

**Modelos estadísticos para valores
extremos y aplicaciones**

Statistical models for tails and applications

Isabel Serra Mochales

Director: Joan del Castillo Franquet

Novembre, 2013

*Doctorat en Matemàtiques
Departament de Matemàtiques
Universitat Autònoma de Barcelona*

Memòria presentada per
aspirar al grau de doctor
en Matemàtiques.

Certifico que la present memòria ha estat realitzada per Isabel Serra Mochales, sota la direcció del Dr. Joan del Castillo Franquet.

Bellaterra, Novembre de 2013.

Firmat: Dr. Joan del Castillo Franquet.



A la meva família

A veces, sonreír es la mejor forma de contribuir a cambiar el mundo.

José Luis Cortés

The statistician cannot excuse himself from the duty of getting his head clear on the principles of scientific inference, but equally no other thinking man can avoid a like obligation.

Ronald A. Fisher

Agraïments

Agraeixo a la Universitat Autònoma de Barcelona i al seu Departament de Matemàtiques per haver-me donat la oportunitat de dedicar-me quasi exclusivament a la recerca els darrers quatre anys i fer possible aquest treball. Tanmateix, i més personalment vull agrair al departament que sempre m'ha acollit fent que em senti com a casa. També a cadascun dels seus membres, que amb les seves complexes i divertides personalitats, et provoquen inquietuds diverses i no només en el món de les matemàtiques. Finalment, vull agrair a la meua família i amics, sense ells mai m'hagués preocupat la resposta a la pregunta: .això per què serveixi, dia a dia, no me n'adonaria de lo relatiu que és el concepte: problema matemàtic.

Bé, i ara que ja he fet els agraïments generals que em garanteixen que no em deixo a ningú, com que uns agraïments acostumen a ser més llargs, continuo. La veritat és que si penso en agrair aquest treball a algú, em ve al cap un munt de gent que ha contribuït fent d'aquesta etapa una de les millors de la meua vida. Del departament començaria per la Safont que em va contagiar el seu respecte per la estadística, la Mercè i les seves lliçons, i podria estar estona... Els *compis* del club de la *birra*, els companys de despatx (i passadís) des de la Consuelo i la Trabal que varen recolzar la meua decisió de fer la tesi en estadística o l'Alsedà que em va fer veure que podia fer-ho, fins a la persona que més anyoro en aquests darrers dies de tesi, el Seco. Podria estar estona parlant de companys i amics que han estat al meu costat. Per exemple, dia a dia les úniques distraccions venen dels

àpats. Amb qui penso primer sempre que arribo al despatx: l'Albert, tot i que de seguida acabem discutint. I amb qui m'encanta dinar: el Clop tot i que acabo tant tipa que m'agafa son. Qui sempre té un berenar a punt: el Ramos, tot i que si no deixem el menú xocolata acabarem com pilotetes. I a pocs llocs he rigut tant com en els barracons sentint converses diverses, tot i que algunes em feien dententa, el Gavira i el Giraldo parlant de topologia, el Bosa i Romero amb les seves excursions... I als qui sempre m'han mirat amb el convenciment que podria fer aquest treball: el Coll, la Mireia, el Miquel, el Gerard...

Ara el grup més proper d'aquests 4 anys, els nous amics i companys del món de l'estadística: el Barrera, la Sara, el David, amb la Laia i la petita Queralt, la Jalila i tot el grup de la UAB, el Valero, el Llorens, l'Anna, l'Ester i la resta del servei. En especial, al Pere Puig, que sense cap dubte, va ser la persona que em va despertar la inquietud per la estadística i el Joan del Castillo que me la ha dirigit.

Al Joan del Castillo vull dedicar-li un apartat sencer. És sens dubte a qui haig d'agrair més coses de la tesi. Li agraeixo el suport científic: m'ha ensenyat, m'ha guiat i m'ha motivat a buscar respostes,... i preguntes. I el personal, ja que m'ha ajudat a créixer en molts aspectes, controlant la meva tossudesca, minimitzant els meus despistes,... I a superar obstacles com el que amablement descriu per: "no ser Shakespeare".

La meva família ha estat fonamental per a aquest treball des de la meva àvia que sempre m'ha animat a estudiar i no deixar-me portar per feines que només m'aportin diners, fins al meu fillet que ha tornat relatiu tot el meu entorn i m'ha ensenyat a fixar-me en lo realment important. Però deixeu que agraeixi a tots els avis i àvies de l'Àlex per cuidar d'ell i deixar-me treballar tranquil·la. No em vull deixar als meus *amores*: la Anna per lo perfeccionista, en Quim perquè és el meu científic preferit, en Joan per lo pragmàtic i la meva nina Samanta per les converses que fan que mai vulgui deixar d'aprendre. I vull acabar amb els meus solets: Salva i Àlex, tot i que si fos per ells poca

tesi hagués fet, ja que sempre m'estan provocant perquè no faci altre cosa que estar amb ells: juguen, riuen... i no em puc resistir.

Índice general

<i>Agraïments</i>	XI
Índice general	XIV
Introducción	XVII
<i>Introduction</i>	XXII
1. Preliminares	1
1.1. Modelización estadística	1
1.1.1. Teoría de la verosimilitud	2
1.1.2. Modelos exponenciales	3
1.2. Teoría de los valores extremos	4
1.2.1. Primer teorema fundamental	5
1.2.2. Segundo teorema fundamental	8
1.3. Modelización estadística de los valores extremos	12
1.3.1. Estimación del índice de la cola y del umbral óptimo	13
1.3.2. Modelos para valores extremos	15
1.4. Análisis de riesgo	16
1.4.1. El riesgo en seguros y finanzas	16
1.4.2. Medidas de riesgo	18
2. Visión crítica de la modelización estadística de valores extremos	21

2.1.	Distribución de Pareto o <i>power-law</i> ?	22
2.2.	Sobre la estimación del umbral óptimo	26
2.3.	Revisión de los modelos para valores extremos	30
2.4.	Ejemplo de POT con enlace la LogNormal	34
2.4.1.	Datos de seguros: pérdidas por incendios en Dinamarca	35
2.4.2.	Análisis de los datos	35
2.4.3.	Modelización de los datos	36
2.4.4.	Consideraciones finales	38
2.5.	Revisión de la clasificación	39
3.	Especificación y estimación en el modelo GPD	41
3.1.	Cálculo del estimador de máxima verosimilitud	42
3.1.1.	Resumen de los métodos de estimación para la GPD	44
3.1.2.	Mala especificación en el modelo de GPD	48
3.2.	Sobre la existencia del estimador de máxima verosimilitud	51
3.3.	Enfoque metodológico	55
3.3.1.	Sobre datos del oleaje en Bilbao	58
3.3.2.	Sobre datos del río Nidd	61
4.	Familia gamma truncada completada	65
4.1.	La distribución colas de gamma completada	67
4.1.1.	FTG como familia exponencial	69
4.1.2.	Modelo FTG y Pareto como modelos para colas	71
4.2.	Herramientas estadísticas y MLE	75
4.2.1.	Momentos y función generatriz	75
4.2.2.	Generación de variables aleatorias con distribución FTG	76
4.2.3.	Estimador de máxima verosimilitud de la FTG	77
4.2.4.	Matriz de información de Fisher esperada	80
4.3.	Estudios numéricos	82
4.3.1.	Análisis de los ciclones tropicales	82
4.3.2.	Análisis de la distribución de pérdidas agregadas	84

5. Propuestas finales y conclusiones	91
5.1. Propuestas finales en modelización	91
5.1.1. Cálculo del índice de la cola	92
5.1.2. Selección del umbral óptimo	94
5.1.3. Modelos para colas	96
5.2. Conclusiones	101
5.3. <i>Conclusions</i>	103
Bibliografía	105

Introducción

En los últimos años han proliferado los escándalos financieros que han puesto de manifiesto irregularidades contables y abusos por parte de ciertos colectivos. Estas pérdidas acostumbran a ser de un valor escalofriante. Barings Bank, el banco más antiguo de Londres, fundado en 1762 cerró sus puertas tras las pérdidas ocasionadas por Nick Leeson que ascendían a 1.4 millardos de dólares.

Pérdidas igual de sorprendentes son causadas por la naturaleza. En la temporada de huracanes del 2005, el huracán Wilma alcanzó tierra en más de una ocasión causando unos daños que se estiman entre 18 y 22 mil millones de dólares, esto posicionó a Wilma entre los 10 huracanes más costosos del Atlántico. Eventos extremos de fenómenos naturales se observan con demasiada frecuencia y nos preguntamos si son valores anómalos o son el resultado normal de años de evolución, de hecho, el cambio climático es un tema de actualidad con repercusiones catastróficas para nuestro entorno. Refiriéndose a la temperatura de la Tierra actual, el climatólogo Shaun Marcott dijo: *Un pico de calor como este no había ocurrido antes, al menos no en los últimos 11.300 años.*

Frases como estas las oímos a diario puesto que no es de extrañar que los valores u observaciones extremas y posiblemente anómalas nos llamen la atención. Los valores extremos los hallamos en muchos ámbitos de las ciencias y su modelización se utiliza en varios campos tales como la hidrología,

los seguros, las finanzas y la ciencia medio ambiental, ver Furlan (2010), Coles&Sparks (2006), Moscadelli (2004). La singularidad de los valores extremos hace que debamos tratarlos de un modo separado al resto de datos que observamos.

En general, cuando construimos modelos a partir de unos datos observados, deberíamos detectar si estos datos presentan valores extremos que deban ser tenidos en cuenta. Por ejemplo, a menudo cuando observamos datos financieros, a nivel descriptivo observamos valores extremos que deben tratarse por separado, porque su distribución es heterogénea con el resto de los datos o porque nuestro interés en modelar los datos radica en estos valores y por tanto queremos prestar especial interés en ellos.

El objeto que analiza los valores extremos desde un punto de vista estadístico son las colas de las distribuciones sobre un umbral (que simplemente llamaremos colas). Generalmente, las colas hacen referencia a aquello que puede suceder una vez de cada mil, en contraposición a la estadística habitual que se fija como mucho en lo que sucede una de cada 20 o 100 veces.

La teoría de valores extremos (EVT) tomó importancia en los años 20 con problemas relacionados principalmente con la hidrología y dieron lugar al primer teorema fundamental en EVT de Fisher-Tippet (1928) y Gnedenko (1948) que caracteriza la distribución asintótica del máximo observado. Otro punto de vista surgió en los años 70 con el segundo teorema fundamental de EVT de Pickands (1975) y Balkema-de Haan (1974) cuando todo parecía resuelto. Este resultado caracteriza la distribución asintótica de las colas como una distribución de la familia Pareto generalizada (GPD). A partir de estos resultados, podemos clasificar una cola por su índice, que denotaremos por ξ y toma valores reales. A partir de este resultado el modelo GPD se consideró el modelo de referencia para modelar colas, ver McNeil *et al.* (2005), Finkenstadt&Rootzén (2003), Coles (2001) y Embrechts *et al.* (1997). Además según este valor clasificamos las colas en ligeras ($\xi < 0$), exponencia-

les o normales ($\xi = 0$) y colas pesadas en otro caso. A partir de aquí, la teoría de valores extremos ha seguido evolucionando y a su vez, a menudo se aparta de las necesidades prácticas, de la modelización estadística, ver Diebold *et al.* (1998).

Actualmente, los ámbitos que presentan más problemas relacionados con valores extremos se clasifican según dónde deriva el riesgo que producen: en el ámbito financiero, en el ámbito medio-ambiental o en el ámbito de la salud. En este trabajo trataremos aplicaciones prácticas en los dos primeros ámbitos.

Últimamente, las herramientas, técnicas y procesos que se utilizan en modelización estadística de valores extremos se están cuestionando, puesto que desde un punto de vista práctico se aprecian limitaciones. Además, el hecho que la GPD caracterice la distribución de una cola ha hecho que este modelo se considere el modelo de referencia, cuando en realidad este modelo en ocasiones produce resultados poco satisfactorios, Dutta&Perry(2006).

Llegado este punto ya podemos enumerar los retos principales de la modelización estadística de los valores extremos. En primer y segundo lugar, la estimación del índice de la cola así como la estimación del umbral óptimo donde enlazar con el modelo GPD. I en tercer lugar, hallar modelos alternativos a la GPD que den resultados satisfactorios. En Coles (2001), Embrechts *et al.* (1997), McNeil *et al.*(2005) y Beirlant *et al.* (2004), hallamos revisiones satisfactorias de estos puntos clave en modelización estadística, pero aún así y como veremos en este trabajo, todavía hay trabajo que hacer.

Este trabajo está dividido en 5 Capítulos. El primero introduciremos algunos preliminares básicos. El Capítulo 2 revisaremos el estado de la modelización estadística de valores extremos de un modo crítico. En esta revisión vamos a mostrar que el problema de estimación de parámetros de la GPD es un obstáculo en el progreso de la modelización y por ello, trataremos este

tema en el Capítulo 3 en el cual hallaremos un nuevo enfoque del modelo que resolverá esta cuestión. De esta forma y con el trabajo de Castillo *et al.* sobre el coeficiente de variación residual podremos concluir en el Capítulo 5 con un protocolo de estimación del umbral óptimo y del índice de la cola que es satisfactorio, manejable y más riguroso, desde un punto de vista teórico, que otros métodos que se usan habitualmente. El reto de hallar nuevos modelos para colas es iniciado en el Capítulo 4 dónde presentaremos un modelo analítico nuevo que nos permitirá fijar los criterios para decidir si un modelo es apto para modelar colas. Finalmente, en el Capítulo 5 hallaremos las conclusiones generales de este trabajo.

El Capítulo 2 contiene una revisión crítica de los principales métodos y herramientas que se utilizan para afrontar los retos de la modelización estadística. En esta revisión, vamos a proponer el uso del coeficiente de variación como herramienta de metodologías alternativas a las clásicas de estimación del umbral y el índice de la cola. La segunda propuesta se basa en modelos para colas que utilicen la metodología clásica del *peaks-over-threshold* (POT), pero en lugar de enlazar con el modelo GPD usar modelos alternativos. Veremos un ejemplo utilizando la Log-Normal truncada. Finalmente, en esta revisión destacaremos el hecho de que a nivel práctico las colas con decaimiento exponencial son muy habituales y, a menudo, no son detectadas.

La motivación para el trabajo expuesto en el Capítulo 3 reside en el hecho de que, a pesar de sus debilidades, necesitamos el MLE en muchos procedimientos de inferencia. Los ajustes de distintos modelos a los mismos datos se comparan con los criterios de información de Akaike y Bayesiano y con la prueba de razón de verosimilitud en modelos anidados, todos ellos basados en el MLE. También es común el uso de la MLE en bondad de ajuste como en Choulakian&Stephens (2001). A pesar de esto, el MLE para la GPD está cuestionado, ver Hosking&Wallis (1987), Castillo&Hadi (1997),

Zhang&Stephens (2009). En consecuencia para la EVT, la problemática del MLE se ha traducido en la aparición de otras metodologías de estimación del índice de la cola y elección del umbral que por su popularidad se han extendido en la literatura a pesar de sus debilidades. Vamos a tratar a fondo el origen del problema del cálculo del MLE y daremos un nuevo enfoque metodológico para utilizar el MLE.

En el Capítulo 4 vamos a presentar un nuevo modelo para colas no ligeras que denominamos, modelo de colas gamma completado, FTG. La FTG es una nueva familia que además de contener las colas de distribuciones gamma contiene nuevas distribuciones que no habían sido consideradas, así como la distribución Pareto y la distribución exponencial. Estas nuevas distribuciones serán el punto clave del modelo. La teoría de la probabilidad para la distribución FTG se verá en detalle y se aplicará en la modelización a dos conjuntos de datos. En resumen, este modelo nos permite hallar distribuciones tan cerca como queramos de la distribución de Pareto, pero todas ellas con cola exponencial lo cual es muy útil para modelar datos que muestran indicios de cola pesada pero que ya sea por el contexto o por evidencias descriptivas, deben ser ajustadas por colas exponenciales.

Por último, en el Capítulo 5 vamos a proponer unas condiciones para modelos para colas que son coherentes con lo que hemos observado en nuestra investigación. Además haremos un repaso del estado de la modelización estadística de los valores extremos a fin de este trabajo, dando lugar a posibles líneas de investigación futuras.

Introduction

In recent years have proliferated financial scandals that have revealed accounting irregularities and abuses by part of certain groups. These losses are usually of a value creepy. Barings Bank, the oldest bank in London, founded in 1762 closed its doors after losses produced by Nick Leeson which amounted to 1.4 billion dollars.

Equally surprising losses are caused by nature. In the 2005 hurricane season, Hurricane Wilma hit earth on more than one occasion caused some damage was estimated between 18 and 22 billion dollars, Wilma was positioned between the 10 costliest Atlantic hurricanes. Extreme events of natural phenomena are observed too often and we wonder if this values are anomalous or the normal result of years of evolution. In fact, climate change is an emerging issue with catastrophic consequences for our environment. Referring to Current Earth's temperature, Shaun Marcott climatologist said: A heat spike like this had not happened before, at least not the last 11,300 years.

Phrases like these we hear daily, since it is no a surprise that the extreme values or outliers, and possibly anomalous, us attract attention. The extreme values are in many fields of science and modeling is used in several fields such as hydrology , insurance, finance and environmental science, see Furlan (2010), Coles&Sparks (2006), Moscadelli (2004). The uniqueness of outliers makes that we must treat them in a separate mode to other observations.

In general, when we build models from data, we should detect if the data have extreme values that must be taken into account. For example, often when we look at financial data, to descriptive level are observed extreme values that should be treated separately, because their distribution is heterogeneous with the rest of the data or because our interest in modeling the data lies in these values.

The main object in analyzing the extreme values, from a statistical viewpoint, are the left truncated distributions or the distributions above thresholds (which are known as tails). Generally, the tails do reference to what can happen once in thousand times, it is in contrast with the usual statistical who is more dedicated to what happen once in 20 or 100 times.

The extreme value theory (EVT) became important in the '20s, from problems mainly related to hydrology and led to the first fundamental theorem in EVT by Fisher - Tippet (1928) and Gnedenko (1948) to characterizing the asymptotic distribution of the maximum in observed data. When everything seemed settled, another point of view emerged in the '70s with the second fundamental theorem in EVT by Pickands (1975) and Balkema -de Haan (1974). This result characterizes the asymptotic distribution of the tails as a distribution in the generalized Pareto family. From these results, we can classify a tail by its index, denoted by ξ that it takes real values. Moreover, this result have led to GPD model to be the reference model to model tails, see McNeil et al.(2005), Finkenstädt&Rootzén (2003), Coles (2001) and Embrechts et al. (1997). Further according to this value of index, the tails are classified in light tails ($\xi < 0$), exponential or normal tails ($\xi = 0$) and otherwise heavy tails. From here, the extreme value theory has continued to evolve and in turn, to often has departs from the practical needs of the modeling statistics, see Diebold et al. (1998).

Currently, the fields that they have more problems in extreme values are classified according to where the risk of occurrence is derived: in the financial

field, in the environmental field or the field of health. We are going to discuss applications practices in the first two areas.

Lately, tools, techniques and processes used in statistical modeling of extreme values are questioned, since from a practical point of view are limitations. Moreover, the fact that the GPD characterizes distribution of the tails has made this model is considered as the reference model, when in fact this model sometimes produces unsatisfactory results, Dutta&Perry (2006).

At this point, we can already list the main challenges in statistical modeling of extreme values. First and second are the estimation of the tail index and the optimal threshold to bind to the GPD model. Third, find alternative models to the GPD with satisfactory results. In Coles (2001), Embrechts et al. (1997), McNeil et al. (2005) and Beirlant et al. (2004), we find a satisfactory review of these points in statistical modeling.

This paper is divided into 5 Chapters . The first contains an introduction for some basic preliminaries . Chapter 2 will review the status of statistical modeling of extreme values in a critical way . In this review we will show that the problem of estimating GPD parameters is an obstacle in the progress of the modeling. Therefore in Chapter 3 we discuss and we find a new approach to solve this question. In this way and with the work of Castillo et al. (2013) on the residual coefficient of variation we conclude in Chapter 5 with a protocol to estimate the optimal threshold and the tail index, manageable and more rigorous , from a theoretic point of view than other commonly used methods. Moreover, the challenge of finding new models for tails is initiated in Chapter 4 where we present a new analytical model that will allow us establish the criteria for deciding whether a model is suitable for modeling tails. Finally , in Chapter 5 we find the conclusions general of this work.

Chapter 2 provides a critical review of the main methods and tools used to meet the challenges of the statistical modeling. In this review, we propose

the use of the coefficient of variation as an alternative methodological tool to estimate the optimal threshold and the index tail. The second proposal is based on models for tails consisting in to consider the classical methodology, peaks-over-threshold (POT), but instead of linking to the GPD model using alternatives models. We will see an example using the truncated log-normal. Finally, in this review we highlight the fact that into practice the tails with decay exponential are very common and often not detected.

The motivation for the work described in Chapter 3 resides in the fact that, despite its weaknesses, we need MLE in many procedures of inference. To compare the goodness-of-fit of different models for the same data uses Akaike and Bayesian information criterium or the likelihood ratio test, all based on the MLE. Also common is the use of the MLE in goodness of fit as Choulakian&Stephens (2001). Despite this, the MLE for GPD is questioned, see Hosking&Wallis (1987), Castillo&Hadi (1997), Zhang&Stephens (2009). Consequently, for the EVT, the MLE problem has resulted in the emergence of other methodologies for estimation tail index and choice of threshold that have spread by its popularity in the literature, despite its weaknesses. We will deal in depth with the origins of MLE calculation problem and we are going to give a new methodological approach.

