7172,21.$$

Puesto que se verifica $AIC_{\text{GPD I}} < AIC_{\text{S}}$, se tiene que el modelo con cola GPD y restricción de diferenciabilidad de primer orden es preferible al modelo con cola Pareto ordinaria.

A su vez, en vista de que $AIC_{\text{GPD II}} > AIC_{\text{C}}$, se tiene que el modelo con cola GPD y restricción de diferenciabilidad de segundo orden es el modelo menos satisfactorio según el criterio AIC, quedando todos los valores de esta medida de bondad de ajuste del siguiente modo:

$$AIC_{\text{GPD I}} < AIC_{\text{S}} < AIC_{\text{C}} < AIC_{\text{GPD II}},$$

o lo que es lo mismo, siguiendo este criterio los modelos estudiados quedan en el siguiente orden decreciente de preferencia:

1. Modelo con cola GPD y restricción de diferenciabilidad de primer orden;
2. Modelo con cola Pareto ordinaria y pesos variables;
3. Modelo con cola Pareto ordinaria y pesos fijos; y
4. Modelo con cola GPD y restricción de diferenciabilidad de segundo orden.

Capítulo 5

Conclusiones y futuros desarrollos

En este trabajo se ha introducido y analizado el conjunto de datos de los siniestros derivados de los incendios daneses, clásico en la literatura actuarial. Durante el análisis preliminar de estos datos, se han presentado algunas nociones básicas necesarias para llevar cabo este análisis exploratorio.

Con el objetivo de encontrar un modelo teórico que explique estos datos, se han introducido y desarrollado en profundidad varios modelos consistentes en concatenar una distribución lognormal con otras distribuciones con colas más pesadas. La razón detrás de usar la distribución lognormal para la cabeza de los datos es que es una distribución que prueba ser muy efectiva para modelizar los siniestros de cuantías moderadas, aunque la cola de esta distribución es demasiado ligera como para capturar correctamente el comportamiento de los siniestros más graves observados en este conjunto de datos de los incendios daneses, en particular, y de los presentes en el sector asegurador, en general.

Concretamente, se ha empezado desarrollando un modelo basado en una cola Pareto ordinaria, sujeto a restricciones de continuidad y diferenciabilidad en el punto de unión de las dos funciones de densidad. Tras el estudio del mismo, se ajustó este modelo a los datos daneses por máxima verosimilitud, comprobando un ajuste satisfactorio. Se evaluó la calidad del ajuste mediante dos pruebas: la prueba χ^2 y la prueba G .

A pesar de este ajuste satisfactorio, el modelo presentaba una restricción relativa a los porcentajes de observaciones dentro de cada una de las dos componentes –lognormal y Pareto– que parece poco realista.

Para superar esta restricción, el modelo anterior se generalizó a uno en el que los porcentajes correspondientes a cada componente eran variables, conservándose las restricciones relativas a la continuidad y diferenciabilidad de primer orden. También se comprobó que es imposible conseguir una restricción de diferenciabilidad de orden dos. Tras la construcción y estudio del nuevo modelo, éste se ajustó a los mismos datos, observándose una mejora al pasar de un modelo a otro.

Para la evaluación de esta mejora se introdujo el test de Wilks del ratio de verosimilitudes y el criterio de información de Akaike. Estos tests indicaron la superioridad del nuevo modelo y su significatividad estadística frente al modelo inicial.

Al evaluar la bondad de ajuste de este segundo modelo se volvieron a usar las pruebas χ^2 y G , y se comprobó que son poco adecuados para juzgar la validez de modelos continuos en vista de su alta sensibilidad a la elección de la partición del rango de datos.

Por último, se cuestionó la validez de la asunción de la cola Pareto, a pesar de que haya probado ofrecer ajustes correctos. Se introdujo brevemente la Teoría del Valor Extremo, y en particular, el Teorema de Pickands–Balkema–de Haan, el cual establece la distribución de los excesos por encima de un umbral bajo unas ciertas hipótesis razonables, la cual recibe el nombre de distribución de Pareto generalizada.

Tras una breve discusión de esta distribución, incluyendo la justificación del nombre *Pareto generalizada*, viendo que esta distribución efectivamente incluye la distribución de Pareto ordinaria, se introducen los dos últimos modelos de este trabajo consistentes en concatenar la cabeza lognormal con la cola Pareto generalizada. El primero de ellos es sometido a las mismas restricciones de continuidad y diferenciabilidad que los modelos anteriores, y el segundo también a una condición de orden dos, buscando aún más suavidad en el punto de corte entre los dos modelos.

El primero de estos dos modelos probará ser el mejor entre los considerados, mientras que el segundo se convertirá en el peor al tratar de explicar los datos, a pesar de tener la misma cantidad de parámetros que el modelo con cola Pareto ordinaria y pesos variables –el segundo en ser introducido en este trabajo– y un parámetro más que el primer modelo formulado en el Capítulo 2.

El test del ratio de verosimilitud indica la significatividad del modelo con cola Pareto generalizada y restricción de diferenciabilidad de orden uno, haciéndolo preferible al modelo con cola Pareto ordinaria, al igual que el

criterio AIC que también se inclina en favor de este modelo. Ambos criterios indican la insuficiencia del modelo con la condición de segundo orden de diferenciabilidad.