In Chapter 4 we present a new model for no-light-tails, we call full truncated gamma, FTG. The FTG is a new family that contains the tails of gamma distributions, some new distributions that had not been considered and the Pareto distribution and the exponential distribution. These new distributions will be the key point of the model. The theory of probability to the FTG distribution is discussed in detail and is applied to the modeling of two sets of data. In summary, this model allows us to find distributions as close as we want to Pareto distribution, but all with exponential tail. This is very useful for modeling data which show signs of heavy tail but for context or for descriptive evidence, should be adjusted as exponential tail.

Finally, in Chapter 5 we will propose some conditions for queuing models that are consistent with what we observed in our research. In addition we will review the state of the statistical modeling of extreme values after of this work, leading to possible future research lines.

Capítulo 1

Preliminares

Los daños causados por el huracán Katrina registró 1833 muertes y la cifra total de daños materiales se estimó en un principio en 108 mil millones de dólares. Podríamos predecir la pérdida ocasionada por una catástrofe meteorológica? Jérôme Kerviel, trabajador de Soci t  G n rale desde 2008 estaf  5000 millones de euros. Podemos estar preparados para cubrir las p rdidas ocasionadas por un empleado estafador? La estad stica se ocupa de estas cuestiones al considerar una secuencia de observaciones como realizaciones de una secuencia de variables aleatorias, y utilizar los datos para estimar la estructura probabil stica de estas variables.

1.1. Modelizaci n estad stica

En este trabajo haremos  nfasis en diferenciar el papel que juega la teor a de la probabilidad y la modelizaci n estad stica en el an lisis de datos u observaciones aparentemente an malos por ser *valores extremos*. Nuestra herramienta de trabajo es la modelizaci n estad stica y esta tiene tres componentes: datos, modelo y un enlace entre estas dos.

Vamos a denotar indistintamente por

$$\mathbf{x} = \{x_1, \dots, x_n\} = \underline{x} \tag{1.1}$$

el conjunto de n , natural finito, observaciones independientes de una variable aleatoria (v.a.) real. Usaremos $\{X_1, \dots, X_n\}$ para denotar un conjunto de n variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.), de las que los datos (1.1) son una realización.

La modelización estadística requiere de modelos que pueden tomar formas muy variadas. En este trabajo un modelo estadístico consistirá en una familia

$$\mathcal{F} = \{f(x; \theta)\} \tag{1.2}$$

donde cada valor de $\theta \in \Theta \subset \mathbb{R}^k$ define una función de distribución (cdf) que denotaremos con letras mayúsculas, o sus correspondientes funciones de densidad (pdf) que denotaremos por letras minúsculas. Los parámetros de las distribuciones, generalmente, los denotamos por letras griegas.

Por último, dada una v.a., los valores extremos que son objeto de nuestro estudio son los valores que toma la variable por encima (o por debajo) de un umbral. Podemos estandarizar, con transformaciones sencillas, y sólo considerar valores extremos aquellos que se hallen por encima de un umbral y tomen valores positivos.

1.1.1. Teoría de la verosimilitud

La modelización estadística siempre que sea posible estará basada en la teoría de la verosimilitud desarrollada por Fisher (1922), la cual es encargada de enlazar datos y modelo. El enlace se describe por la función de verosimilitud

$$L(\theta; \mathbf{x}) \tag{1.3}$$

y se calibra el modelo hallando los parámetros de la familia que maximizan la función de verosimilitud. Estos se denominan los estimadores de máxima verosimilitud (MLE) y se denota por $\hat{\theta}$. El MLE proporciona una calibración que no depende de la parametrización del modelo. Por lo tanto podemos decir que hemos estimado la función de distribución. Esta propiedad de la teoría

de la verosimilitud hace que no deba ser relevada por otra teoría de enlace que no la cumpla.

La verosimilitud se utiliza también para comparar el ajuste de diversos modelos. Si tenemos un modelo y un submodelo se utiliza el test de razón de verosimilitud (LRT). En otros casos se utiliza el criterio de Akaike (AIC) o el Bayesian information criterion (BIC)

1.1.2. Modelos exponenciales

Hacia los años 20 Fisher introdujo entre sus muchos trabajos, la teoría de los modelos exponenciales. A pesar de la importancia de estos modelos apenas hallamos algunos libros que desarrollen metodologías sobre estos: Barndorff-Nielsen (1978), Brown (1986) y Letac (1992). La importancia de estos modelos está justificada por el hecho de que la mayoría de las familias de distribuciones que se acostumbra a utilizar en modelización estadística corresponden a modelos exponenciales.

Un modelo exponencial completo, generado por la medida de *Lebesgue* en $[0, \infty)$, con estadístico canónico $T(x)$ es el conjunto de todas las densidades

$$\exp(\theta \cdot T(x)) / C(\theta) \tag{1.4}$$

para $\theta \in D$, donde D denota el conjunto más grande de parámetros tales que la transformada de Laplace

$$C(\theta) = \int_0^\infty \exp(\theta \cdot T(x)) dx. \tag{1.5}$$

converge. En caso que este conjunto no sea maximal el modelo no será completo. D se denomina el *dominio natural de parámetros*. El *dominio de las medias* es la imagen del interior de D por la aplicación gradiente

$$\theta \mapsto \nabla k(\theta) \tag{1.6}$$

dónde $k(\theta) = \log C(\theta)$. Si D es un conjunto abierto, entonces las ecuaciones de verosimilitud tienen una única solución siempre que la media de las ob-



Figura 1.1: Barndorff-Nielsen.

servaciones transformadas por el estadístico canónico estén en el *dominio de las medias*, Barndorff-Nielsen (1978).

Dada una muestra $\mathbf{x} = \{x_1, \dots, x_n\}$ tal que el valor muestral del estadístico T , $t(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n T(x_i)$, está en el interior del dominio de las medias, entonces el estimador de máxima verosimilitud de la muestra está en el interior del dominio natural de parámetros.

Ejemplo 1.1.1 (Distribución LogNormal). Podemos definir la distribución LogNormal de parámetros 0 y 1, como el modelo exponencial completo generado por la medida $d\mu = dx/x$ en $[0, \infty)$, con estadístico canónico $T(x) = (\ln(x), \ln^2(x))$

1.2. Teoría de los valores extremos

La teoría de valores extremos (EVT) tomó importancia en los años 20 con problemas relacionados principalmente con la hidrología, por ejemplo, conocer el máximo oleaje en cierta región en el 99% de los casos. Este tipo de cuestiones quedaron en gran medida resueltos con el primer teorema fundamental en EVT de Fisher - Tippet (1928) y Gnedenko (1948). No obstante, quedaban otras cuestiones por resolver, por ejemplo, cuál era el oleaje



Figura 1.2: Maurice René Fréchet, Emil Julius Gumbel y Waloddi Weibull.

esperado en el peor 1 % de los casos. No fue hasta los años 70 con el segundo teorema fundamental de EVT de Pickands (1975) y Balkema - de Haan (1974) cuando estas cuestiones parecían haber quedado resueltas. Y de hecho a nivel teórico puede que sí, pero a la práctica veremos que no. Actualmente, problemas relacionados con datos de finanzas y seguros y problemas relacionados con el cambio climático no disponen de soluciones satisfactorias.

1.2.1. Primer teorema fundamental

El primer teorema fundamental de la EVT consiste en la caracterización de la distribución del estadístico de orden: máximo. Esta caracterización viene dada por la distribución de valores extremos generalizados (GEV).

Definición 1.2.1 (GEV: *generalized extreme value*). La función de distribución de la GEV estándar viene dada por

$$H(x; \xi) = \begin{cases} \exp(-(1 + \xi x)^{-1/\xi}), & \xi \neq 0, \\ \exp(-e^{-x}), & \xi = 0, \end{cases} \quad (1.7)$$

dónde $1 + \xi x > 0$. La familia tri-paramétrica GEV corresponde a la familia de distribuciones descritas por

$$H(x; \xi, \mu, \sigma) = H((x - \mu)/\sigma; \xi) \quad (1.8)$$

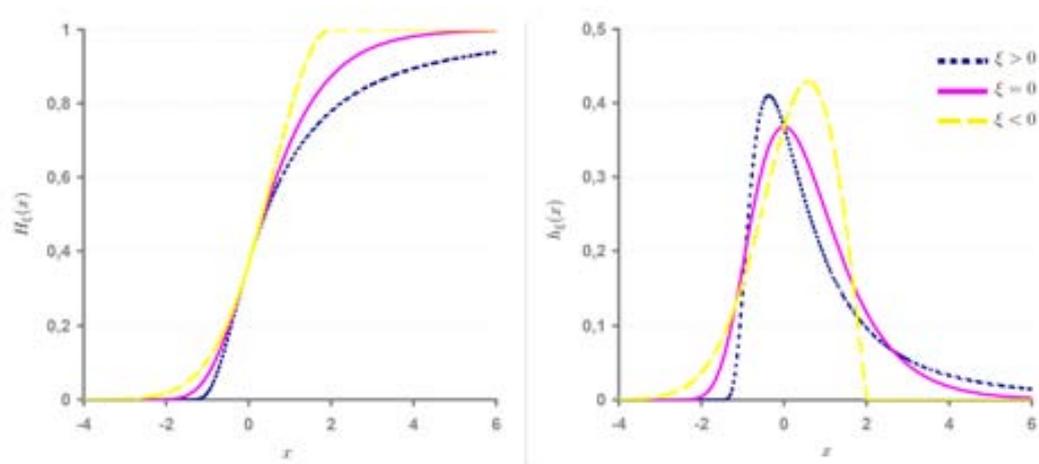


Figura 1.3: Funciones de densidad y funciones de distribución de la familia GEV estándares. Ejemplos de Fréchet ($\xi = 0,5$), Gumbel ($\xi = 0$) y Weibull ($\xi = -0,5$).

con $\mu \in \mathbb{R}$ un parámetro de localización y $\sigma > 0$ un parámetro de escala.

El parámetro ξ es conocido como parámetro de forma (*shape*). La GEV es la unificación de tres familias que se pueden identificar por el parámetro de forma: Fréchet (si $\xi > 0$), Gumbel (si $\xi = 0$) y Weibull (si $\xi < 0$).

Para x fijo,

$$\lim_{\xi \rightarrow 0} H(x; \xi) = H(x; 0),$$

lo cual facilita el uso de esta familia en modelización estadística. Este límite se cumple por ambos lados a pesar que para $\xi < 0$ corresponda a distribuciones con soporte compacto, ver la Figura 1.3.

El papel que juega la familia GEV en la teoría de los valores extremos es análogo al de la distribución normal en la teoría sobre el límite de suma de variables. Más concretamente, dado X_1, \dots, X_n v.a.i.i.d con distribución dada por F y sea $S_n = X_1 + \dots + X_n$ su suma, entonces la distribución normal estándar caracteriza la distribución de S_n (debídamente estandarizada) cuando n tiende a infinito. Si consideramos $M_n = \max(X_1, \dots, X_n)$ su



Figura 1.4: Sir Ronald Aylmer Fisher, Leonard Henry Caleb Tippett y Boris Vladimirovich Gnedenko.

bloque de máximos, entonces la distribución de GEV estándar caracteriza la distribución de M_n (debídamente estandarizada) cuando n tiende a infinito. Remarcamos que $P(M_n \leq x) = F^n(x)$.

Definición 1.2.2 (MDA). Sea F una cdf y M_n su bloque de máximos. Si existen sucesiones de números reales (d_n) i (c_n) , con $c_n > 0$ para todo $n \in \mathbb{N}$, tal que

$$\lim_{n \rightarrow \infty} P((M_n - d_n)/c_n \leq x) = \lim_{n \rightarrow \infty} F^n(c_n x + d_n) = H(x) \quad (1.9)$$

para alguna función de distribución no-degenerada (es decir, no concentrada en un punto), entonces decimos que F pertenece al máximo dominio de atracción de H , $F \in MDA(H)$.

Teorema 1.2.3 (Fisher-Tippet, Gnedenko). Sea F una función de distribución tal que $F \in MDA(H)$, entonces H puede estandarizarse de manera única (variando las sucesiones (d_n) i (c_n)) para obtener $H(\cdot; \xi)$ en la familia GEV estándar, para algún $\xi \in \mathbb{R}$.

Ejemplo 1.2.4. Consideramos la distribución exponencial, cuya función de distribución viene dada por

$$F(x) = 1 - \exp(-\beta x),$$

para $\beta > 0$ y $x \geq 0$. Si tomamos secuencias para estandarizar: $c_n = 1/\beta$ para todo n y $d_n = \ln n/\beta$, podemos calcular directamente la distribución límite:

$$\lim_{n \rightarrow \infty} F^n(c_n x + d_n) = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n} \exp(-x)\right)^n = \exp(-e^{-x}), \quad x \in \mathbb{R} \quad (1.10)$$

por lo tanto, concluimos que $F \in MDA(H_0)$.

1.2.2. Segundo teorema fundamental

Un segundo refinamiento de la teoría de los valores extremos parte de considerar no sólo el bloque máximo sino la distribución de todos los valores que exceden un umbral fijado. El mayor defecto del método de bloques máximos es la cantidad de datos que menospreciamos, puesto que sólo nos quedamos con los máximos. Es por esto que a nivel práctico es mucho más usado el método de los excesos de umbral, en el que está basado este segundo teorema fundamental. Del mismo modo que la GEV caracteriza los bloques máximos, la distribución generalizada de Pareto (GPD) caracteriza los excesos de umbral.

Definición 1.2.5 (GPD: generalized Pareto distribution). La distribución de Pareto generalizada (GPD) introducida por Pickands (1975) es la familia bi-paramétrica de funciones de distribución dada por

$$G(x; \xi, \psi) = \begin{cases} 1 - (1 + \xi x/\psi)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-x/\psi), & \xi = 0, \end{cases} \quad (1.11)$$

dónde $\xi \in \mathbb{R}$ y $\psi > 0$ son los parámetros de forma y escala, respectivamente. Para $\xi < 0$ el rango de x es $0 < x < -\psi/\xi$ y $x > 0$ para $\xi > 0$. Denotaremos por $g(x; \xi, \psi)$ los elementos de la correspondiente familia de funciones de densidad.

Observación 1.2.6. Para una v.a. X con función de distribución en la familia GPD podemos expresar la esperanza como $E(X) = \psi/(1 - \xi)$ y la varianza por $V(X) = \psi^2/((1 - \xi)^2(1 - 2\xi))$ siempre que $\xi < 0,5$, puesto que el r -ésimo momento es finito si i sólo si $1/\xi > r$.



Figura 1.5: Vilfredo Pareto.

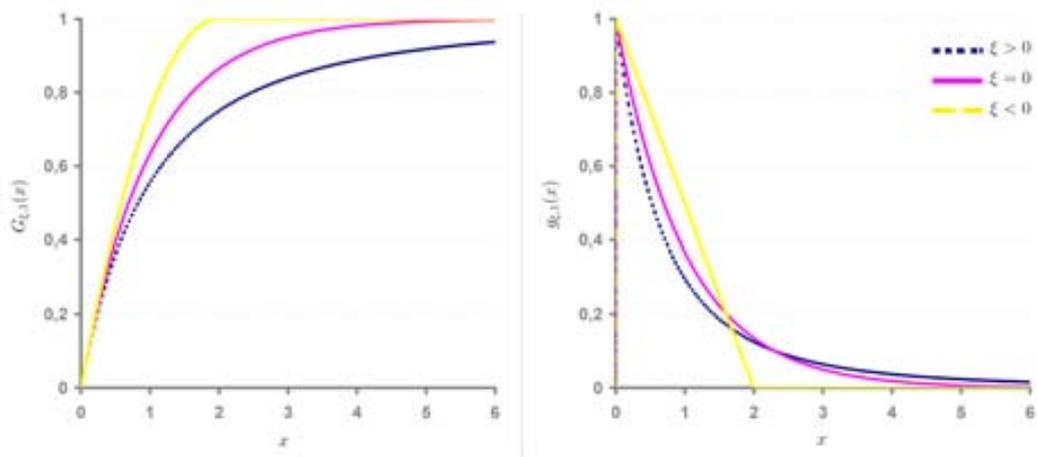


Figura 1.6: Densidades y distribuciones de GPD con parámetro de escala 1 y parámetros de forma: $\xi = 0,5$, $\xi = 0$ y $\xi = -0,5$.



Figura 1.7: James Pickands III, August Aimé Balkema y Laurens de Haan.

Definición 1.2.7 (función de distribución de los excesos). Sea X una v.a. con distribución F y sea $u > 0$ un umbral fijado. Denotamos X_u los valores trasladados a 0 de la variable X que exceden un umbral u

$$(X - u \mid X > u).$$

Y del mismo modo, dada una muestra $\mathbf{x} = \{x_1, \dots, x_n\}$, denotaremos por \mathbf{x}_u la muestra truncada en u y trasladada al 0,

$$\{x_i - u \mid x_i > u\}.$$

La distribución de X_u viene dada por

$$F_u(x) = P(X - u \leq x \mid X > u) = \frac{F(x + u) - F(u)}{1 - F(u)} \quad (1.12)$$

para $0 \leq x_F - u$, dónde $x_F \leq \infty$ denota el límite derecho de la distribución F .

Ejemplo 1.2.8. Si F es la función de distribución de una distribución exponencial, entonces se comprueba fácilmente que $F_u(x) = F(x)$ para todo $x > 0$ y todo $u > 0$, puesto que es la conocida ley de *falta de memoria* de dicha distribución.

Proposición 1.2.9. *Sea F una función de distribución de la familia GPD, es decir, $F = G(\cdot; \xi, \psi)$ para algún $\xi \in \mathbb{R}$ y $\psi > 0$, entonces*

$$F_u(x) = G(x; \xi, \psi + \xi u)$$

dónde $0 \leq x < \infty$ si $\xi \leq 0$ y $0 \leq x \leq -(\psi/\xi) - u$ si $\xi < 0$.

Observamos que la Proposición 1.2.9 nos muestra como la GPD es una familia cerrada bajo la transformación en distribución de los excesos.

Teorema 1.2.10 (Pickands - Balkema - de Haan). *Existe una función $\psi(u)$ tal que*

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G(x; \xi, \psi(u))| = 0,$$

si y sólo si $F \in MDA(H_\xi)$, $\xi \in \mathbb{R}$.

Por lo tanto, toda distribución tiene un umbral a partir del cual la distribución de sus excesos está tan cerca de una distribución GPD como queramos. Además, el parámetro de forma $\xi \in \mathbb{R}$, que caracteriza la GPD, queda fijado según la distribución inicial, ya que caracteriza la distribución (en forma estandarizada) de sus bloques de máximos.

Fijado un criterio para elegir que umbral cumple nuestros requisitos de ajuste de la GPD, lo definiremos como umbral óptimo. En general, el umbral óptimo será tomado como el umbral que garantice un ajuste del modelo GPD mejor que los modelos alternativos que se estén considerando. Estos modelos alternativos dependerán del contexto de los datos y serán comparados, si es posible con LRT y en otro caso con AIC o BIC.

Utilizaremos el concepto de *cola* para definir los excesos de un umbral fijado. Por el Teorema 1.2.10 están clasificadas según el valor del parámetro de forma que denominaremos índice de la cola. Si $\xi < 0$ diremos cola ligera (o cola de soporte compacto), si $\xi = 0$ diremos cola exponencial (o normal) y si $\xi > 0$ diremos cola pesada (o cola Pareto).

Por lo tanto, la clasificación principal de los valores extremos de una distribución, muestra o variable es la que queda determinada de forma unívoca por el tipo de cola.

1.3. Modelización estadística de los valores extremos

La modelización estadística de los valores extremos ha sido enriquecida durante muchos años por metodologías y técnicas provenientes de muchos ámbitos de la ciencia y las ciencias sociales.

Evidentemente, la modelización estadística de los valores extremos tiene casi tantas bifurcaciones como la modelización estadística ordinaria: modelización univariante, multivariante, con dependencia o sin ella, etc... pero en cualquier caso hay un ingrediente básico.

Este ingrediente imprescindible de la modelización de extremos tiene su analogía en la teoría estadística ordinaria. Del mismo modo que en la teoría estadística ordinaria utilizamos la esperanza y la varianza para describir gran parte de la distribución de una variable, en la estadística de los valores extremos utilizamos el índice de la cola. Remarcamos que el índice de la cola viene dado por el parámetro de forma, $\xi \in \mathbb{R}$ el cual caracteriza tanto los bloques de máximos como los excesos de un umbral, ver el Teorema 1.2.10. A modo genérico, la esperanza y la varianza disponen de buenos estimadores: la media y la varianza muestral; en cambio, el índice de la cola no dispone de un estadístico muestral de tan buena calidad, y este es el primer reto en la modelización estadística.

Es evidente, que el mejor modelo para los excesos de un umbral, desde un punto de vista teórico, es la GPD puesto que tenemos certeza matemática de que existe un umbral a partir del cual una distribución de la familia GPD

ajusta tan bien como queramos nuestros datos. No obstante, hallar el umbral óptimo a nivel práctico resulta otro reto relevante.

1.3.1. Estimación del índice de la cola y del umbral óptimo

Recalamos que hallar la forma, o MDA, de una familia de distribuciones resulta un problema más o menos viable y analítico, pero hallar la forma de los excesos de una muestra resulta mucho más complejo. Existen numerosos métodos pero cabe destacar el más utilizado y reformulado en la literatura de los valores extremos: el método Hill.

El método Hill

El método Hill es el método más aceptado para calcular el índice de la cola de unos datos. No obstante, cabe remarcar que este método está diseñado para estimar el índice de la cola en la subfamilia Fréchet.

Sean $X_{(1)}, \dots, X_{(n)}$ los estadísticos de orden, la forma estándar del estimador de Hill es

$$\hat{\xi}_{i,n}^{Hill} = \frac{1}{i} \sum_{j=1}^i \ln \frac{x_{(j)}}{x_{(i)}}, \quad 2 \leq i \leq n.$$

así, $\hat{\xi}_{i,n}^{Hill}$ estima el parámetro ξ que determina el índice de la cola. La estrategia general es representar el estimador de Hill para diferentes valores de i . Esto se llama el Hill-plot y corresponde a la representación del conjunto

$$\{ (i, \hat{\xi}_{i,n}^{Hill}) \mid i = 2 \dots n \}$$

La metodología para hallar el índice de la cola de los datos extremos consiste en hallar una zona suficientemente constante en el Hill-plot.

Este método no-paramétrico utiliza ciertas mejoras que cabe destacar. Por un lado, es relevante representar en el Hill-plot las bandas de confianza que

se construyen fácilmente a partir de:

$$\sqrt{i}(\hat{\alpha}_{i,n}^{Hill} - \alpha) \xrightarrow{d} N(0, \alpha^2)$$

dónde $\hat{\alpha}_{i,n}^{Hill} = (\hat{\xi}_{i,n}^{Hill})^{-1}$ y $\alpha = \xi^{-1}$, ver. Cabe añadir que buscar zonas estables del Hill-plot es más efectivo si escalamos el eje de abscisas, dando mayor amplitud a las zonas correspondientes a muestras más pequeñas, o más concretamente, escalar el eje en función del *porcentaje de ocupación*, ver Dress *et al.* (2000).

Otros métodos

Otros métodos que cabe destacar son: el método de la media de los excesos y el método de los bloques.

La metodología de la media de los excesos es análoga a la del método Hill, pero en lugar de considerar el estimador Hill para cada umbral consideramos la media residual. La simplicidad de este método no ha sido suficiente para relevar al método Hill puesto que ciertas distribuciones de la familia GPD no tienen esperanza finita. A pesar de ello, hallamos trabajos recientes en el que podemos ver la potencia de este método frente a la del método Hill para la subfamilia Weibull, ver Dierckx *et al.* (2008).

El método de los bloques resulta viable siempre que dispongamos de un gran volumen de datos y preferentemente agrupado de forma natural por bloques de modo que podamos tomar el máximo de cada bloque. Por ejemplo, el valor máximo diario de un activo. En este caso, si pudiéramos suponer independencia entre dichos máximos, podríamos ajustar el índice de la cola a través de la estimación del parámetro de forma de la familia GEV.

Redactar: Coles calcula els estimadors màxim versemblants per un seguit de sostres i analitza la seva estabilitat.

1.3.2. Modelos para valores extremos

Existen dos caminos en la modelización de datos de valores extremos univariantes e independientes. Podemos modelar todos los datos utilizando modelos que ya contengan distribuciones con cierta variedad para el índice de la cola (para el caso en que lo desconocemos) y calibrarlo según los datos. O bien, podemos modelar sólo los excesos de un umbral fijado.

El método tradicional

Sea X_1, \dots, X_n v.a.i.i.d. Si conocemos el índice de la cola deberemos tenerlo en cuenta para decidir el modelo y esto puede resultar una dificultad. Notemos que la forma $\xi = 0$ es común a muchos modelos conocidos, pero otros valores del MDA limitan mucho la elección. La modelización global de los datos considerando familias de distribuciones óptimas para modelar a su vez valores extremos suele escasear, por ejemplo, tenemos la g-and-h conocida en el ámbito financiero.

El método POT

El modelo Peaks Over Threshold (POT) consiste en enlazar la distribución empírica hasta un umbral fijado con un modelo para los excesos de dicho umbral. Observamos que la distribución que usemos para modelar los excesos debe ser nula para valores negativos.

Sea X_1, \dots, X_n una muestra, u un umbral fijado y sea F la distribución que usaremos para modelar los excesos de dicho umbral, llamemos u el umbral de enlace. La distribución que modela nuestros datos con la metodología POT es

$$F^{POT,u}(x) = \begin{cases} n_x/n, & x \leq u \\ n_u/n + (1 - n_u/n) F(x - u), & x > u \end{cases}$$

dónde $n_s = \#\{X_i | X_i \leq s\}$. El uso de la GPD como familia para modelar los excesos es lo más utilizado puesto que desde un punto de vista teórico existe

un umbral a partir del cual tenemos garantías de que el ajuste de los valores extremos es realmente bueno.

1.4. Análisis de riesgo

El *análisis de riesgo* es el estudio de las causas de los probables eventos extremos y las consecuencias que estos puedan producir. Este tipo de análisis es ampliamente utilizado como herramienta de gestión en estudios financieros y de seguros para identificar y evaluar riesgos, en un sentido perjudicial, para la entidad o empresa involucrada. También se utiliza en ámbitos como el medio ambiente para modelar y predecir eventos catastróficos como terremotos, huracanes y otros fenómenos meteorológicos. El análisis de riesgo también se halla en numerosos otros ámbitos como, por ejemplo, la sanidad, la gestión empresarial y el biológico.

1.4.1. El riesgo en seguros y finanzas

En el ámbito financiero, el primer paso del análisis es identificar los activos a proteger o evaluar. La evaluación de riesgos involucra comparar el nivel de riesgo detectado durante el proceso de análisis con criterios de riesgo establecidos previamente. Los resultados obtenidos del análisis van a permitir aplicar alguno de los métodos para el tratamiento de los riesgos, que involucra identificar el conjunto de opciones que existen para tratar los riesgos, evaluarlas, preparar planes para este tratamiento y ejecutarlos.

Desde el acuerdo de Basilea II se clasificaron los riesgos financieros en varios tipos, esencialmente: riesgo de crédito, riesgo de mercado y riesgo operacional, ver BIS (2005).

Definición 1.4.1. El riesgo operacional (RO) consiste en el riesgo de pérdida debido al fallo, o uso inadecuado, de procesos, personas y sistemas o por

eventos externos. Se incluye riesgo legal y se excluye el riesgo estratégico o comercial y el riesgo de la reputación.

Para medir el riesgo operacional se utiliza el escenario de *Loss Distribution Approach* (LDA). Este tiene tres componentes:

- Frecuencia de pérdidas: una función de distribución del número de pérdidas en un periodo T (generalmente, es un año). Generalmente, consideramos una distribución de Poisson con parámetro el valor esperado para el número de pérdidas en dicho periodo, denotemos N a una v.a con dicha distribución.
- Severidad de la pérdida: una distribución para la severidad de una pérdida por encima de un umbral fijado, u . Denotaremos la v.a. de la severidad de una pérdida por L . Remarcamos que denota un incidente por el cual una entidad sufre daños que pueden medirse con un valor monetario.
- Pérdidas agregadas: una distribución para la severidad de las pérdidas en el tiempo T , que combina las dos anteriores. Generalmente, una v.a. con dicha distribución viene dada por

$$S = \sum_{i=1}^N L_i$$

Además, supondremos que las pérdidas sobre un umbral fijado suficientemente grande son independientes.

La función de distribución para la severidad no está fijada por el acuerdo de Basilea, por lo tanto es un reto para la estadística matemática la elección y calibración del modelo, ya que por el contexto del problema la solución teórica la deberíamos hallar en la teoría de valores extremos, pero a la práctica esta teoría no resulta apropiada. Aunque la GPD proporcione ajustes razonables para la mayoría de situaciones, no se obtienen capitales reguladores o coberturas razonables, ver Dutta&Perry (2006).

1.4.2. Medidas de riesgo

Existen varias medidas del riesgo. El *value-at-risk* (VaR) es el valor que proviene de considerar el cuantil γ de las pérdidas agregadas, S_γ , por lo tanto,

$$VaR_\gamma = S_\gamma$$

el CVaR es el valor que proviene de considerar el valor esperado para la severidad de una pérdida cuyo valor exceda a S_γ

$$CVaR_\gamma = E(S|S > S_\gamma)$$

Para terminar veamos un ejemplo para ilustrar como el percentil 95 de una distribución normal no es un umbral suficientemente alto para asumir cola exponencial, o más generalmente cola GPD.

Un ejemplo

En la gestión de los riesgos de mercado de los grandes bancos, así como las autoridades bancarias, utilizan el VaR para la distribución de pérdidas y ganancias (P/L). El modelo básico asume distribución normal para P/L y las estimaciones se calculan para el 95 (o 99) percentil de la pérdida, con horizonte de diez días (o un día).

Artzner *et al.* (1999) sugirieron el CVaR como medida preferente de riesgo para el VaR. El VaR nos dice lo máximo que podemos esperar a perder si un evento *extremo*, es decir, en la cola, no se produce, y el CVaR nos dicen lo que podemos esperar perder si ocurre un evento en dicha cola.

Supongamos que queremos estimar el $CVaR_{0,95}$. En el supuesto de que P/L se distribuye normalmente con media 0 y desviación estándar 1, el $VaR_{0,95}$ redondeado a tres decimales es 1,645. La cola a partir del umbral a de una distribución normal de media μ y varianza σ^2 , $N(\mu, \sigma^2)$, tienen una distribución normal truncada, $TN(a, \mu, \sigma^2)$, cuya función de densidad viene

dada por

$$f_a(x; \mu, \sigma) = \frac{\varphi((x - \mu)/\sigma)}{\sigma(1 - \Phi(\nu))}, \quad x > a,$$

dónde $\varphi(x) = \Phi'(x)$ es la función de densidad de la distribución normal estándar y $\nu = (a - \mu)/\sigma$. Los primeros dos momentos de $TN(a, \mu, \sigma^2)$ vienen dados por

$$\begin{aligned} E(x) &= \mu + \sigma\lambda(\nu) \\ V(x) &= \sigma^2(1 - \lambda(\nu)^2 + \nu\lambda(\nu)) \end{aligned} \tag{1.13}$$

donde $\lambda(\nu) = \varphi(\nu)/(1 - \Phi(\nu))$ es la inversa del *Mill's ratio*. Por lo tanto, el CVaR es 2.063 por (1.13) y el exceso de capital requerido con respecto al VaR es de $CVaR - VaR = 0,418$.

Cuadro 1.1: Excesos de capital requeridos con respecto al $VaR_{0,95}$ bajo diferentes suposiciones sobre la distribución de la cola.

Normal	Exponencial	GPD(0.1)	GPD(0.125)	GPD(0.25)	GPD(0.5)
0.418	0.485	0.539	0.554	0.646	0.970

La cola de una función de distribución normal se clasifica como cola exponencial. Vamos a considerar el modelo GPD, como modelo alternativo para la cola y con el fin de asumir una continuación continua de la distribución normal estándar en el percentil 95 percentil, hacemos coincidir el punto de partida de la normal truncada de distribución con el valor correspondiente de la distribución GPD

$$f_v(v; 0, 1) = g(0; \xi, \psi)$$

Entonces, por (1.2.6) el exceso de capital requeridos con respecto a VaR, para pérdidas esperadas bajo suposición de cola pesada en el modelo GPD, es

$$CVaR(\xi) - VaR = \frac{1}{(1 - \xi) CVaR}$$

El caso límite, cuando ξ tiende a 0 corresponde a la cola exponencial.

La Tabla 1.1 muestra el exceso de capital requerido, para la pérdida esperada en la cola, con respecto al VaR, bajo normalidad, exponencialidad y algunas distribuciones del modelo GPD con colas pesadas. Observamos con que rapidez podemos duplicar el capital

Capítulo 2

Visión crítica de la modelización estadística de valores extremos

Desde un punto de vista matemático o probabilista la teoría de los valores extremos puede considerarse completa. Hemos visto además que se dispone de dos enfoques alternativos: el comportamiento límite de los máximos que dan lugar a la distribución GEV y el comportamiento límite de observaciones sobre un umbral que dan lugar a la distribución GPD. Desde el punto de vista estadístico las cosas son menos satisfactorias pues las propiedades con muestras pequeñas de los teoremas límite anteriores son poco conocidas, ver Diebold *et al.* (1998).

Los retos principales en EVT se resumen en encontrar el índice de la cola que caracteriza los valores extremos de una muestra o distribución y hallar el umbral óptimo dónde enlazar una distribución de la familia Pareto generalizada. Asimismo, aparecen otros problemas como la estrechez de modelos para valores extremos. Existen estrategias y modelos en la teoría de valores extremos que a nivel teórico están muy cerca de resolver dichos retos, no obs-

tante, en modelización estadística estas metodologías están sujetas a muchas críticas.

En general, la teoría asintótica llevada a la práctica manifiesta errores o inexactitudes que no siempre podemos controlar, evaluar o predecir. Esto se hace latente en modelización estadística y, en general, se acentúa en el caso de modelización estadística de valores extremos.

A continuación, vamos a discernir sobre algunas de estas metodologías y modelos que constituyen la teoría de los valores extremos y por tanto, fundamentan su modelización estadística. A la vez, iremos introduciendo los fundamentos de nuevas metodologías estadísticas para valores extremos.

2.1. Distribución de Pareto o *power-law*?

Los problemas de modelización estadística de los valores extremos están diluidos en numerosas disciplinas científicas y sociales, esto se ha traducido en una gran diversificación del lenguaje utilizado para hallar el índice de la cola y el umbral óptimo. Esto no supondría un problema si no fuera porque estos lenguajes se traducen en diferentes parametrizaciones de la familia GPD y por consiguiente las estimaciones de estos parámetros, con métodos diferentes a la máxima verosimilitud, pueden dar lugar a errores no controlados. No obstante, recurrir a la estimación por máxima verosimilitud para el índice de la cola no siempre parece viable como discutiremos a continuación.

La familia Pareto, incluida en el familia GPD como el subconjunto de distribuciones con parámetro de forma $\xi > 0$, corresponde a una familia bi-paramétrica que se parametriza por $\alpha = 1/\xi$ y $\sigma = \psi/\xi$.

Definición 2.1.1 (Familia Pareto). La familia Pareto corresponde a la familia de funciones de densidad dadas por

$$p(x; \alpha, \sigma) = \alpha \sigma^{-1} (1 + x/\sigma)^{-\alpha-1}, \quad (2.1)$$

para $x > 0$ donde $\alpha > 0$ y $\sigma > 0$.

La familia Pareto está relacionada con la familia *power-law*, Sornette (2006), Clauset *et al.* (2009). Incluso, se abusa del término Pareto para designar a la familia *power-law*.

Definición 2.1.2 (Familia *power-law*). La familia *power-law* corresponde a la familia de funciones de densidad dadas por

$$p_w(y; \alpha, \sigma) = \alpha \sigma^{-1} (y/\sigma)^{-\alpha-1} \quad (2.2)$$

dónde el rango de y es $[\sigma, \infty)$ y $\alpha > 0$.

Puede parecer que si tomamos $y = (x + \sigma)$ los dos modelos son el mismo, pero no es cierto. Estas dos familias están relacionadas y esta relación se puede interpretar fácilmente si destacamos que las dos corresponden a subfamilias de la familia de tres parámetros Pareto tipo II, Arnold (1983)

Definición 2.1.3 (Familia Pareto tipo II). La familia Pareto tipo II corresponde a la familia de funciones de densidad dadas por

$$p_{II}(z; \alpha, \mu, \sigma) = \alpha \sigma^{-1} \left(1 + \frac{z - \mu}{\sigma} \right)^{-\alpha-1}$$

dónde el rango de z es $[\mu, \infty)$ y α, μ y σ son positivos.

Efectivamente, la familia Pareto corresponde a la subfamilia de la Pareto tipo II que se obtiene de $\mu = 0$ y la familia *power-law* corresponde a la subfamilia de la Pareto tipo II que se obtiene de $\mu = \sigma$. Más concretamente, si $\{x_i\}$ es una muestra de Pareto, entonces $\{y_i\} = \{x_i + \sigma\}$ es *power-law* y $\{z_i\} = \{x_i + \mu\}$ es Pareto tipo II.

Por lo tanto, tenemos tres modelos diferentes con el valor de α como candidato a caracterizar el tipo de valores extremos para el caso de cola pesada. Es correcto que, bajo la hipótesis del parámetro de forma ξ positivo, la distribución de Pareto permite determinar el índice de la cola. Pero a menudo se usa el valor obtenido de la estimación del parámetro α del modelo *power-law* para estimar el índice de la cola de los datos.

Sea $\{x_{(1)}, \dots, x_{(n)}\}$ con $x_i > 0$ para $i = 1 \dots n$ una muestra de Pareto con parámetros α y σ . La función de log-verosimilitud viene dada por:

$$lp(x_i; \alpha, \sigma) = n \left(\log(\alpha\sigma^{-1}) - (\alpha + 1) \frac{1}{n} \sum_{i=1}^n \log(1 + x_i/\sigma) \right) \quad (2.3)$$

dónde $\sigma > 0$, $\alpha > 0$ y $x_i > 0$. Así, las ecuaciones de máxima verosimilitud son

$$\begin{aligned} 1/\alpha &= \frac{1}{n} \sum_{i=1}^n \log(1 + x_i/\sigma) \\ \alpha/(1 + \alpha) &= \frac{1}{n} \sum_{i=1}^n 1/(1 + x_i/\sigma) \end{aligned} \quad (2.4)$$

estas no siempre tienen solución. La existencia del MLE para la familia GPD es un punto complejo que trataremos en profundidad en el Capítulo 3. Si bien es cierto que esta debilidad del modelo motiva la existencia de otras metodologías de estimación y que añade inseguridad a la elección del índice de la cola y el umbral óptimo. No obstante, cabe insistir en que en caso de que exista debería ser tomado como el estimador de referencia.

Ahora veamos en el modelo power-law cual es la situación. Sea $\{y_1, \dots, y_n\}$ positivos una muestra de power-law con parámetros α y σ . La función de log-verosimilitud viene dada por:

$$lp(y_i; \alpha, \sigma) = n \left(\log(\alpha\sigma^{-1}) - (\alpha + 1) \frac{1}{n} \sum_{i=1}^n \log(y_i/\sigma) \right) \quad (2.5)$$

dónde $0 < \sigma < y_{(1)}$, $\alpha > 0$ y $y_i > 0$. Para α fijado esta función es creciente y toma su máximo para $\sigma = y_{(1)}$, el mínimo de la muestra. Además tenemos una ecuación de verosimilitud,

$$1/\alpha = \frac{1}{n} \sum_{i=1}^n \log(y_i/\sigma) \quad (2.6)$$

que determina el MLE del parámetro α , puesto que siempre tiene solución única para σ dado. Por lo tanto, el MLE para la familia uni-paramétrica power-law (con σ fijado) siempre existe como solución de la ecuación de verosimilitud, ver Ejemplo 2.1.4.

Por lo tanto, dada una muestra de power-law $\{y_1, \dots, y_n\}$, no parece trivial recuperar el valor de σ con el cual la correspondiente muestra dada

por $\{x_i = y_i - \sigma\}$ tenga distribución Pareto y por consiguiente estimar el valor de α que determina el índice de la cola bajo hipótesis del parámetro de forma $\xi > 0$, ver Ejemplo 2.1.4.

Observamos que el estimador de Hill corresponde al MLE de la familia uni-paramétrica de la *power-law*, que si bien también cumple una propiedad análoga a la Proposición 1.2.9, no debe confundirse con el MLE de la Pareto.

Finalmente, el MLE de la Pareto tipo II es $\mu = \min\{x_i\}$ y $(\hat{\alpha}, \hat{\sigma})$ los mismos que obtenemos de la Pareto.

Ejemplo 2.1.4. Tomamos una muestra pequeña $\{x_1, x_2\}$ como

$$\{10, 1000\}$$

. Suponemos que son observaciones independientes con distribución en la familia Pareto. El MLE es $\hat{\alpha} = 0,41$ y $\hat{\sigma} = 12,95$. Por lo tanto, el índice de la cola queda determinada por el parámetro de forma $\xi = 2,46$. Ahora consideramos la correspondiente muestra $\{y_i = x_i + \sigma\}_{i=1,2}$, es decir,

$$\{10 + 12,95, 1000 + 12,95\}$$

que sabemos que por hipótesis siguen siendo independientes y ahora con distribución *power-law*. Si fijamos en el modelo *power-law* el valor de $\sigma = 12,95$, obtenemos fácilmente el correcto estimador MLE para α , es decir, $\hat{\alpha} = 0,41$. Ahora bien, si pretendemos tomar los datos como un modelo bi-variente obtenemos $\hat{\sigma} = 22,95$, $\hat{\alpha} = 0,20$ que correspondería a un valor muy diferente del índice de la cola $\xi = 5,03$, el doble de la correcta estimación con el modelo Pareto.

El método para hallar el umbral óptimo más popularmente aceptados son el estimador de Hill y él de la media de los excesos. Los dos estimadores han estado sujetos a críticas desde su aparición. No obstante, la modificación de estos estimadores y de las metodologías asociadas los han llevado a ser los métodos más utilizados con la restricción de considerar el estimador Hill

más apropiado para el caso de cola Pareto, o equivalentemente con valores extremos en la clase MDA de la subfamilia Fréchet y el estimador de la media de los excesos, para el caso de cola ligera, o equivalentemente con valores extremos en la clase MDA de la subfamilia Weibull, ver Guillou&Hall (2001).

La principal crítica al método Hill radica en que su formulación no tiene su fundamento en un método de estimación como el MLE o el método de los momentos. La del exceso de medias radica en su definición, ya que requiere la existencia del primer momento y no todas las clases de valores extremos de la familia Fréchet lo tienen, ver Definición 1.2.5.