En definitiva, en este trabajo se han construido varios modelos que son útiles para explicar un conjunto de datos tan complejo como es el de los incendios daneses utilizando un número bajo de parámetros, siendo el ajuste satisfactorio tanto en la cabeza como en la cola de la distribución. Uno de estos modelos permite hacer uso de la Teoría del Valor Extremo, al usar para la cola la distribución de Pareto generalizada.

Una ventaja de este modelo frente a las técnicas aplicadas usualmente al ajustar una distribución de Pareto generalizada es la *no* necesidad de estimar explícitamente el parámetro de corte θ , lo cual puede ser bastante complicado, sobre todo en las situaciones en las que aparecen los llamados “*horror Hill plots*” [Resnick, 2007] –gráficos del estimador de Hill [Hill, 1975] muy irregulares–, dejándose la elección del parámetro de corte al usual criterio de máxima verosimilitud. Puede ser interesante desarrollar un estudio que compare las estimaciones clásicas del parámetro de corte θ mediante gráficos de estimador de Hill frente a este enfoque de máxima verosimilitud.

Otro posible análisis interesante y útil a realizar sería la determinación mediante la simulación de la distribución del estadístico de Kolmogorov-Smirnov para evaluar la bondad de ajuste.

Con respecto a la propia estructura de los modelos estudiados en este trabajo, es posible cambiar la distribución lognormal de la cabeza por una distribución Weibull y realizar un estudio similar al expuesto aquí. Para una introducción de estos modelos veáse [Cooray, 2009; Scollnik & Sun, 2012]. Nótese que según Cooray [2009] desgraciadamente es imposible construir modelos análogos con cabezas de distribución Gamma o inversa Gaussiana.

Por último, algo que no ha tenido éxito en la literatura pero nos parece tremendamente interesante es el enfoque propuesto en [Frigessi, Haug & Rue, 2002] para mezclar las dos funciones de densidad. En vez de pasar de forma brusca, aún con las condiciones de suavidad, de una densidad a otra en el valor del parámetro de corte θ , Frigessi et al. proponen lo que llaman *mixture dinámicas*, que consiste en *mezclar* las dos densidades de un modo más suave, utilizando una función de peso variable en el parámetro x , de tal modo que en valores bajos de esta variable x la función de peso le dará mayor importancia a la densidad de la cabeza, y en valores altos de x , la componente de la cola tendrá cada vez más peso, de tal suerte que en el infinito la función de peso ha de valer 1, dándole todo el peso posible a la componente de la cola.

Bibliografía

- Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6), 716–723.
- Arlot, S. & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, 4, 40–79.
- Balkema, A. A. & De Haan, L. (1974). Residual life time at great age. *The Annals of probability*, 792–804.
- Cooray, K. (2009). The weibull–pareto composite family with applications to the analysis of unimodal failure rate data. *Communications in Statistics—Theory and Methods*, 38(11), 1901–1915.
- Cooray, K. & Ananda, M. M. (2005). Modeling actuarial data with a composite lognormal-pareto model. *Scandinavian Actuarial Journal*, 2005(5), 321–334.
- Davison, A. C. (2003). *Statistical models*, volume 11. Cambridge University Press.
- Davison, A. C. (2013). *SMPracticals: Practical for use with Davison (2003) Statistical Models*. R package version 1.4-2.
- Dutang, C., Goulet, V., & Pigeon, M. (2008). actuar: An r package for actuarial science. *Journal of Statistical Software*, 25(7), 38.
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling Extremal Events, volume 33 of Applications of Mathematics* (New York ed.). Springer-Verlag, Berlin.

- Frigessi, A., Haug, O., & Rue, H. (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5(3), 219–235.
- Group, P. . R. A. & Securities, C. (2009). *its: Irregular Time Series*. R package version 1.1.8.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3(5), 1163–1174.
- Hoey, J. (2012). The Two-Way Likelihood Ratio (G) Test and Comparison to Two-Way Chi Squared Test. *ArXiv e-prints*.
- Hyndman, R. J. & Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, 50(4), 361–365.
- Joanes, D. & Gill, C. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1), 183–189.
- Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). *Loss models: from data to decisions*, volume 715. John Wiley and Sons.
- Luebeck, G. & Meza, R. (2013). *Bhat: General likelihood exploration*. R package version 0.9-10.
- McDonald, J. (2014). *Handbook of Biological Statistics (3rd ed.)*. Baltimore, Maryland: Sparky House Publishing.
- McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN bulletin*, 27(01), 117–137.
- NIST/SEMATECH, Croarkin, C., Tobias, P., & Zey, C. (2001). *Engineering statistics handbook*.
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302), 157–175.

- Pickands, J. (1975). Statistical inference using extreme orderstatistics. *Annals of Statistics*, 3, 119–131.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.
- Scollnik, D. P. (2007). On composite lognormal-pareto models. *Scandinavian Actuarial Journal*, 2007(1), 20–33.
- Scollnik, D. P. & Sun, C. (2012). Modeling with weibull-pareto models. *North American Actuarial Journal*, 16(2), 260–272.
- Seal, H. L. (1980). Survival probabilities based on pareto claim distributions. *Astin Bulletin*, 11(01), 61–71.
- Teodorescu, S. (2010). On the truncated composite lognormal-pareto model. *Math Rep*, 12, 71–84.
- Vernic, R. & Teodorescu, S. (2009). On composite pareto models.
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1), 60–62.
- Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. *Supplement to the Journal of the Royal Statistical Society*, 1(2), 217–235.