2.2. Sobre la estimación del umbral óptimo

Sea X una v.a. con distribución F , por el Segundo Teorema fundamental 1.2.10 existe $u_0 > 0$, tal que $F_{u_0}(x) = G(x; \xi_0, \psi_0)$ para algún $\xi_0 \in \mathbb{R}$ i $\psi_0 > 0$. Por la Proposición 1.2.9, X_u es una v.a. con función de distribución $G(x; \xi_0, \psi_0 + u\xi_0)$ para todo $u > u_0$. Por la Observación 1.2.6,

$$me_X(u) = E(X - u | X > u) = \frac{\psi_0 + u\xi_0}{1 - \xi_0} \quad (2.7)$$

por lo tanto, me es una función lineal para $u > u_0$ y esto justifica el método ME-plot para hallar el umbral óptimo. Por otro lado, usando el Teorema de Karamata, ver detalles en McNeil *et al.* (2005, pg 287),

$$\lim_{u \rightarrow \infty} me_{\log X}(\log u) = E(\log X - \log u | \log X > \log u) = \xi \quad (2.8)$$

por lo tanto, $me_{\log X}(\log u)$ es aproximadamente constante para $u > u_0$ y esto justifica el método de Hill-plot para hallar el umbral óptimo.

Castillo *et al.* (2013), proponen una nueva herramienta para tratar los valores extremos: el coeficiente de variación residual, considerado como un proceso aleatorio en términos del umbral. Dada una muestra $\mathbf{x} = \{x_{(1)}, \dots, x_{(n)}\}$ de valores positivos ordenados denominamos CV-plot la representación del

conjunto

$$k \mapsto cv(\{x_{(j)} - x_{(k)} : j \geq k\})$$

Por la Observación 1.2.6, el coeficiente de variación de una variable con distribución GPD es independiente del parámetro de escala, viene dado por

$$cv = 1/\sqrt{(1 - 2\xi)} \quad (2.9)$$

por lo tanto el coeficiente de variación residual es constante para $u > u_0$

$$cv(u) = cv(X - u | X > u) = \frac{1}{1 - 2\xi} \quad (2.10)$$

y esto justifica el uso del método CV-plot para la elección del umbral. Las mismas críticas que sufre el ME-plot pueden aplicarse al método CV-plot, pero tiene la ventaja que no depende del parámetro de escala σ y, por tanto, hallar una zona estable para la elección del umbral es tan fácil como lo es en el Hill-plot. Además el CV-plot para colas exponenciales dispone de un test de hipótesis, Castillo *et al.*(2013) para hallar el umbral óptimo, veremos un ejemplo de su uso en el Capítulo 3.

Podemos concluir que el CV-plot es un método que con la misma simplicidad que el Hill-plot o el ME-plot, es desde un punto de vista teórico igual, o mejor, que estos para aproximar el valor del parámetro de forma de la cola.

En general, dado un método, MET , para la estimación del índice de la cola de unos datos $\{x_{(1)}, \dots, x_{(n)}\}$ de valores positivos. Una metodología estándar para hallar el umbral óptimo es tomar por umbral cada valor de la muestra ordenada, $u = x_{(i)}$ y estimar el índice de la cola de la muestra truncada $\mathbf{x}_u, \hat{\xi}_i^{MET}$. Esto se puede representar gráficamente como el conjunto de puntos

$$\{(i, \hat{\xi}_i^{MET}) \mid i = 1..n\} \quad \text{o} \quad \{(x_{(i)}, \hat{\xi}_i^{MET}) \mid i = 1..n\}$$

La zona constante del gráfico indicará la localización aproximada del umbral óptimo y aproximará el índice de la cola, basándonos en la propiedad descrita en la Proposición 1.2.9.

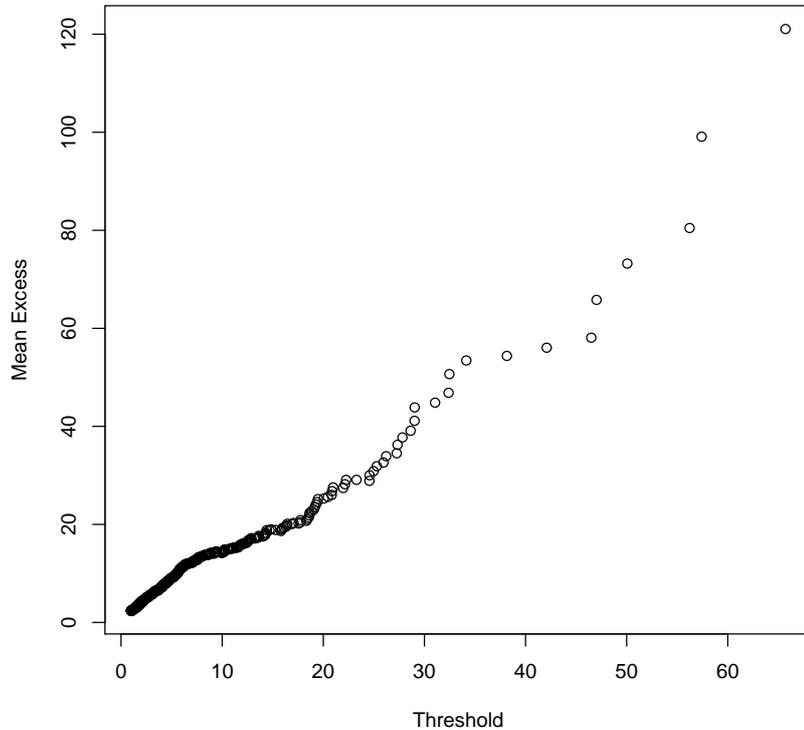


Figura 2.1: ME-plot de los datos sobre incendios en Dinamarca.

Utilizar como *MET* el MLE puede parecer la mejor metodología y puede sorprender que no sea usado este método más a menudo, pero debemos recordar que el método MLE para GPD es un tema complejo. Por un lado, la solución no-explicita de las ecuaciones de verosimilitud, actualmente no supone un problema puesto que es un ejercicio numérico, pero en los orígenes de la teoría de valores extremos, sí. Por otro lado, la posible no solución de las ecuaciones de verosimilitud, esto quedará resuelto en el Capítulo 3. Además, hay que destacar que de la estimación de los parámetros de la GPD también obtenemos una estimación del parámetro de escala ψ y por tanto

2.2. Sobre la estimación del umbral óptimo

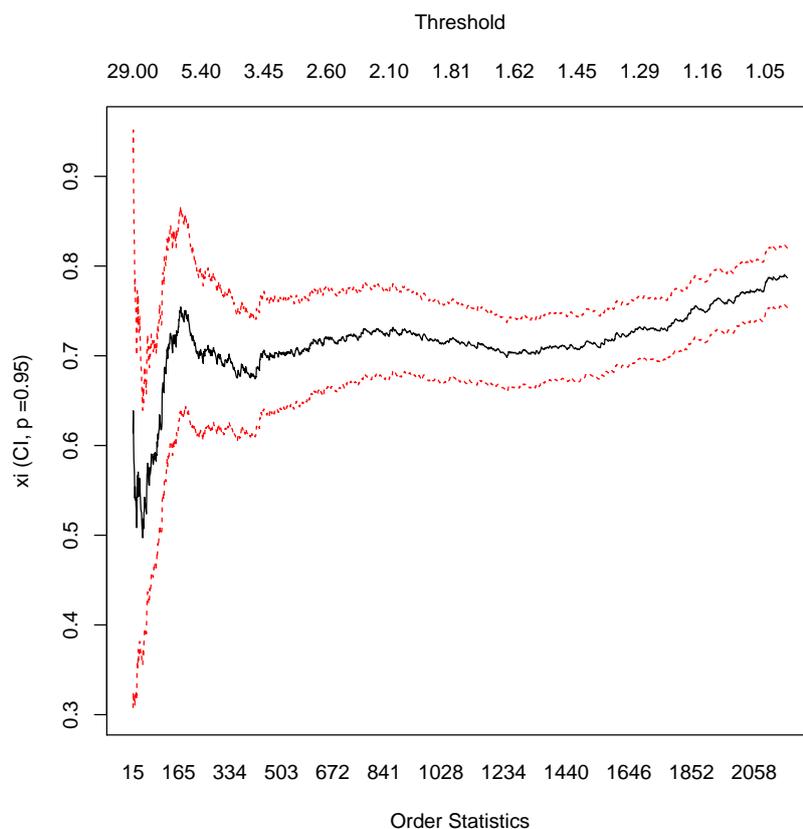


Figura 2.2: Hill-plot de los datos sobre incendios en Dinamarca.

también deberíamos tenerlo en cuenta en la elección del umbral óptimo, este problema halla una solución eficaz en Coles (2001, pg 83). Por todo esto, no es de extrañar que tengamos otros métodos para la elección del umbral, como por ejemplo, el ME-plot, ver Figura 2.4, el Hill-plot, ver Figura 2.2 o el nuevo CV-plot, ver Figura 2.3.

2.3. Revisión de los modelos para valores extremos

Sea $\{x_{(1)}, \dots, x_{(n)}\}$ una muestra ordenada con una función de distribución arbitraria. Si queremos modelar estos datos o, al menos, sus valores extremos, es evidente, que el mejor modelo desde un punto de vista teórico, es el POT con la GPD para los extremos, puesto que tenemos certeza matemática de que existe un umbral a partir del cual una distribución de la familia GPD ajusta tan bien como queramos nuestros datos. Ahora bien, todos los métodos para hallar el umbral óptimo, así como determinar el índice de la cola, trabajan bajo la suposición de que $u^* < x_{(n)}$ y esto no tiene porque cumplirse.

En caso de que no se cumpla $u^* < x_{(n)}$, el uso del modelo GPD carecería totalmente de fundamento teórico y muchos otros modelos serían más apropiados.

Más aún, si el conjunto $\{x_i \mid x_i > u^*\}$, es pequeño, tampoco deberíamos confiar demasiado en el modelo GPD, puesto que podríamos tener insuficientes datos para calibrar correctamente el modelo.

Y más aún, si por ejemplo queremos analizar el 20% de los valores extremos de nuestros datos y el umbral óptimo cae mucho más allá del 1% mayor, el método POT consistirá esencialmente en la distribución empírica y sólo este 1% estará ajustado por un modelo teórico. Este modelo podría ser insuficiente para resolver nuestro análisis.

Nuestra experiencia con datos reales nos ha llevado a restringirnos a modelos basados en POT, puesto que a menudo nos encontramos con datos que ya han sido truncados, por ejemplo, se suelen registrar magnitudes medio ambientales extremas, e ignoramos las medidas pequeñas, de ciertos fenómenos como cantidad de lluvia, magnitud de terremotos,... También sucede en finanzas, por ejemplo, se analizan para su modelización pérdidas por fraude que superen cierto umbral prefijado puesto que son las que se conocen en detalle e interesa modelar. También el hecho de que el análisis de nuestros

datos esté enfocado en resolver cuestiones que se centran en los extremos de nuestros datos nos lleva a considerar la metodología POT, puesto que queremos un ajuste lo más preciso posible para valores extremos y el resto carece de interés. Ahora bien, en lugar de utilizar como se hace clásicamente la GPD en el método POT, nosotros proponemos utilizar otros modelos.

Definición 2.3.1. Sea F una función de distribución para variables positivas. Consideramos la correspondiente familia de funciones de distribución de los excesos de F , $\{F(\cdot; u)\} = \{F_u \mid u > 0\}$, ver Definición 1.2.7. El modelo POT con enlace F consistirá en el modelo POT con función de distribución de enlace una de la familia de los excesos de F , denotaremos $F^{POT, u_0}(\cdot; u)$, dónde u_0 es el umbral de enlace.

Ejemplo 2.3.2. Consideramos la función de distribución de una variable aleatoria Y LogNormal, $LN(\mu, \sigma)$, es decir, $Y = \exp(X)$ para X con función de distribución normal, $N(\mu, \sigma)$ y consideramos la correspondiente familia tri-paramétrica de funciones de distribución de sus excesos, descritas para $y \geq 0$ como

$$F(y; \mu, \sigma, u) = \frac{\phi\left(\frac{\ln(y+u)-\mu}{\sigma}\right) - \phi(\alpha)}{1 - \phi(\alpha)} \quad (2.11)$$

dónde $\alpha = \frac{\ln u - \mu}{\sigma}$, ϕ es la función de distribución de la Normal estándar. Y las funciones de densidad vienen dadas por

$$f(y; \mu, \sigma, u) = \frac{1}{\sigma(y+u)} \varphi\left(\frac{\ln(y+u)-\mu}{\sigma}\right) / (1 - \phi(\alpha)) \quad (2.12)$$

dónde φ es la función de densidad de la normal estándar.

Efectivamente, una primera propuesta como conclusión a la revisión de los modelos estadísticos para valores extremos es ampliarlos y considerar modelos POT con distribución de enlace los excesos de las distribuciones de un modelo prefijado. Observamos que considerar distribuciones de excesos de enlace proporciona flexibilidad en la elección del umbral de enlace. Si los modelos para valores extremos carecían de diversificación con el clásico uso de la

GPD, este nuevo punto de vista abre quizás demasiado las posibilidades. Pero no es así, puesto que para empezar debemos considerar modelos que ajusten los excesos de nuestros datos, para ello un buen indicador es el coeficiente de variación de los excesos, ver la Figura 2.3. Más allá de estas restricciones descriptivas veremos más restricciones para los modelos estadísticos para valores extremos en el Capítulo 5.

Sea X una v.a. con función de distribución F_X , el coeficiente de variación de sus excesos del umbral u viene dado por

$$cv^2(X_u) = \frac{V(X|X > u)}{(E(X|X > u) - u)^2} = \frac{\frac{u_2}{u_0} - \left(\frac{u_1}{u_0}\right)^2}{\left(\frac{u_1}{u_0} - u\right)^2} = \frac{u_2u_0 - u_1u_1}{(u_1 - u_0u)^2} \quad (2.13)$$

donde $u_n = \int_u^\infty x^n f_X(x)dx$ y f_X es la función de densidad. Para varias distribuciones conocidas tenemos calculados los u_n en Jawitz (2004).

Teorema 2.3.3. *El coeficiente de variación de los excesos de la distribución LogNormal tiende a 1 si el umbral tiende a infinito.*

Demostración. Sea Y una distribución LogNormal con parámetros μ y σ y sea f_Y la correspondiente función de densidad. Para calcular el coeficiente de variación residual o de los excesos del umbral u debemos calcular u_k para $k = 0, 1, 2$ dónde $u_k = \int_u^\infty y^k f_Y(y)dy$, ver (2.13). Estos se obtienen del caso general que detallamos a continuación para $k > 0$ puesto que $u_0 = 1 - \phi(\alpha)$

$$u_k = \int_u^\infty t^k f_Y(t)dt = \int_u^\infty t^k \frac{1}{t\sigma} \varphi\left(\frac{\ln t - \mu}{\sigma}\right)dt \quad (2.14)$$

con el cambio de variable $t = \exp(\sigma x + \mu)$ obtenemos

$$u_k = \int_\alpha^\infty (\exp(\sigma x + \mu))^k \varphi(x)dx \quad (2.15)$$

Y si sustituimos por la expresión de φ obtenemos

$$u_k = \frac{1}{\sqrt{2\pi}} \int_\alpha^\infty \exp(k\sigma x + k\mu - \frac{1}{2}x^2)dx \quad (2.16)$$

Finalmente, por completación de cuadrados y para el caso de la $k = 1$ y $k = 2$, obtenemos

$$u_1 = \frac{\exp\left(\mu + \frac{\sigma^2}{2}\right)(1 - \phi(\alpha - \sigma))}{(1 - \phi(\alpha))} \quad (2.17)$$

y

$$u_2 = \frac{\exp(2\mu + 2\sigma^2)(1 - \phi(\alpha - 2\sigma))}{(1 - \phi(\alpha))} \quad (2.18)$$

Por lo tanto, como $cv^2(Y_u) = \frac{u_2 u_0 - u_1 u_1}{(u_1 - u_0 u)^2}$, ver (2.13), obtenemos

$$cv^2(Y_u) = \frac{\frac{\exp(2\mu+2\sigma^2)(1-\phi(\alpha-2\sigma))}{(1-\phi(\alpha))} - \left(\frac{\exp(\mu+\frac{\sigma^2}{2})(1-\phi(\alpha-\sigma))}{(1-\phi(\alpha))}\right)^2}{\left(\frac{\exp(\mu+\frac{\sigma^2}{2})(1-\phi(\alpha-\sigma))}{(1-\phi(\alpha))} - u\right)^2} \quad (2.19)$$

que se puede reducir usando la llamada Q -función $Q(z) := 1 - \phi(z)$. Además, denotaremos por ω_1 y ω_2 las expresiones $\exp(\mu + \frac{\sigma^2}{2})$ y $\exp(2\mu + 2\sigma^2)$, respectivamente.

$$cv^2(Y_u) = \frac{\omega_2 Q(\alpha - 2\sigma)Q(\alpha) - (\omega_1 Q(\alpha - \sigma))^2}{(\omega_1 Q(\alpha - \sigma) - uQ(\alpha))^2} \quad (2.20)$$

Además la Q -función se puede acotar con las desigualdades de *Mill's ratio*:

$$l(x)\varphi(x) < Q(x) < u(x)\varphi(x), \quad x > 0.$$

De funciones $u(x)$ y $l(x)$ tenemos varias, por ejemplo, una acotación clásica viene dada por $l(x) = \frac{x}{1+x^2}$ y $u(x) = \frac{1}{x}$. No obstante hay más recientes, por ejemplo, en Baricz (2008) encontramos:

$$l(x) = \frac{2}{\sqrt{x^2 + 4} + x} \quad y \quad u(x) = \frac{4}{\sqrt{x^2 + 8} + 3x} \quad (2.21)$$

Vamos a proceder a la acotación del coeficiente de variación de los excesos de u usando las desigualdades de *Mill's ratio*. La cota superior es

$$cv^2(Y_u) < \frac{\omega_2 u(\alpha - 2\sigma)\varphi(\alpha - 2\sigma)u(\alpha)\varphi(\alpha) - (\omega_1 l(\alpha - \sigma)\varphi(\alpha - \sigma))^2}{(\omega_1 l(\alpha - \sigma)\varphi(\alpha - \sigma) - u(\alpha)\varphi(\alpha))^2 u} \quad (2.22)$$

que simplificando corresponde a

$$cv^2(Y_u) < \frac{u(\alpha)u(\alpha - 2\sigma) - l(\alpha - \sigma)l(\alpha - \sigma)}{(l(\alpha - \sigma) - u(\alpha))^2} \quad (2.23)$$

y esta acotación considerando las funciones dadas por (2.21) tiende a 1 si α tiende a infinito. Análogamente, obtenemos la cota inferior

$$cv^2(Y_u) > \frac{\omega_2 l(\alpha - 2\sigma)\varphi(\alpha - 2\sigma)l(\alpha)\varphi(\alpha) - (\omega_1 u(\alpha - \sigma)\varphi(\alpha - \sigma))^2}{(\omega_1 u(\alpha - \sigma)\varphi(\alpha - \sigma) - l(\alpha)\varphi(\alpha))^2 u} \quad (2.24)$$

que simplificando corresponde a

$$cv^2(Y_u) > \frac{l(\alpha)l(\alpha - 2\sigma) - u(\alpha - \sigma)u(\alpha - \sigma)}{(u(\alpha - \sigma) - l(\alpha))^2} \quad (2.25)$$

y esta acotación también tiende a 1 si α tiende a infinito considerando las funciones dadas por (2.21). Finalmente, dado que para μ y σ fijados, tenemos que u tiende a infinito si y sólo si α tiende a infinito, obtenemos

$$\lim_{u \rightarrow \infty} cv^2(Y_u) = 1 \quad (2.26)$$

□

Es interesante notar que el coeficiente de variación de los excesos de una distribución puede tomar formas muy diversas a pesar de la poca flexibilidad que ofrecen las distribuciones de sus excesos cuando el umbral tiende a infinito. Efectivamente, cuando el umbral tiende a infinito la distribución de sus excesos tiende a una distribución en la familia GPD y por tanto queda fijado el índice de la cola y a su vez el correspondiente coeficiente de variación. Por ejemplo, vemos en la Figura 2.3 como el coeficiente de variación de los excesos de la distribución LogNormal se halla alejado del valor del límite 1, ver Teorema 2.3.3.

2.4. Ejemplo de POT con enlace la LogNormal

A continuación compararemos dos modelos para valores extremos: el clásico POT con GPD y el POT con enlace LogNormal, para unos datos de seguros sobre las pérdidas por incendios en Dinamarca. Estos datos son un buen ejemplo de riesgo operacional y han sido usados en este contexto por varios autores, ver Embrechts *et al.*(1997). El hecho que se correspondan a datos representativos de de riesgo operacional nos lleva a considerar el modelo LogNormal como alternativa al modelo clásico GPD, Carrillo&Suárez (2006) apuntan al uso del modelo LogNormal con $\mu = 5$ y $\sigma = 2$. Nuestra

innovación radica en considerar en lugar de la LogNormal como modelo, usar los excesos de la LogNormal con la metodología POT. La elección de este modelo como candidato al ajuste está acreditado desde un punto de vista descriptivo como veremos en la figura .

2.4.1. Datos de seguros: pérdidas por incendios en Dinamarca

Vamos a utilizar los datos daneses sobre las pérdidas de seguros de incendio a partir de enero de 1980 a diciembre de 1990 como un proxy de pérdidas por riesgo operacional, ver Degen *et al.*(2007). La unidad es millones de coronas danesas y el rango de los datos originales es (0.313,263.250), pero como se hace en este ámbito, vamos a utilizar las pérdidas de más de un millón de coronas puesto que los datos para valores inferiores se tiene evidencia de que podrían estar incompletos. Este conjunto de datos se ha sido analizado por varios autores, por ejemplo, Resnick (1997) contrasta el ajuste de algunos modelos y verifica la independencia de los datos. Embrechts *et al.* (1997) utiliza esta muestra debido a que es un buen ejemplo de una muestra de cola pesada, dónde la elección del umbral y el MDA no es trivial. Todos los autores remarcan la inestabilidad de los resultados cuando atacan este problema. En general, estos problemas son comunes en el ámbito de las finanzas de riesgo operacional y dificulta el cálculo del capital regulatorio, de hecho Degen *et al.* (2007) utilizan estos datos como modelo de riesgo operacional.

2.4.2. Análisis de los datos

En las Figuras 2.2 y 2.4 podemos observar que ni el Hill-plot ni el ME-plot nos dan suficiente información para la elección del umbral óptimo tal y como apuntan otros autores. Siguiendo nuestra metodología cabe suponer que esto nos indica que el umbral óptimo es irrelevante y que deberíamos considerar otros modelos. Si observamos el CV-plot en la Figura 2.3 vemos

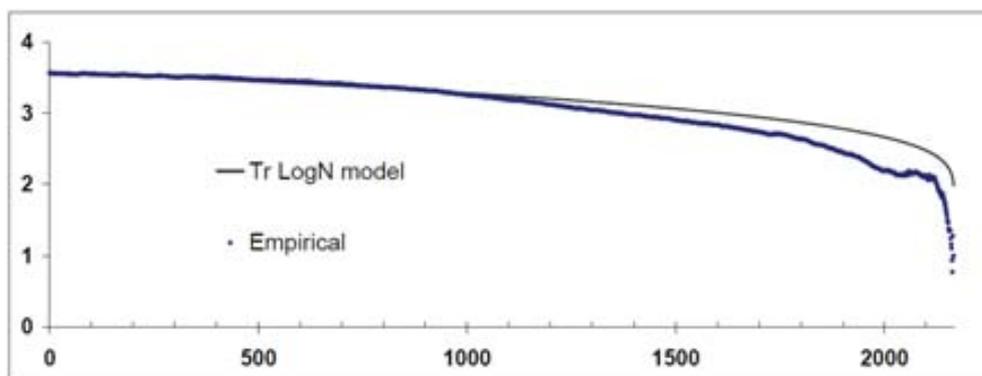


Figura 2.3: El CV-plot teórico para el modelo TrLogN junto el CV-plot residual de los datos sobre pérdidas por incendios en Dinamarca.

que tampoco muestra una zona constante dónde hallar el umbral óptimo, pero además observamos que no deberíamos descartar una cola exponencial por el rápido decaimiento del coeficiente de variación en valores extremos, decae más rápido que oscila.

Con este ejemplo hemos visto la utilidad del coeficiente de variación en el modelo de las colas de las distribuciones.

2.4.3. Modelización de los datos

La Figura 2.4 muestra que el cv empírico no es constante, como sucedería con datos distribuidos por una GPD y en cambio presenta un ajuste excelente al coeficiente de variación de los excesos de la función de distribución LogNormal. Esto junto al hecho que la cola de la distribución LogNormal es exponencial nos hace concluir que la cola de nuestros datos no es pesada y que el modelo POT con enlace LogNormal es un modelo apropiado.

La Tabla 2.1 muestra que el método POT con enlace LogNormal es tan razonable como usado con el modelo GPD. Más aún, mirando el CVaR, Tabla 2.2, vemos que el nuevo modelo es más realista que el clásico modelo

2.4. Ejemplo de POT con enlace la LogNormal

Cuadro 2.1: Estimación del VaR de las pérdidas por incendios en Dinamarca bajo suposición de cola GPD y TRLogN, usando diferentes umbrales, u , 512, 254 y 109.

	Obs.	GPD model			TrLogN model		
		$u = 3,1$	$u = 5$	$u = 10$	$u = 3,1$	$u = 5$	$u = 10$
90 %	5.5	5,7	5,6	5,5	5,7	5,6	5,5
95 %	10.0	9,2	9,3	10,1	9,3	9,4	10,1
99 %	26.0	27,5	27,5	27,3	27,1	27,5	27,4
99.9 %	131.6	129,2	121,4	94,5	114,2	108,1	97,1

con la GPD. Esto puede ser observado también en el CV-plot en la Figura 2.4 que corresponde a la cola (254) de los datos con un umbral inicial $u = 10$ transformados a cola exponencial bajo hipótesis de ajuste por el modelo GPD y el modelo TrLogN transformadas a exponencialidad.

Cuadro 2.2: Estimación del CVaR de las pérdidas por incendios en Dinamarca bajo suposición de cola GPD y TRLogN, usando diferentes umbrales, u , 512, 254 y 109.

	Obs.	GPD model			TrLogN model		
		$u = 3,1$	$u = 5$	$u = 10$	$u = 3,1$	$u = 5$	$u = 10$
90 %	15.6	17,7	17,1	15,8	16,2	15,9	15,9
95 %	24.1	28,3	27,0	24,0	25,3	24,6	23,9
99 %	58.6	83,5	76,6	58,3	68,7	64,1	58,4
99.9 %	186.8	389,8	332,1	192,5	265,3	223,3	185,1

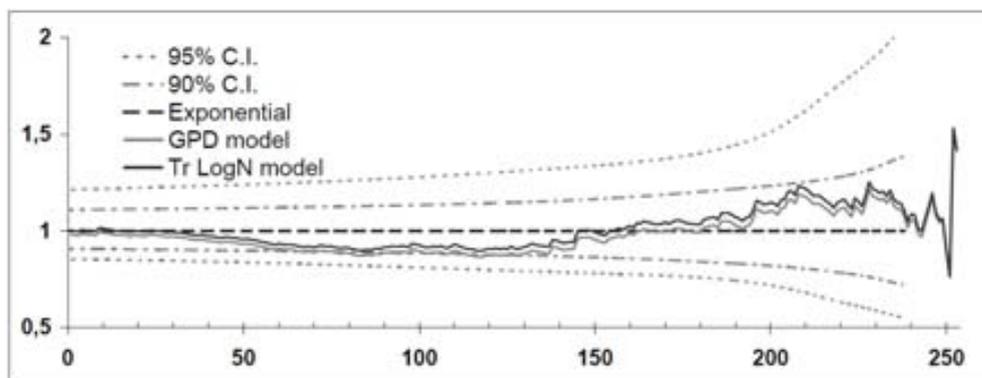


Figura 2.4: CV-plot de los datos sobre incendios en Dinamarca transformados a cola exponencial, bajo ajuste del modelo GPD y TrLogN.

2.4.4. Consideraciones finales

El buen ajuste con el modelo TrLogN sugiere la posibilidad de que el índice de la cola es 0, es decir la cola es exponencial. En efecto, a pesar de la convergencia del coeficiente de variación de los excesos de la distribución LogNormal a 1, vemos que dicha convergencia es muy lenta, ver la Figura 2.4. Embrechts *et al.* (1997) prueban que la distribución LogNormal es sub-exponencial pero no es de variación regular, esto nos motiva a buscar modelos para el riesgo operacional en esta clase. Nuestra experiencia con datos reales nos lleva a concluir que las colas exponenciales son mucho más frecuentes de lo que consideramos, esto lo veremos con más detalle en el Capítulo 3. En cualquier caso, nuevos modelos y metodologías son necesarios y ponemos énfasis en el enfoque del CV-plot como herramienta descriptiva para especificar el modelo y simultáneamente hallar el umbral óptimo como alternativa al método Hill y al ME-plot. Y a la vez, hemos introducido la necesidad de pautas justificadas para nuevos modelos para valores extremos.

2.5. Revisión de la clasificación

Además de la clasificación de las colas en ligeras, exponenciales y pesadas, es importante introducir un nuevo tipo de cola, se trata de la cola uniforme, para $\xi = -1$. Efectivamente, corresponde a una cola ligera, pero su comportamiento resulta casi tan atráctor como el tipo exponencial.

Proposición 2.5.1. *Sea $f(x)$ la función de densidad de una v.a. con soporte $[0, x_F]$ y $0 < f(x_F) < \infty$. Entonces, tiene cola con distribución uniforme.*

Demostración. De McNeil *et al.* (2005)

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_u(x)| = 0 \quad (2.27)$$

donde G_u denota la distribución uniforme en $[0, x_F - u]$. Dado que f es una función continua en x_F , para cada $\epsilon > 0$, existe $\delta_\epsilon > 0$ tal que si s satisface $x_F - s < \delta_\epsilon$, entonces

$$c - \epsilon < f(s) < c + \epsilon \quad (2.28)$$

donde $c = f(x_F)$. Tomando integrales, sobre el intervalo cerrado $[u, x_F]$, en la igualdad (2.28) obtenemos

$$(x_F - u)(c - \epsilon) < 1 - F(u) < (x_F - u)(c + \epsilon) \quad (2.29)$$

Por lo tanto,

$$\left| \frac{f(s)}{1 - F(u)} - \frac{1}{x_F - u} \right| < \left(\frac{c + \epsilon}{c - \epsilon} - 1 \right) \frac{1}{(x_F - u)}$$

para todo s tal que $x_F - s < \delta_\epsilon$. Observamos que para todo $\epsilon_0 > 0$ existe u tal que $|F_u(x) - G_u(x)| < \epsilon$ para todo $x \in [0, x_F - u]$, ya que

$$\begin{aligned} |F_u(x) - G_u(x)| &= \left| \int_0^x f_u(x) dx - 1/(x_F - u) dx \right| \leq \\ &\leq \int_0^x |f(s)/(1 - F(u)) - 1/(x_F - u)| dx \leq (c + \epsilon)/(c - \epsilon) - 1 < \epsilon_0 \end{aligned}$$

tomando $\epsilon = 2c/(2 + \epsilon_0)$ y $u = x_F - \delta_\epsilon$. Observamos que, la construcción de u es minimal, entonces la ecuación (2.27) se cumple. \square

Capítulo 2. Visión crítica de la modelización estadística de valores extremos

La Proposición 2.5.1 muestra que si una familia de distribuciones truncada por la derecha, en un punto donde la función de densidad sea diferente de cero, es usada como modelo para colas, entonces sólo estaremos considerando una clase de colas: las colas uniformes, y habremos fijado el índice de la cola en -1 , por ejemplo, Aban *et al.* (2006).

Capítulo 3

Especificación y estimación en el modelo GPD

La motivación de este Capítulo se encuentra en el hecho de que a pesar de las debilidades del MLE, lo necesitamos en muchos procedimientos inferenciales. Los ajustes de varios modelos a los mismos datos son comúnmente comparados con AIC, BIC o LRT en modelos anidados, todos de ellos basado en el MLE. La GPD ha sido reparametrizada con el parámetro de forma $\kappa = -\xi$. Un resumen de los resultados teóricos para el MLE en la GPD se puede encontrar en Davison&Smith (1990). El MLE existe para grandes muestras siempre que $\kappa < 1$ y es asintóticamente eficiente si $\kappa < 0,5$. Para $\kappa > 1$ con probabilidad que tiende a 1 no hay ningún máximo local. Estos resultados serán ampliados en este Capítulo. En la Sección 2 se propone un algoritmo simple, rápido y estable, que siempre proporciona un MLE. Este algoritmo se compara con los métodos de Zhang y Stephens, ZSE, y el de Song y Song, SSE, en términos de eficiencia y sesgo, ninguno de ellos es definitivamente mejor. De acuerdo con el comportamiento de la funciones de densidad de probabilidad de la GPD se pueden distinguir tres sub-modelos separados para la distribución exponencial ($\kappa = 0$) y la distribución uniforme ($\kappa = 1$). El estudio pone de relieve el problema de la mala especificación del

sub-modelo. Realizando una simulación de una de los tres sub-modelos de GPD, los parámetros estimados pueden fácilmente pertenecer a otro, como se muestra en la Tabla 3.2. Los mejores resultados de correcta clasificación siempre se obtienen con MLE o ZSE. La Sección 3.2 muestra la importancia de la distribución uniforme en el GPD, esta separa las funciones de densidad monótonas decrecientes y acotadas superiormente de las funciones de densidad monótonas crecientes que tienden a infinito en el soporte superior, ver la Figura 3.3. Por otra parte, en el Teorema 3.2.2, se demuestra que existe el MLE global para el sub-modelo $0 \leq \kappa \leq 1$. El resultado muestra la importancia de este sub-modelo en el estudio de las distribuciones para colas ligeras utilizadas en hidrología. En la Sección 3.3, veremos algunos resultados numéricos y analíticos de las secciones anteriores que nos van a permitir sugerir una nueva metodología que hace hincapié en el papel de los puntos de transición $\kappa = 0$ y $\kappa = 1$. La metodología se aplica a dos ejemplos polémicos ampliamente estudiados en la literatura .

3.1. Cálculo del estimador de máxima verosimilitud

La GPD, ver la Definición 1.2.5, es la familia bi-paramétrica con densidad dada por

$$g(x; \kappa, \psi) = \begin{cases} \psi^{-1}(1 - \kappa x/\psi)^{1/\kappa-1}, & \kappa \neq 0, \\ \psi^{-1} \exp(-x/\psi), & \kappa = 0, \end{cases} \quad (3.1)$$

para $\kappa > 0$ el rango de x es $0 < x < \psi/\kappa$ y $x > 0$ para $\kappa < 0$. Vamos a definir σ como ψ/κ . Observamos que hemos reparametrizado el parámetro de forma por $\kappa = -\xi$ para facilitar la lectura de este Capitulo, puesto que vamos a trabajar en profundidad el caso $\xi < 0$. La GPD contiene la distribución uniforme si $\kappa = 1$ y la distribución exponencial cuando $\kappa = 0$ (y tomado como límite).

3.1. Cálculo del estimador de máxima verosimilitud

Sea x_1, \dots, x_n una muestra aleatoria de la GPD y sean $x_{(1)} \leq \dots \leq x_{(n)}$ sus estadísticos de orden, la función de log-verosimilitud viene dada por

$$l(\kappa, \psi) = n \left(-\log(\psi) + (1/\kappa - 1) \frac{1}{n} \sum_{i=1}^n \log(1 - \kappa x_i / \psi) \right) \quad (3.2)$$

dónde $\psi > 0$ para $\kappa \leq 0$ y $\psi > \kappa x_{(n)}$ para $\kappa > 0$.

La función de log-verosimilitud parametrizada por κ y σ es denotada por l^* , es decir, $l^*(\kappa, \sigma) = l(\kappa, \kappa\sigma)$ y dónde $\sigma < 0$ para $\kappa \leq 0$ y $\sigma > x_{(n)}$ para $\kappa > 0$. Observamos que σ es el valor máximo de la distribución para el caso de soporte compacto. Esta reparametrización es más eficiente desde un punto de vista numérico, puesto que las variables no aparecen mezcladas al definir el dominio de definición. Indicamos que la ecuación de verosimilitud $\partial l^*(\kappa, \sigma) / \partial \sigma = 0$ es equivalente a $\kappa = \kappa(\sigma)$ dónde

$$\kappa(\sigma) = -\frac{1}{n} \sum \log(1 - x_i / \sigma) \quad (3.3)$$

Por lo tanto, la correspondiente función profile-verosimilitud viene dada por

$$l_p^*(\sigma) = n (-\log(\kappa(\sigma)\sigma) + \kappa(\sigma) - 1) \quad (3.4)$$

Calcular el MLE usando las ecuaciones de verosimilitud no siempre es posible. De hecho, el MLE como solución de las ecuaciones de verosimilitud existe siempre que $\kappa < 1$ y si $\kappa < 0,5$ entonces es asintóticamente normal y eficiente, ver Smith (1985) y Davison&Smith (1990).

Sin embargo, el MLE existe como veremos en la Sección 3.2. Aún más, proponemos utilizar el siguiente código de R, véase R Development Core Team (2010), como un algoritmo de aproximación para obtener un valor óptimo del MLE desde el punto de vista numérico.

```
#to estimate de maximum likelihood (MLE)
#of a sample x by GPD(k,psi)
eGPD<-function(x){
fk<-function(sigma) -mean(log(1-x/sigma))
```

```
fp<-function(sigma) length(x)*(-log(fk(sigma)*sigma)+fk(sigma)-1)
int<-c(-100*max(x),100*max(x))
sigma<-optimize(fp,interval=int,maximum=T)$maximum
list(k=fk(sigma),psi=fk(sigma)*sigma)}
```

Este código busca el valor de σ para maximizar la función de profile-verosimilitud (3.4). Observamos que la función `fk` y `fp` que aparecen en el código corresponde a las ecuaciones (3.3) y (3.4), respectivamente. Este código debe ser tomado como un esquema del procedimiento a seguir en el cálculo del MLE.

Nuestro método de estimación es rápido de cálculo, muy estable y no depende de los valores iniciales. Más aún, la simplicidad del código proporciona soluciones claras.

El principal punto de este código es la función de R `function: optimize` la cual busca en el intervalo desde ‘lower’ hasta ‘upper’ para un mínimo o un máximo. Esta rutina es una de las muchas rutinas de optimización que ofrece el paquete `stats`. El método utilizado es una combinación de la búsqueda de la Sección áurea y sucesiva interpolación parabólica, fue diseñada para su uso con funciones continuas. La función nunca se evalúa en dos puntos demasiado cercanos, para más información ver Brent (1973).

3.1.1. Resumen de los métodos de estimación para la GPD

Métodos tradicionales para estimar los parámetros de la GPD han sido revisados por Hosking&Wallis (1987). El más relevante método de estimación alternativo al MLE son el método de los momentos, el método de los momentos probability-weighted y el método momentos-verosimilitud considerado por Zhang (2007). Actualmente, la mejor alternativa para MLE son los métodos propuestos por Zhang&Stephens (2009) y mejorado por Zhang

3.1. Cálculo del estimador de máxima verosimilitud

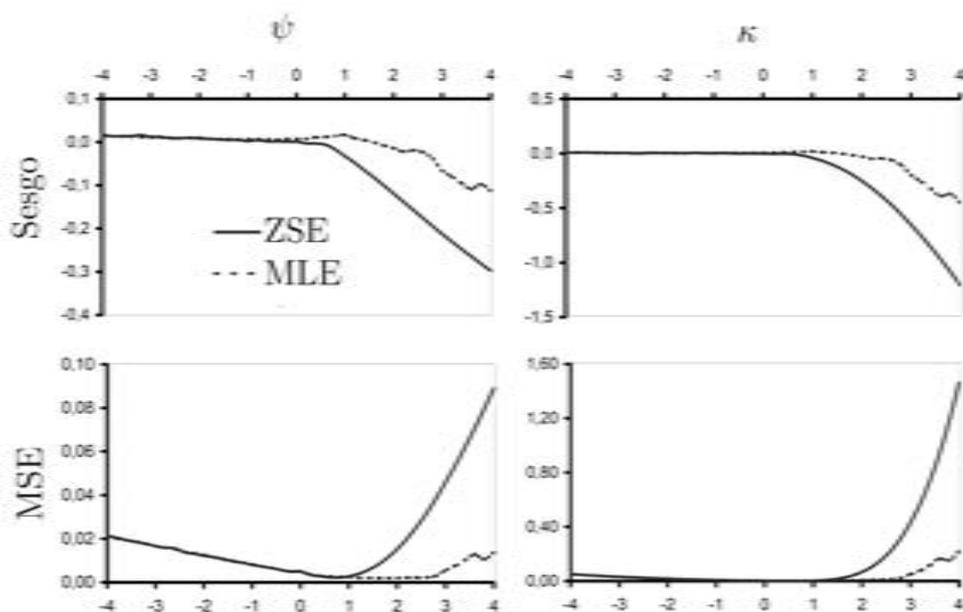


Figura 3.1: El sesgo y el MSE, para cada uno de los métodos ZSE y MLE, para cada parámetro ψ y κ de muestras de tamaño 500.

(2010) , aquí lo denotaremos por ZSE y el método propuesto por Song&Song (2012) que denotaremos por SSE.

El interés de buscar métodos de estimación alternativos al MLE para GPD se debe a los problemas de cálculo que presenta este método y al afán de corregir el sesgo de este estimador así como mejorar la eficiencia. Los problemas de cálculo se pueden resolver con una rutina apropiada, por ejemplo, la rutina anterior en R: EGPD. Por otro lado, los métodos de estimación que pretenden corregir el sesgo de MLE para GPD lo hacen localmente, es decir, sólo para un rango de los parámetros, como dice Zhang y Stephens (2009).

De hecho, ZSE también trabaja a nivel local, pero en un intervalo más amplio. ZSE exactamente muestra una mejora en el sesgo y eficiencia respecto los métodos alternativos para $-6 < \kappa < 0,5$, véase la Figura 1 en Zhang

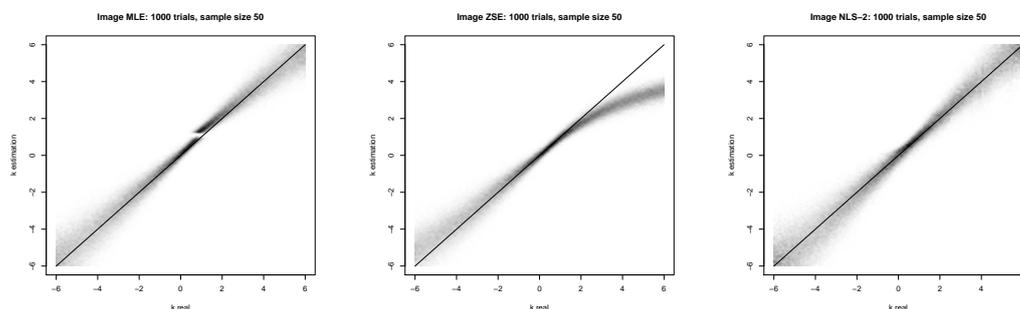


Figura 3.2: El gráfico de sombras de la estimación del parámetro κ de la GPD por cada uno de los métodos de estimación: ML, SS i ZS.

(2010). Sin embargo, la mejora con respecto a MLE en el sesgo tal vez no es relevante, ya que para κ o ψ mayor que 1, el sesgo de ZSE crece rápidamente, consulte la Figura 3.1.

ZSE mejora la eficacia si esta es medida como el ratio de la cota de Crámer-Rao y la varianza del error cuadrático medio (MSE). Sin embargo, podemos ver en la Figura 3.1 que la mejora con respecto a MLE no es significativa si se mide directamente con MSE y para $\kappa > 1,5$ o $\psi > 1$ el MLE tiene significativamente menor MSE que ZSE. La Figura 3.1 corresponde a muestras de tamaño 500.

Para muestras de tamaño pequeño, el sesgo y la eficiencia de los métodos actuales: ZSE, SSE y MLE están ilustrados con detalle en el gráfico de sombras de la Figura 3.2. Los gráficos de sombras consisten en representar con la escala de grises la frecuencia observada en cada intervalo de la partición del rango del parámetro κ , el eje horizontal representa el valor del parámetro con el cual hacemos la simulación y el eje vertical el valor estimado obtenido con cada uno de los métodos MLE, ZSE y SSE. La Figura 3.2 se obtiene para 1000 ensayos para cada ancho del intervalo de 0,1 del parámetro κ de muestras de tamaño 50 de GPD con $\psi = 1$, los valores de κ para simular son uniformemente distribuido en cada intervalo de la partición.

3.1. Cálculo del estimador de máxima verosimilitud

Como muestra la Figure 3.2, el sesgo del MLE no es muy diferente de los otros métodos propuestos. Además, verificamos que el MLE muestra problemas con muestras pequeñas y κ cerca de 1, concluimos que este tipo de muestras deberá ser analizado cuidadosamente. De hecho, la Sección 3.3 propone una nueva metodología para resolver este problema. Finalmente, observamos que ZSE no funciona fuera del rango $-6 < \kappa < 1$. Todo esto podría dar lugar a considerar como mejor método el SSE, no obstante, veremos en la Sección 3.1.2 que no es así.

Cuadro 3.1: La media (y el s.e.) del tiempo computacional en segundos.

Tamaño	10^2	10^3	10^4	10^5
ZSE	0.003 (0.006)	0.012 (0.009)	0.197 (0.016)	5.470 (0.082)
SSE	0.172 (0.099)	0.582 (0.419)	3.544 (2.260)	50.892 (28.543)
MLE	0.005 (0.006)	0.014 (0.005)	0.098 (0.017)	0.952 (0.199)

La Tabla 3.1 muestra los resultados de comparar el tiempo de computación de cada algoritmo: ZSE, SSE y nuestro algoritmo para MLE. Debido al muy corto tiempo de cálculo requerido para estimar los parámetros con una muestra de tamaño pequeño, medimos el tiempo para 100 iteraciones y calculamos el promedio, usando un AMD PC con M300 (2 GHz) y 4 GB de memoria. En la Tabla 3.1 vemos la media y el error estándar del tiempo de computación para cada método de estimación y cada medida de muestra. Los valores son obtenidos considerando 100 muestras variando κ entre -4 y 4, puesto que el tiempo depende de este valor, especialmente el método SSE.

Observamos en la Tabla 3.1 que el método SSE es más lento que los otros métodos, para valores de κ positivos es mucho más significativo. De hecho SSE es al menos 30 veces más lento que MLE. El MLE es el más rápido y eficiente desde un punto de vista computacional, ya que muestra menos variabilidad. ZSE es algo más lento que MLE para muestras pequeñas, pero para muestras grandes la diferencia se acentúa.

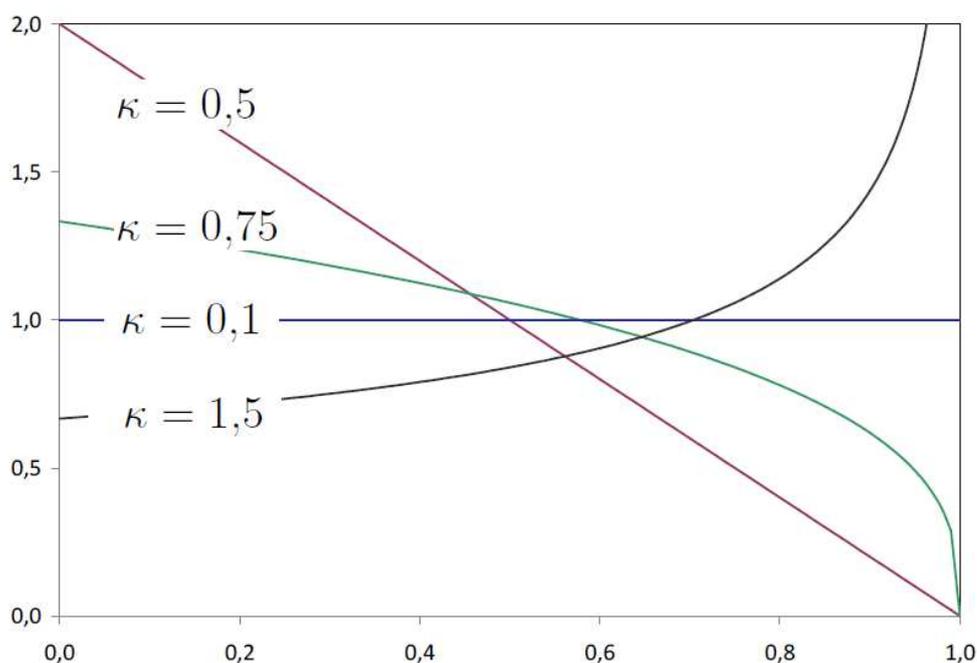


Figura 3.3: Funciones de densidad de la familia GPD para valores de $\kappa > 0$ y $\sigma = 1$.

3.1.2. Mala especificación en el modelo de GPD

El parámetro de forma de GPD incluye dos puntos críticos correspondiente a $\kappa = 0$ y $\kappa = 1$. El primero corresponde a la distribución exponencial y está bastante trabajado. Por $\kappa = 1$ corresponde a la distribución uniforme y no está suficientemente estudiado, consulte la Sección 3.2. Por lo tanto y en principio, desde el punto de vista de su calibración, la GPD debería ser considerado como tres modelos separados por la distribución exponencial y uniforme, respectivamente. En la Tabla 3.2, estos tres modelos se denotan por A, B y C, respectivamente. Resumiendo, para $\kappa < 0$ la GPD corresponde al modelo de Pareto, para $\kappa = 0$ la distribución exponencial, para $\kappa \in (0, 1)$ una familia de funciones de densidad decrecientes con soporte compacto $(0, \sigma)$, para $\kappa = 1$ el distribución uniforme y para $\kappa > 1$ una familia de densidades crecientes con soporte finito. Un problema de cualquier método de estimación

del GPD es la mala especificación entre estos tres modelos y puesto que existe una gran heterogeneidad entre ellos, este problema no puede ser ignorado.

La Tabla 3.2 muestra un resumen de la mala especificación de los métodos MLE, SSE y ZSE. Las muestras aleatorias se han generado a partir de la GPD para cada combinación de valores de los tamaños de muestra $n = 15, 25, 50, 100$ y parámetros de forma $\kappa = -0,2, -0,1, 0,1, 0,2, 0,8, 0,9, 1,1, 1,2$. El parámetro de escala se establece en 1, ya que el modelo es invariante bajo el cambio de escala. Para cada combinación de valores de n y κ , se han generado 10.000 muestras de GPD, y en la tabla vemos el porcentaje de veces que el parámetro estimado κ es negativo, entre 0 y 1, o mayor que 1. Ponemos énfasis en el hecho que el algoritmo para calcular MLE de la Sección 3.1 nos permite construir la tabla la clasificación ya que converge para todas las muestras.

En relación con el modelo de Pareto, Tabla 3.2 muestra el SSE como la mejor para muestras de tamaño pequeño, pero al aumentar el tamaño de la muestra la clasificación no mejora como sucede con otros métodos. Por ejemplo, si el tamaño de muestra es 100 y $\kappa = -0,2$, el SSE tiene 16% de mala especificación y esto es dos veces la mala clasificación del MLE para este caso.

Con respecto al caso de $\kappa > 1$, MLE clasifica mejor que los otros métodos, ya que el error de especificación es de aproximadamente 4% para cada tamaño de muestra y para $\kappa = 1,2$ es de alrededor de 2%. Por otra parte, SSE es un poco mejor que ZSE, pero SSE clasifica erróneamente como modelo de Pareto. De hecho, para los tamaño de la muestra 15, alrededor de un 15%, se clasifican por SSE como $\kappa < 0$.

Para κ entre 0 y 1, SSE y ZSE tienden a clasificarse erróneamente como caso Pareto. De hecho, más del 30% son clasificados como $\kappa < 0$ para tamaños de muestra de menos de 50 y $\kappa = 0,1$ y más del 10% para $\kappa = 0,2$. Sin embargo, MLE tiende a mal clasificar para $\kappa > 1$, de hecho, para $\kappa \geq 0,8$ más del 40% de muestras de tamaño menor de 50 son clasificados erróneamente

como $\kappa > 1$. SSE y ZSE también clasifican erróneamente como $\kappa > 1$, pero menos que el MLE.

La conclusión principal es que, dado que ningún método está preparado para garantizar la clasificación correcta entre estos tres diferentes modelos A, B y C, entonces la estimación de los parámetros se debe hacer posteriormente a la elección del modelo y más aún si se trata de muestras pequeñas, ver detalles en la Sección 3.3.

Cuadro 3.2: Los porcentajes de la clasificación de cada sub-modelo para algunos tamaños de muestra y valores de κ . El modelo A corresponde al sub-modelo de GPD para $\kappa < 0$, el modelo B para $\kappa \in [0, 1]$ y el modelo C para $\kappa > 1$. Los valores en negrita corresponden a categorías bien clasificadas.

κ	A			B						C		
	-0,1			0,1			0,9			1,1		
n	A	B	C	A	B	C	A	B	C	A	B	C
ZSE												
15	65.8	34.1	0.1	44.1	55.5	0.4	0.9	76.5	22.6	0.4	57.7	41.9
25	71.0	29.0	0.0	41.7	58.3	0.0	0.1	79.8	20.1	0.0	52.7	47.3
50	76.5	23.6	0.0	30.9	69.1	0.0	0.0	83.7	16.3	0.0	40.5	59.5
100	84.2	15.8	0.0	19.9	80.1	0.0	0.0	90.7	9.3	0.0	31.0	69.0
SSE												
10	85.3	14.5	0.3	73.8	25.6	0.7	20.7	50.0	29.4	14.7	40.7	44.6
25	73.4	26.6	0.0	54.2	45.8	0.0	3.7	61.0	35.3	1.8	40.4	57.8
50	67.5	32.6	0.0	39.4	60.6	0.0	0.1	69.3	30.6	0.0	34.4	65.6
100	70.1	30.0	0.0	29.1	70.9	0.0	0.0	77.1	22.9	0.0	27.3	72.7
MLE												
15	41.4	50.2	8.4	18.8	64.6	16.6	0.1	11,2	88.8	0.0	4.1	95.9
25	49.2	49.9	0.9	17.4	80.1	2.5	0.0	16.1	83.9	0.0	4.8	95.2
50	63.0	37.0	0.0	15.2	84.8	0.0	0.0	30.1	69.9	0.0	4.8	95.2
100	75.0	25.0	0.0	9.7	90.3	0.0	0.0	49.8	50.2	0.0	4.4	95.6

3.2. Sobre la existencia del estimador de máxima verosimilitud

El hecho de que la GPD sea principalmente usada para modelar las colas de distribuciones hace que debamos poner especial interés en clasificar correctamente el sub-modelo. En esta Sección vamos a mostrar que la existencia del MLE para la GPD no es un problema si se especifica previamente el modelo y observemos que esta previa especificación resulta muy natural cuando queremos hallar el índice de una cola, puesto que lo natural es saber de antemano si nuestros datos son de cola pesada o ligera, y por tanto de soporte finito, y en este caso saber si a la cola le corresponde una densidad de creciente o decreciente.

Para $\kappa \in (-\infty, 0]$, Castillo&Daoudi (2009) probaron que el MLE para el modelo GPD existe. Además observaron que el máximo global se toma en el interior del dominio de κ si el coeficiente de variación de la muestra es mayor que 1. Si el coeficiente de variación es menor que 1, en $\kappa = 0$ hay un máximo local y los autores insisten en que desde un punto de vista empírico, es global. Véase también Kozubowski *et al.* (2009).

Para $\kappa \geq 0$ consideramos

$$\partial l / \partial \psi = -(n/\psi) \left(1 - (1/\kappa - 1) \frac{1}{n} \sum_{i=1}^n \log(1 - \kappa x_i / \psi) \right) \quad (3.5)$$

Simplificando y añadiendo 1 a cada sumando se obtiene que $\partial l / \partial \psi = 0$ es equivalente a

$$1/(1 - \kappa) = \frac{1}{n} \sum 1/(1 - \kappa x_i / \psi). \quad (3.6)$$

Para $0 \leq \kappa \leq 1$, Choulakian&Stephens (2001) muestran que fijado $\kappa < 1$ existe una solución única para $\partial l / \partial \psi = 0$ y es un máximo que denotaremos por $\hat{\psi}(\kappa)$. A continuación, vamos a analizar la curva $(\kappa, \hat{\psi}(\kappa))$ para $\kappa \in (0, 1)$, que denominaremos curva de Choulakian-Stephens.

Proposición 3.2.1. *La función, $\hat{\psi}(\kappa)$, para $\kappa \in (0, 1)$ es continua, monótona creciente y diferenciable.*

Demostración. La continuidad y diferenciability de la función se obtiene directamente de aplicar el teorema de la función implícita a $g(\kappa, \psi) = 1/(1 - \kappa) - \frac{1}{n} \sum 1/(1 - \kappa x_i/\psi) = 0$ definida en el producto de intervalos abiertos $(0, 1) \times (\bar{x}, x_{(n)})$. Debemos probar que las derivadas parciales son diferentes de cero para cada κ . Vamos a calcularlas:

$$g_1(\kappa, \hat{\psi}(\kappa)) = \frac{1}{\kappa} \left(\frac{\kappa}{(1 - \kappa)^2} - \frac{1}{n} \sum \frac{\kappa x_i / \hat{\psi}(\kappa)}{(1 - \kappa x_i / \hat{\psi}(\kappa))^2} \right) \quad (3.7)$$

$$g_2(\kappa, \hat{\psi}(\kappa)) = \frac{1}{\hat{\psi}(\kappa)} \frac{1}{n} \sum \frac{\kappa x_i / \hat{\psi}(\kappa)}{(1 - \kappa x_i / \hat{\psi}(\kappa))^2} \quad (3.8)$$

Por supuesto, g_2 es estrictamente positiva para todo κ . Vamos a probar que $g_1 < 0$. De desarrollar $(1 - a)^2$ obtenemos la equivalencia

$$2 \frac{a}{(1 - a)^2} = \left(\frac{1}{1 - a} \right)^2 + \left(\frac{a}{1 - a} \right)^2 - 1 \quad (3.9)$$

para todo a . Por otro lado,

$$1/(1 - \kappa) = \frac{1}{n} \sum \frac{1}{1 - \kappa x_i / \hat{\psi}(\kappa)} \Rightarrow \left(\frac{1}{1 - \kappa} \right)^2 < \frac{1}{n} \sum \left(\frac{1}{1 - \kappa x_i / \hat{\psi}(\kappa)} \right)^2 \quad (3.10)$$

y

$$\kappa/(1 - \kappa) = \frac{1}{n} \sum \frac{\kappa x_i / \hat{\psi}(\kappa)}{1 - \kappa x_i / \hat{\psi}(\kappa)} \Rightarrow \left(\frac{\kappa}{1 - \kappa} \right)^2 < \frac{1}{n} \sum \left(\frac{\kappa x_i / \hat{\psi}(\kappa)}{1 - \kappa x_i / \hat{\psi}(\kappa)} \right)^2 \quad (3.11)$$

Ambas igualdades son consecuencia directa de $g(\kappa, \hat{\psi}(\kappa)) = 0$. Entonces,

$$2 \frac{1}{n} \sum \frac{\kappa x_i / \hat{\psi}(\kappa)}{(1 - \kappa x_i / \hat{\psi}(\kappa))^2} = \quad (3.12)$$

$$= \frac{1}{n} \sum \left(\frac{1}{1 - \kappa x_i / \hat{\psi}(\kappa)} \right)^2 + \frac{1}{n} \sum \left(\frac{\kappa x_i / \hat{\psi}(\kappa)}{1 - \kappa x_i / \hat{\psi}(\kappa)} \right)^2 - 1 > \quad (3.13)$$

3.2. Sobre la existencia del estimador de máxima verosimilitud

$$> \left(\frac{1}{1-\kappa}\right)^2 + \left(\frac{\kappa}{1-\kappa}\right)^2 - 1 = 2\frac{\kappa}{(1-\kappa)^2} \quad (3.14)$$

y por consiguiente $g_1 < 0$. Finalmente,

$$\hat{\psi}'(\kappa) = -\frac{g_1(\kappa, \hat{\psi}(\kappa))}{g_2(\kappa, \hat{\psi}(\kappa))} > 0 \quad (3.15)$$

para todo κ , entonces la función es creciente para todo $\kappa \in (0, 1)$, ver la Figura 3.4. □

Teorema 3.2.2. *Considera el modelo GPD para $0 \leq \kappa \leq 1$, entonces el MLE global existe. Más aún, $\bar{x} \leq \hat{\psi}(\kappa) \leq x_{(n)}$,*

$$\lim_{\kappa \rightarrow 0} \hat{\psi}(\kappa) = \bar{x} \quad \text{and} \quad \lim_{\kappa \rightarrow 1} \hat{\psi}(\kappa) = x_{(n)}. \quad (3.16)$$

Demostración. La función, $\hat{\psi}(\kappa)$ define el MLE para el parámetro ψ para cada κ fijado en $(0, 1)$ y es una función continua por la Proposición 3.2.1. Dada una muestra arbitraria, para $\kappa = 0$ y $\kappa = 1$ el MLE para el parámetro ψ es \bar{x} y $x_{(n)}$, respectivamente. Vamos a probar que los límites en (3.16) se cumplen y por tanto, $\hat{\psi}(\kappa)$ es continua sobre el conjunto cerrado $[0, 1]$. Además, el conjunto $\{(\kappa, \hat{\psi}(\kappa)) \mid 0 \leq \kappa \leq 1\}$ es la curva de Choulakian-Stephens y es un subconjunto cerrado del producto de conjuntos cerrados $[0, 1] \times [\bar{x}, x_{(n)}]$. l es una función continua y en particular, lo es sobre la curva Choulakian-Stephens, entonces en el conjunto compacto existe un máximo global.

Observamos que $\hat{\psi}(\kappa)$ está implícitamente definida por (3.6), entonces $1/(1-\kappa) = \frac{1}{n} \sum 1/(1-\kappa x_i/\hat{\psi}(\kappa))$. Si $\kappa = 1$ entonces $1/(1-\kappa x_i/\hat{\psi}(\kappa)) = 0$ para algún $i = 1 \cdots n$ por lo tanto $\hat{\psi}(1) = x_i$. Puesto que $x_i \leq \psi$ para todo $i = 1 \cdots n$, $\hat{\psi}(1) = x_{(n)}$. Observamos que

$$1/(1-\kappa) = \frac{1}{n} \sum 1/(1-\kappa x_i/\hat{\psi}(\kappa)) \quad (3.17)$$

implica

$$\kappa/(1-\kappa) - \frac{1}{n} \sum (\kappa x_i/\hat{\psi}(\kappa))/(1-\kappa x_i/\hat{\psi}(\kappa)) \quad (3.18)$$

y por consiguiente

$$1/(1 - \kappa) = \frac{1}{n} \sum (x_i/\hat{\psi}(\kappa))/(1 - \kappa x_i/\hat{\psi}(\kappa)). \quad (3.19)$$

De la última igualdad, $\hat{\psi}(0) = \frac{1}{n} \sum x_i = \bar{x}$. □

Observamos que pruebas de simulación muestran que la unicidad ocurre con probabilidad uno.

Proposición 3.2.3. *Toda solución de las ecuaciones de máxima verosimilitud para GPD satisface $\kappa < 1$.*

Demostración. El caso de GPD con soporte compacto ($\kappa > 0$) cumple $x_{(n)} < \sigma$ o equivalentemente, $\psi > \kappa x_{(n)}$, ver la función de log-verosimilitud (3.2). En particular, $1 - \kappa x_i/\psi > 0$ para todo $1 \leq i \leq n$, entonces

$$\frac{1}{n} \sum_{i=1}^n 1/(1 - \kappa x_i/\psi) > 0 \quad (3.20)$$

Por lo tanto, si la ecuación de log-verosimilitud se cumple $\partial l/\partial \psi = 0$, equivale a la ecuación (3.6), de dónde se deduce que $\kappa < 1$. □

El modelo GPD para $\kappa \geq 1$ no tiene MLE como solución de las ecuaciones de verosimilitud como consecuencia de la Proposición 3.2.3. No obstante, esto no contradice la existencia de MLE para el modelo GPD para $\kappa \geq 1$.

Proposición 3.2.4. *El MLE del modelo GPD para $\kappa \in [1, \infty)$ con κ fijado es $\hat{\psi} = \kappa x_{(n)}$. Observamos que define una nueva función $\hat{\psi}(\kappa) = \kappa x_{(n)}$.*

Demostración. Si $\kappa > 1$ está fijado, entonces la función de log-verosimilitud definida por la ecuación (3.2) depende solamente de ψ y la $dl/d\psi$, ver la función (3.5), es estrictamente negativa para cualquier $\kappa > 1$ fijado. De este modo, la función de log-verosimilitud tiene el máximo absoluto en el borde del dominio de ψ , es decir, $\hat{\psi} = \kappa x_{(n)}$. □

De este modo, la función de log-verosimilitud muestra que el MLE satisface $\hat{\sigma} = x_{(n)}$ y el problema es el valor de $\hat{\kappa}$. Este problema se resuelve añadiendo una pequeña tolerancia para el error de la estimación. Por ejemplo, soluciones convencionales para problemas no-regulares como este existen, ver Cheng&Iles (1987), Cheng&Amin (1983), Smith (1985) y Hall&Wang (2005) para enfoques metodológicos.

El valor de la función de log-verosimilitud en $(\kappa, \hat{\psi}(\kappa))$ es infinito ∞ para todo $\kappa > 1$. Para determinar un máximo para la función de log-verosimilitud sobre $E = \{(\kappa, \psi) \mid \kappa > 1, \psi = \kappa x_{(n)}\}$ la evaluaremos en un entorno de este conjunto. Fijamos un error, $\epsilon > 0$ y consideramos el conjunto $E_\epsilon = \{(\kappa, \psi) \mid \kappa > 1, \psi = \kappa(x_{(n)} + \epsilon)\}$ donde el valor de la función de log-verosimilitud es finito. Consideramos la parametrización con σ para $E_\epsilon = \{(\kappa, \sigma) \mid \kappa > 1, \sigma = x_{(n)} + \epsilon\}$ y para la función de log-verosimilitud $l_\sigma(\kappa, \sigma) = l(\kappa, \psi/\kappa)$, podemos ver que $\partial l_\sigma / \partial \kappa = 0$ equivale a $\kappa = -\frac{1}{n} \sum_{i=1}^n \log(1 - x/\sigma)$. Entonces para cada $\epsilon > 0$ el valor máximo de la función de log-verosimilitud, $l(\kappa, \psi)$, ocurre en el punto: $(-\frac{1}{n} \sum_{i=1}^n \log(1 - x/(x_{(n)} + \epsilon)), (-\frac{1}{n} \sum_{i=1}^n \log(1 - x/(x_{(n)} + \epsilon))) (x_{(n)} + \epsilon))$. Por supuesto, para menores valores de $\epsilon > 0$ obtenemos mayores valores para la función de log-verosimilitud y mayores valores para κ y ψ . Generalmente, el crecimiento de los parámetros es muy muy lento, entonces desde un punto de vista computacional parecen estacionarios.

3.3. Enfoque metodológico

Las colas exponenciales representadas en el modelo GPD por $\kappa = 0$ corresponden a las colas de las distribuciones más conocidas como la distribución normal, la distribución gamma, ver McNeil *et al.* (2005). Las colas uniformes aparecen muchas menos veces en la literatura de lo esperado. Las colas uniformes son tan difíciles de identificar como las colas exponenciales. Desde un punto de vista práctico, podemos ver que la distribución exponencial

incluida en el modelo GPD para $\kappa = 0$ y la distribución uniforme incluida para $\kappa = 1$ ejercen de atractores dentro del modelo puesto que ambas son mucho más frecuentes de lo que se consideran. Por lo tanto, dan pie a seleccionar erróneamente el modelo y por lo tanto conducen a modelizaciones con resultados no realistas.

Para introducir el nuevo enfoque metodológico permitan hacer algunas observaciones sobre propiedades de la GPD. La independencia del parámetro de escala del coeficiente de variación, ver (2.9), y la invariancia del parámetro de forma para las colas de la GPD, ver la Proposición 1.2.9, permiten considerar el coeficiente de variación empírico como la medida para detectar el umbral óptimo. La nueva metodología se basa en estas propiedades.

A continuación, introduciremos una nueva metodología para clasificar el tipo de cola. Esta metodología proporciona las herramientas para obtener *el modelo más realista para la cola de una muestra*.

1er paso: Calcular el MLE para algunos (o todos) los posibles umbrales. Por ejemplo, podemos usar el algoritmo de la Sección 3.1 para calcular el MLE.

2do(a) paso: Si el valor de $\hat{\kappa}$ estimado oscila para entre valores positivos y negativos, o simplemente, son cercanos a cero, entonces debemos considerar la posibilidad de que la cola sea exponencial. Para contrastar esta hipótesis calculamos el intervalo de confianza exacto para $\hat{\kappa}$ usando simulación basado en el coeficiente de variación. Además para contrastar esta hipótesis, podemos calcular el intervalo de verosimilitud para $\hat{\psi}$ para cada umbral y no rechazar el modelo exponencial como el modelo para la cola si para cada umbral se halla dentro.

2do(b) paso: Si el valor de $\hat{\kappa}$ varía alrededor de $\kappa = 1$ entonces considerar la posibilidad de cola uniforme. Para contrastar esta hipótesis calculamos el intervalo de confianza exacto para $\hat{\kappa}$ usando simulación basado en el coeficiente de variación. Además para contrastar esta hipótesis, podemos calcular

el intervalo de verosimilitud para $\hat{\psi}$ para cada umbral y no rechazar el modelo uniforme como el modelo para la cola si para cada umbral se halla dentro.

2nd(c) paso: En otro caso, el sub-modelo está claro. Y debería ser fijado por el investigador, especialmente si se trata de muestras de tamaño pequeño.

Observamos que el algoritmo para calcular el intervalo de confianza para κ está basado en la propiedad que el coeficiente de variación de la GPD sólo depende del parámetro de forma. De hecho, el intervalo de confianza asintótico para el coeficiente de variación se obtiene de

$$\sqrt{n}(cv(n) - c) \rightarrow N(0, \sigma^2) \quad (3.21)$$

donde la varianza es calculada por el método delta y el $cv(n)$ denota el coeficiente de variación empírico de muestras de tamaño n , ver resultados en la Tabla 3.3 para el caso de la cola uniforme ($\kappa = 1$). Más aún, para el caso $\kappa = 0$ en lugar de hacer un prueba para cada umbral, es posible hacer una prueba global, ver Castillo *et al.* (2013). Recordemos que el intervalo de confianza exacto para el MLE de ψ para la distribución exponencial está dada por

$$2n\bar{x}/\chi_{1-\alpha/2, 2n}^2 < \psi < 2n\bar{x}/\chi_{\alpha/2, 2n}^2. \quad (3.22)$$

Finalmente, el intervalo de verosimilitud para el MLE de σ viene dado por

$$x_{(n)} < \sigma < x_{(n)}/\alpha^{(1/n)}. \quad (3.23)$$

Insistimos que hay mucha diferencia entre una cola pesada y entre una función de densidad creciente o decreciente. Por lo tanto, especialmente para muestras pequeñas, es importante especificar previamente el modelo.

Con los resultados mostrados y la metodología descrita en esta Sección tenemos las herramientas suficientes para clasificar cualquier cola. Después, vamos a analizar dos muestras para ejemplificar los dos puntos críticos en el modelo GPD, $\kappa = 0$ y $\kappa = 1$. Estos ejemplos aparecen en algunos artículos como ejemplos de problemas para estimar los parámetros de GPD. La

Cuadro 3.3: Para algunos tamaños de muestra, la media y la varianza de la diferencia entre el coeficiente de variación empírico y el coeficiente de variación de la distribución uniforme ($\kappa = 1$), así como el intervalo de confianza. La última fila corresponde a los resultados asintóticos.

n	μ	σ^2	i.c. 99% para cv	
20	0.031	0.189	0.373	0.827
50	0.019	0.183	0.446	0.729
100	0.015	0.180	0.483	0.682
200	0.010	0.180	0.510	0.650
500	0.005	0.180	0.534	0.623
1000	0.005	0.178	0.547	0.608
Asintótico	0	0.178	0.577	0.577

solución de los ejemplos no estaba clara hasta ahora. Observamos que la situación en los siguientes ejemplos no son típicos, estos son casos raros pero pueden ocurrir. Más aún, para las muestras con dudas sobre la hipótesis de cola exponencial, que es trabajado por muchos autores, consideramos una nuevo y potente prueba, ver Castillo (2013).

Por último, tenga en cuenta que la metodología propuesta no es innovadora (en el caso global), los objetivos son proponer un algoritmo para calcular MLE y para resolver el conflicto en $\kappa = 0$ y $\kappa = 1$, el último caso no ha estado trabajado hasta ahora.

3.3.1. Sobre datos del oleaje en Bilbao

Para ilustrar las ventajas de la nueva metodología, vamos a presentar un ejemplo del mundo real analizada inicialmente en Castillo&Hadi (1997), que consiste en el zero-crossing períodos medios horarios (en segundos) de las olas del mar medido en el Bahía de Bilbao, España. Más tarde, este conjunto

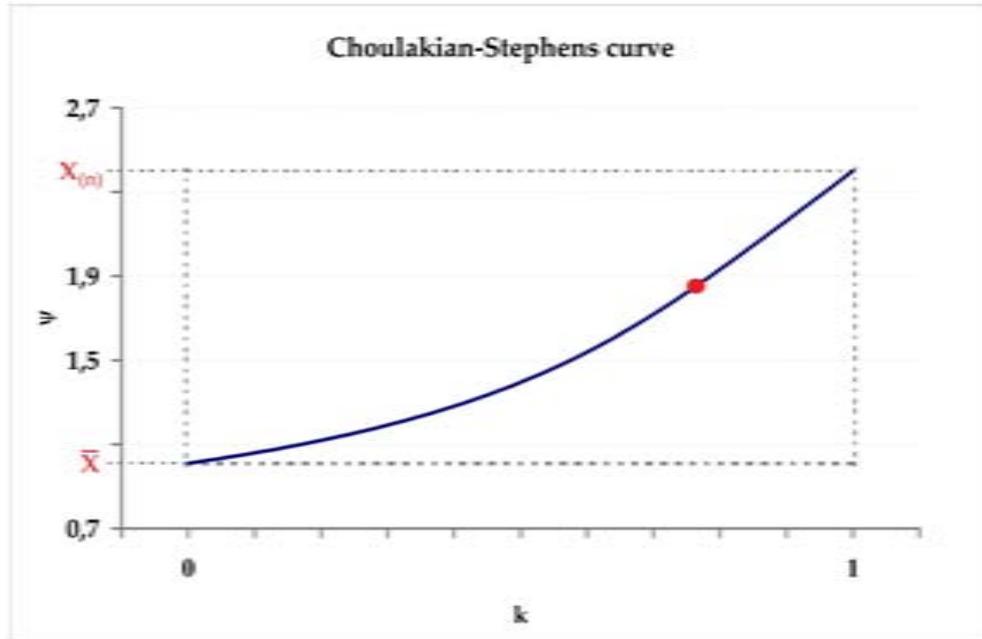


Figura 3.4: Ejemplo de la curva de Choulakian-Stephens. Datos del oleaje en Bilbao a partir de 7.5

de datos fue revisado en Luceño (2006) y en Zhang&Stephens (2009). Sólo las 197 observaciones con periodos por encima de 7 se tuvieron en cuenta. Modelamos estos datos con POT con enlace GPD utilizando umbrales en $t = 7, 7.5, 8, 8.5, 9, 9.5$ siguiendo la metodología de los autores mencionados anteriormente. Todos están de acuerdo para decir que el MLE no existe para los últimos tres casos.

Observamos que el MLE existe, como hemos visto en la Sección 3.2, pero como solución a un problema no-regular. En efecto, el MLE no es una solución de la ecuaciones de verosimilitud, sin embargo, el algoritmo de la Sección 3.1 nos proporciona una solución, consulte la Tabla 3.4. Hacemos hincapié en que Grimshaw (1993) propone un método para estimar los parámetros y obtiene los mismos resultados que en la tabla.

Cuadro 3.4: Para cada umbral de los datos de oleaje de Bilbao, la tabla proporciona la estimación de los parámetros de la GPD usando el algoritmo de la Sección 3.1. Se muestra el intervalo de verosimilitud para los niveles 95 % y 99 % para σ . En cada caso, M denota el correspondiente máximo. Los valores subrayados en el marco MLE corresponden a los casos que no existe solución de las ecuaciones de verosimilitud.

	SSE			ZSE			MLE		
	κ	ψ	$\sigma (M)$	κ	ψ	$\sigma (M)$	κ	ψ	$\sigma (M)$
7	0.84	2.4	2.9 (9.90)	0.81	2.4	3.0 (9.95)	0.86	2.5	2.9 (9.91)
7.5	0.56	1.6	2.8 (10.3)	0.71	1.8	2.5 (9.99)	0.77	1.9	2.4 (9.92)
8	0.63	1.4	2.3 (10.3)	0.77	1.5	2.0 (9.96)	0.86	1.7	1.9 (9.91)
8.5	0.77	1.2	1.5 (10.0)	0.83	1.2	1.5 (9.95)	<u>1.06</u>	<u>1.5</u>	1.4 (9.90)
9	0.80	0.8	1.0 (10.0)	0.88	0.8	0.9 (9.94)	<u>1.19</u>	<u>1.1</u>	0.9 (9.90)
9.5	-0.63	0.2	-0.4 (9.15)	1.01	0.4	0.4 (9.93)	<u>1.53</u>	<u>0.6</u>	0.4 (9.90)

El problema del MLE para los casos no-regulares está intensamente trabajado, los principales resultados están en Smith (1985) , Cheng&Amin (1983) , Cheng&Iles (1987) y Hall&Wang (2005). La Tabla 3.5 proporciona los parámetros estimados de la GPD por cada método de corrección del MLE propuesto por estos autores. Además, en cada caso, se muestra un valor correspondiente al soporte de la distribución GPD, σ , y la consecuente propuesta para el máximo oleaje, M . Podemos observar que en todos los casos , el valor de σ está dentro del intervalo de verosimilitud para la distribución uniforme. Por otra parte, para cada umbral considerado, la Tabla 3.6 proporciona el intervalo de verosimilitud para los niveles de confianza 95 % y 99 % para σ en el caso Uniforme, $\kappa = 1$.

Observar que para el MLE en la Tabla 3.4 y en cada caso no podemos rechazar la hipótesis de la uniformidad en la cola de los datos. Después de la

Cuadro 3.5: Los parámetros estimados para la GPD con los datos del oleaje en Bilbao, para cada umbral, utilizando diferentes estimadores. En todos los casos se dan los valores correspondientes para el soporte de la distribución, σ , y el máximo, M estimado por el modelo.

	Smith			Cheng&Illes			Hall&Wang		
	κ	ψ	$\sigma (M)$	κ	ψ	$\sigma (M)$	κ	ψ	$\sigma (M)$
7	0.84	2.4	2.9 (9.90)	0.86	2.5	2.9 (9.91)	0.87	2.5	2.9 (9.90)
7.5	0.77	1.9	2.4 (9.90)	0.77	1.9	2.4 (9.92)	0.80	1.9	2.4 (9.90)
8	0.83	1.6	1.9 (9.90)	0.86	1.7	1.9 (9.91)	0.87	1.7	1.9 (9.90)
8.5	0.90	1.3	1.4 (9.90)	1.02	1.4	1.4 (9.90)	0.96	1.3	1.4 (9.90)
9	0.95	0.9	0.9 (9.90)	1.13	1.0	0.9 (9.90)	1.04	0.9	0.9 (9.90)
9.5	1.07	0.4	0.4 (9.90)	1.38	0.6	0.4 (9.90)	1.22	0.5	0.4 (9.90)

metodología que hemos propuesto en la Sección 3.1, la distribución uniforme debe ser considerado el modelo más realista para la cola de los datos. En particular, este resultado muestra que para modelar el global de los datos podría considerarse un modelo truncado a la derecha, ver Proposición 2.5.1 .

3.3.2. Sobre datos del río Nidd

Estos datos han sido explorados por muchos autores desde Hosking (1987) y Davison&Smith (1990) hasta Papastathopoulos&Tawn (2012). La cuestión es decidir si la cola es exponencial o no, los últimos autores enfocan el problema cambiando el modelo de la cola.

La Tabla 3.7 muestra la variabilidad en la estimación del parámetro κ . Mirando el intervalo de confianza y teniendo en cuenta la metodología expuesta en el presente Capitulo, no podemos rechazar la cola exponencial. Por

otra parte, la cola exponencial no se rechaza para los datos mayores de 9, ya que los intervalos de confianza incluyen el MLE de parámetros.

Para probar la cola exponencial, existe otra metodología de Castillo, *et al.* (2012) y las conclusiones son las mismas. Por ejemplo, teniendo en cuenta la notación de los autores, para los datos mayor que 70 el valor del estadístico para contrastar la cola exponencial es $T_3 = 16,95$ y la cola exponencial se rechaza, pero para $T_2 = 1,42$ no se rechaza la cola exponencial y esto corresponde a los datos mayores de 90. De hecho la metodología de los autores en Castillo, *et al.* (2012) incluye el uso del CV-plot para contrastar la hipótesis de cola exponencial, ver Figura 3.5

Cuadro 3.6: Para el caso uniforme ($\kappa = 1$) y para cada umbral de los datos de oleaje en Bilbao, la tabla proporciona el intervalo de confianza exacto correspondiente y el intervalo de verosimilitud para los niveles de confianza 95 % y 99 % de κ y σ , respectivamente. En todos los casos, M denota el máximo oleaje correspondiente.

	n	e.i. para $\kappa = 1$				l.i. para $\sigma (M)$	
		95 %		99 %		95 %	99 %
7	179	0.72	1.36	0.64	1.51	2.95 (9.95)	2.98 (9.98)
7.5	154	0.70	1.41	0.62	1.55	2.45 (9.95)	2.47 (9.97)
8	106	0.64	1.50	0.55	1.72	1.96 (9.96)	1.98 (9.98)
8.5	69	0.57	1.65	0.47	1.93	1.46 (9.96)	1.50 (10.00)
9	41	0.47	1.93	0.35	2.37	0.97 (9.97)	1.01 (10.01)
9.5	17	0.27	2.83	0.13	4.09	0.48 (9.98)	0.53 (10.03)

Cuadro 3.7: El MLE de los datos de Nidd y el intervalo exacto y de confianza para los parámetros κ y ψ , respectivamente.

	MLE		n	e.i. for $\kappa = 0$		c.i. for ψ in exponential case	
	κ	ψ		95 %	99 %	95 %	99 %
70	-0.32	21.64	138	(-0.17,0.29)	(-0.23,0.4)	(26.72,37.32)	(25.43,39.48)
80	-0.34	25.22	86	(-0.17,0.26)	(-0.22,0.35)	(30.24,46.2)	(28.43,49.66)
90	-0.24	33.55	57	(-0.16,0.25)	(-0.21,0.33)	(33.91,57.11)	(31.48,62.53)
100	0.00	50.62	39	(-0.15,0.23)	(-0.2,0.31)	(37.98,71.42)	(34.78,79.9)
110	0.07	56.38	31	(-0.15,0.22)	(-0.2,0.3)	(38.09,77.44)	(34.55,88.01)
120	0.25	71.64	24	(-0.14,0.21)	(-0.19,0.28)	(39.55,88.76)	(35.47 , 102.97)
130	0.14	59.43	22	(-0.14,0.2)	(-0.18,0.27)	(35.5,82.64)	(31.7,96.63)
140	0.24	65.58	18	(-0.13,0.19)	(-0.18,0.26)	(34.68,88.48)	(30.66,105.54)

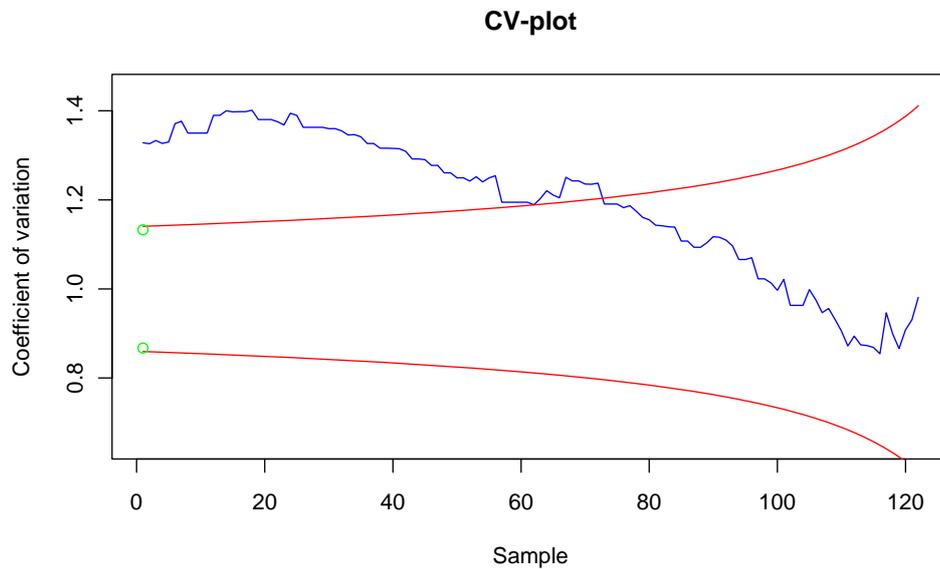


Figura 3.5: El CV-plot de los datos del río Nidd

Capítulo 4

Familia gamma truncada completada

La distribución gamma es una de las familias más estudiadas de distribuciones, ya que Fisher (1922) lo utilizó para demostrar que el MLE es más eficiente que el método de momentos. En general, para la teoría de supervivencia y fiabilidad la distribución de dos parámetros gamma truncada por la derecha es considerada desde Chapman (1956), que también hizo referencia a la dificultad de considerar el caso de origen desconocido en la gamma truncada. Den Broeder (1955) considera la distribución gamma truncada izquierda, pero con el parámetro conocido de escala. Stacy (1962) introdujo la distribución de tres parámetros gamma generalizada que incluye, como casos especiales, la gamma de dos parámetros y la distribución de dos parámetros de Weibull. Harter (1967) extiende el modelo a una familia de cuatro parámetros, mediante la inclusión de un parámetro de localización. Hedge&Dahiya (1989) obtuvieron condiciones necesarias y suficientes para la existencia del MLE de los parámetros de una distribución gamma truncada por la derecha. La distribución gamma truncada por la derecha con origen desconocido es un modelo no regular, por lo tanto, el MLE no es necesariamente el mejor método de estimación y el estimador uniforme de mínima varianza puede ser

mejor, ver Dixit&Phal (2005). Para la simulación de distribuciones truncadas por la derecha o por la izquierda, ver Philippe (1997).

En este Capitulo presentamos un nuevo modelo que muestra la relación entre la distribución de Pareto y la distribución gamma. Esto hace que este último, extendido apropiadamente, sea útil para explicar algunas anomalías que surgen en el uso práctico de la teoría de los valores extremos. Por otra parte, el hecho de que el modelo incluya la distribución de Pareto hace posible que la elección del modelo se pueda realizar con inferencia por verosimilitud.

Los resultados se aplican a ciertos fenómenos que pueden ajustarse por la distribución de ley de potencia o la distribución de Pareto, pero, al mismo tiempo, presentan una desviación de estas leyes para valores muy grandes debido a la interferencia que un límite global produce. El primer ejemplo se refiere a la influencia de la variabilidad climática en la ocurrencia de ciclones tropicales. El segundo al análisis de distribuciones de pérdidas agregadas asociadas a la modelización cuantitativa del riesgo operacional.

La nueva distribución FTG, introducida en la Sección 2, es una familia cerrada por parámetro de escala y tomando distribuciones condicionales más allá de un umbral x . Este hecho proporciona una interpretación clara de los tres parámetros α, θ, ρ . La distribución de FTG $\alpha > 0$ es la distribución gamma truncada izquierda trasladada al origen, por lo tanto, todas las distribuciones se apoyarán en $(0, \infty)$. La distribución FTG para $\alpha \leq 0$ aparece como una familia exponencial, ver Barndorff - Nielsen (1978), Brown (1986), Letac (1992).

La Sección 3 describe las propiedades estadísticas más básicas de la FTG, como los momentos función generatriz, un método de simulación y el herramientas estándar para MLE. En la Sección 4, proporcionamos aplicaciones de la FTG a ejemplos que por lo general están equipados con la distribución de Pareto. el primero en la influencia de la variabilidad climática y el calentamiento global en la ocurrencia de ciclones tropicales, ver Corral *et al.* (2010). El segundo uno en el análisis de pérdida agregada distribuciones

asociadas a la modelización cuantitativa de operativa riesgo, ver Degen, *et al.* (2007).

4.1. La distribución colas de gamma completada

La distribución de *colas de gamma completada* (*FTG*) es la familia tri-paramétrica de funciones de densidad continuas con soporte en $(0, \infty)$, definidas, por ahora, para $\nu \in \mathbb{R}, \theta > 0, \rho > 0$ por

$$f(x; \nu, \theta, \rho) = \theta (\rho + \theta x)^{\nu-1} \exp(-(\rho + \theta x)) / \Gamma(\nu, \rho). \quad (4.1)$$

dónde $\Gamma(\nu, \rho)$ es la (superior) función gamma incompleta, ver Abramowitz y Stegun (1972),

$$\Gamma(\nu, \rho) = \int_{\rho}^{\infty} t^{\nu-1} e^{-t} dt, \quad (4.2)$$

en particular $\Gamma(\nu, 0) = \Gamma(\nu)$ es la función gamma. La distribución FTG extiende para algunos valores de los parámetros en el borde tal y como veremos en esta Sección.

Si $\nu > 0, \theta > 0$ y $\rho = 0$, la familia (4.1) claramente extiende a la función de densidad de la distribución gamma, definida por

$$g(x; \nu, \theta) = \theta^{\nu} x^{\nu-1} \exp(-\theta x) / \Gamma(\nu) \quad (4.3)$$

Para $\nu > 0$ la FTG es la distribución gamma truncada izquierda relocalizada al origen, es decir, la familia de los excesos de la distribución gamma. Recordemos que el término *cola* es usado en el sentido de los excesos sobre un umbral. La Proposición 4.1.1 establece exactamente lo que estamos diciendo.

Proposición 4.1.1. *Si $\nu > 0$ and $\rho > 0$, entonces (4.1) es la función de densidad de los excesos de la distribución gamma en el umbral $\sigma > 0$, con $\sigma = \rho/\theta$.*

$$g_{\sigma}(x; \nu, \theta) = \theta^{\nu} (x + \sigma)^{\nu-1} \exp(-\theta(x + \sigma)) / \Gamma(\nu, \sigma\theta) \quad (4.4)$$

Demostración. Sea $G(x)$ la función de distribución correspondiente a $g(x; \nu, \theta)$, ver (4.3). Consideramos $\sigma > 0$. Usando $dt = \theta dx$, obtenemos

$$1 - G(\sigma) = \int_{\sigma}^{\infty} \theta^{\nu} x^{\nu-1} \exp(-\theta x) dx / \Gamma(\nu) = \int_{\sigma\theta}^{\infty} t^{\nu-1} \exp(-t) dt / \Gamma(\nu) \quad (4.5)$$

Por lo tanto, la función de densidad de los excesos de $g(x; \nu, \theta)$ en σ es (4.4), dado que (4.2). Finalmente, para $\sigma = \rho/\theta$ obtenemos (4.1). \square

La distribución FTG está relacionada con el modelo de tri-paramétrico Pareto tipo III from Arnold (1983), caracterizado por la función de supervivencia

$$\bar{F}(t) = (1 + t/\phi)^{-\lambda} \exp(-\theta t). \quad (4.6)$$

Más aún, la Pareto tipo III se considera un modelo para datos de supervivencia, Davis (1979). Este hecho resulta natural dado que el modelo Pareto tipo III es la mixtura de dos modelos FTG, ver Proposición 4.1.2.

Proposición 4.1.2. *Sea $p_{III}(\theta, \lambda, \phi)$ la función de densidad de la Pareto tipo III y f la función de densidad de la FTG(ν, θ, ρ), ver (4.1) entonces*

$$p_{III}(\theta, \lambda, \phi) = c f(-\lambda, \theta, \theta\phi) + (1 - c) f(-\lambda + 1, \theta, \theta\phi) \quad (4.7)$$

dónde $c = \lambda(\theta\phi)^{\lambda} \exp(\theta\phi) \Gamma(-\lambda, \theta\phi)$.

Demostración. La función de densidad p_{III} es menos la derivada de la función de densidad (4.1). Para verificar (4.7) usamos la propiedad

$$-a\Gamma(a, b) + \Gamma(a + 1, b) = \exp(-b)b^a$$

para todo a, b . Esta propiedad es obtenida de la derivación por partes aplicada a la definición de $\Gamma(a, b)$, ver (4.2). \square

Observamos que el modelo Pareto tipo III es un caso particular del modelo seis-paramétrico que obtenemos de la mixtura de dos modelos FTG.

4.1.1. FTG como familia exponencial

Para σ fijado, el modelo Pareto es un modelo exponencial completo con estadístico suficiente $\log(1 + x/\sigma)$.

La familia gamma de funciones de densidad definidas por (4.3) es un modelo exponencial completo sobre $(0, \infty)$, generado por la medida de Lebesgue en $(0, \infty)$ con estadístico suficiente $(x, \log(x))$ y el dominio de parámetros es exactamente, $\{(\nu, \theta) : \nu > 0, \theta > 0\}$.

Asimismo, para $\sigma > 0$ fijado, la FTG corresponde al modelo exponencial completo generado por la medida de Lebesgue en $(0, \infty)$ con estadístico suficiente $T(x) = (x/\sigma, \log(1 + x/\sigma))$. Lo inesperado es que el dominio natural de parámetros de la FTG es mayor que el correspondiente a la extensión del modelo gamma.

Proposición 4.1.3. *Sea $\sigma > 0$ fijado en (4.4), la transformada de Laplace del estadístico suficiente $(x, \log(x + \sigma))$*

$$C(\theta, \nu - 1) = \int_0^\infty \exp(-\theta x + (\nu - 1) \log(x + \sigma)) dx$$

converge para el conjunto de parámetros dado por $\Theta = \{\nu \in \mathbb{R}, \theta > 0\}$ y para $\Theta_0 = \{\nu < 0, \theta = 0\}$.

Demostración. En primer lugar, asumimos $\theta > 0$ y tomamos $y = x + \sigma$,

$$C(\theta, \nu - 1) = e^{\theta \sigma} \int_\sigma^\infty \exp(-\theta y) y^{\nu-1} dy = \theta^{-\nu} e^{\theta \sigma} \Gamma(\nu, \theta \sigma)$$

la integral converge para $\nu \in \mathbb{R}$, dado que no hay singularidades en el cero y la exponencial domina cualquier función de potencia para valores grandes de x .

Si $\theta = 0$, entonces para $\nu < 0$ la siguiente integral también converge

$$C(0, \nu - 1) = \int_0^\infty (x + \sigma)^{\nu-1} dx = -\sigma^\nu / \nu$$

□

La Proposición 4.1.3 muestra que dado $\sigma > 0$ el dominio natural de parámetros para (4.4) es $D = \Theta \cup \Theta_0$. La distribución FTG (4.1) es la extensión de (4.4) para $\nu < 0$ y se ve que extiende para los parámetros en el borde Θ_0 .

Finalmente, dada una muestra $\mathbf{x} = \{x_1, \dots, x_n\}$ para calcular el MLE podemos usar el procedimiento: determinar σ y si $t(\mathbf{x})$ está en el interior del dominio de las medias, usarlo para resolver las ecuaciones de verosimilitud. Por lo tanto, debemos determinar el dominio de las medias.

Proposición 4.1.4. *El dominio de las medias de la FTG es*

$$\{(x, y) ; x > 0, \log(1+x) > y > x/(1+x)\}$$

para todo parámetro ν, θ y ρ .

Demostración. El modelo FTG es invariante por cambios de escala, por lo tanto es suficiente considerar $\rho/\theta = \sigma = 1$ y siguiendo los resultados de Castillo&Puig (1999) escribimos

$$T(x) = (x, \log(1+x)) = (x, S(x)).$$

Dado que $-S(x)$ es una función convexa, $\Theta = \mathbb{R}^+ \times \mathbb{R}$ y $\Theta_0 = \{0\} \times \mathbb{R}^+$, entonces por la Proposición 4 en Castillo&Puig (1999)

$$S(\mu) \geq E(S(X)) \geq E(S(X_\mu^0)) \quad (4.8)$$

dónde X_μ^0 es una variable aleatoria con función de densidad de la forma $f(x; b) = \exp(-(b)S(x))/C(b)$, $b > 0$, tal que $E(X_\mu^0) = \mu$. Dada una muestra $\mathbf{x} = \{x_1, \dots, x_n\}$ de distribución FTG, las ecuaciones de verosimilitud se pueden describir en términos del estadístico suficiente (\bar{x}, \bar{s}) por

$$E(X) = \frac{\partial k}{\partial a}(\theta, \beta) = \bar{x}, \text{ and } E(S(X)) = \frac{\partial k}{\partial p}(\theta, \beta) = \bar{s},$$

dónde $\beta = -(\nu - 1) > 1$. Consideramos una v.a. X^0 con distribución de Pareto y la función de densidad es:

$$p(x; \beta, 1) = \beta (1+x)^{-\beta-1} = \exp(-\beta S(x))/C(\beta)$$

entonces

$$E(X^0) = \int_0^\infty \beta x (1+x)^{-\beta-1} dx = \frac{1}{\beta-1}.$$

Para cada muestra, existe $\beta > 1$ tal que

$$E(X^0) = \frac{1}{\beta-1} = \bar{x} \Leftrightarrow \beta = \frac{\bar{x}+1}{\bar{x}}$$

Finalmente, el dominio de las medias descrito en (4.8) corresponde a

$$\log(1+\bar{x}) \geq \bar{s} \geq \frac{\bar{x}}{1+\bar{x}}$$

dado que $E(S(X_{\bar{x}}^0)) = \int_0^\infty \beta \log(1+x) (1+x)^{-\beta-1} dx = 1/\beta$ para todo β . \square

4.1.2. Modelo FTG y Pareto como modelos para colas

Insistimos que la familia Pareto y la power-law no son la misma. Observamos que la función de densidad de la power-law tiene soporte variable, en cambio, el modelo Pareto, el soporte está fijado en $(0, \infty)$. Una muestra de distribución Pareto consiste en números positivos, pero una muestra de power-law consiste en números reales mayores que σ .

Para $\nu \leq 0$, la distribución FTG, (4.1), no ha sido estudiada, para nuestro conocimiento, y procederemos a hacerlo.

Teorema 4.1.5. *Sea $\sigma = \rho/\theta > 0$ fijado en (4.1) y $\nu < 0$. Si ρ tiende a cero, entonces la función de densidad (4.1) tiende a la función de densidad de la Pareto en norma L^1 . Más aún, la convergencia extiende a los momentos, siempre que los correspondientes momentos para la distribución de Pareto sean finitos.*

Demostración. Observamos que si ρ tiende a 0, entonces θ tiende a 0, puesto que $\sigma = \rho/\theta > 0$ está fijado. Usando $\theta = \rho/\sigma$,

$$f(x; \nu, \theta, \rho) = \rho^\nu \sigma^{-1} (1+x/\sigma)^{\nu-1} \exp(-\rho(1+x/\sigma)) / \Gamma(\nu, \rho)$$

converge puntualmente a la función de densidad de la Pareto, dado que la propiedad (5.1.23) de Abramowitz&Stegun (1972), bajo la supuestos,

$$\lim_{\rho \rightarrow 0} \rho^\nu / \Gamma(\nu, \rho) = -\nu \quad (4.9)$$

se cumple. Observamos que para ρ pequeño

$$\begin{aligned} f(x; \nu, \theta, \rho) &= \rho \sigma^{-1} (1 + x/\sigma)^{\nu-1} \exp(-\rho(1 + x/\sigma)) / \Gamma(\nu, \rho) \\ &\leq -2\nu \sigma^{-1} (1 + x/\sigma)^{\nu-1} = 2p(x; -\nu, \sigma) \end{aligned}$$

puesto que del límite (4.9) podemos considerar la acotación $\rho^\nu / \Gamma(\nu, \rho) \leq -2\nu$. Finalmente, del teorema de convergencia dominada obtenemos la convergencia en L^1 . Más todavía, cuando los momentos de la distribución de Pareto son finitos, la convergencia se extiende a estos momentos. Observamos que $\nu = -\alpha$. \square

De hecho la (4.1) también puede verse como una versión *weighted* de la distribución de Pareto, con el peso $w(x) = \exp(-\theta x)$, esto también se conoce como la *exponencial tilting* de la distribución, ver Barndorff-Nielsen and Cox (1994).

En resumen, la familia (4.1) ha sido extendida al borde del conjunto de parámetros correspondiente a la distribución gamma y la distribución de Pareto, $\{\nu > 0, \theta > 0, \rho = 0\}$ y $\{\nu < 0, \theta = 0, \sigma > 0\}$, respectivamente. Por consiguiente, podemos asumir que la familia (4.1) incluye la distribución gamma, la distribución de los excesos de gamma ($\nu > 0$), su extensión a modelo exponencial completo ($\nu < 0$) y la distribución de Pareto, ver Figura 4.1. El hecho de que incluya excesos de una familia muy utilizada en modelización de colas como es la familia gamma y que incluya la distribución de Pareto, el modelo fundamental para modelar colas pesadas, hace que consideremos la FTG como una nueva familia apropiada para modelar colas pesadas, esto quedará justificado más rigurosamente y de un modo más genérico en el Capítulo 5.

Proposición 4.1.6. *La función de distribución correspondiente a la familia (4.1) es*

$$F(x; \nu, \theta, \rho) = 1 - \Gamma(\nu, \rho + \theta x) / \Gamma(\nu, \rho)$$

y para la distribución de Pareto debemos considerar el caso límite, correspondiente a $P(x; \alpha, \sigma) = 1 - (1 + x/\sigma)^{-\alpha}$ con $\alpha = -\nu$.

Demostración. Para $\theta > 0$, tomamos $y = \rho + \theta t$

$$F(t; \nu, \theta, \rho) = \int_0^{\rho+\theta t} y^{\nu-1} \exp(-y) dy / \Gamma(\nu, \rho) = 1 - \Gamma(\nu, \rho + \theta t) / \Gamma(\nu, \rho)$$

y observamos que para $\rho = 0$ se cumple dado que $\Gamma(\nu, \rho) = \Gamma(\nu)$ y que esto corresponde a la distribución gamma. \square

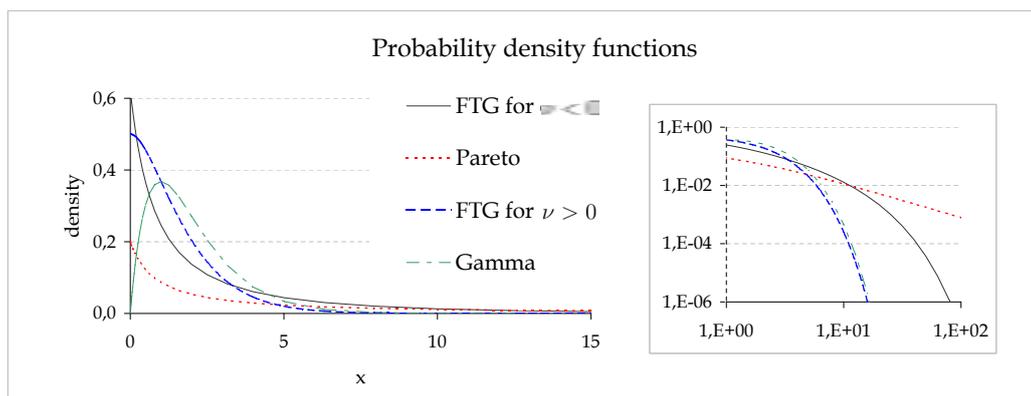


Figura 4.1: La figura de la izquierda muestra algunas funciones de densidad de la familia FTG. La FTG con $\nu = 2$, $\sigma = 1$ y $\theta = 1$ corresponde a los excesos de una distribución gamma y la FTG para $\nu = -0,2$, $\sigma = 1$ y $\theta = 0,1$ no. Para el conjunto de parámetros en el borde de su dominio, hemos considerado la distribución gamma con $\nu = 2$ y $\theta = 1$ y la distribución Pareto con $\nu = -0,2$ y $\sigma = 1$. La figura de la derecha muestra el mismo gráfico en escala logarítmica de las colas de dichas distribuciones. Podemos ver el decaimiento exponencial excepto para el caso Pareto.

Proposición 4.1.7. *Sea X una v.a. con distribución $FTG(\nu, \theta, \rho)$, entonces*

- *Para $\lambda > 0$, la v.a. λX tiene distribución $FTG(\nu, \theta/\lambda, \rho)$.*
- *Para cualquier umbral, $u > 0$, los excesos, X_u , tienen distribución $FTG(\nu, \theta, \rho + \theta u)$.*

Demostración. El primer resultado se deduce observando la función de densidad de λX para $\lambda > 0$,

$$f(x/\lambda; \nu, \theta, \rho) / \lambda \tag{4.10}$$

insistimos que para $\theta = 0$ y $\nu = -\alpha$ es

$$p(x/\lambda; \alpha, \sigma) / \lambda = \alpha(\lambda\sigma)^{-1} (1 + x/(\lambda\sigma))^{-\alpha-1} = p(x; \alpha, \lambda\sigma).$$

Y la segunda se deduce de la definición. Para $\theta > 0$ es

$$\frac{f(x+u; \nu, \theta, \rho)}{1 - F(u)} = \frac{\theta(\rho + \theta u + \theta x)^{\nu-1}}{\Gamma(\nu, \rho + \theta u)} \exp(-(\rho + \theta u + \theta x)) = f(x; \nu, \theta, \rho + \theta u)$$

y para $\theta = 0$ y $\nu = -\alpha$ es

$$\frac{p(x+u; \alpha, \sigma)}{1 - P(u)} = \frac{\alpha(1 + x/(\sigma + u))^{-\alpha-1}}{(\sigma + u)^{-1}} = p(x; \alpha, \sigma + u).$$

□

De la Proposición se deduce que la distribución FTG además de ser una familia con parámetro de escala, también es cerrada por truncamiento, o excesos. Esto son propiedades que hace que la consideremos apropiada para modelar colas y además nos proporciona una interpretación de los parámetros. El parámetro $\beta = 1/\theta$ es el parámetro de escala y el parámetro ρ es el parámetro que caracteriza el truncamiento o excesos. El parámetro ($\nu = -\alpha$) podría ser interpretado en términos de la distribución de Pareto como el peso de la cola, puesto que el modelo incluye la incluye como caso límite. Así, cada uno de los tres parámetros separadamente tiene una clara interpretación.

4.2. Herramientas estadísticas y MLE

Con el actual *software* de manipulación numérica especializados para el análisis estadístico no es difícil trabajar con la distribución FTG. La función gamma incompleta, $\Gamma(\nu, \rho)$ y sus derivadas son fácilmente accesibles. La diferenciación simbólica nos permite obtener los momentos de una distribución a partir de la función generatriz. Algoritmos de simulación y optimización también están disponibles. El trabajo de pioneros como Chapman (1956) debe ser visto de esta manera.

4.2.1. Momentos y función generatriz

La distribución FTG tiene función generatriz de momentos en el interior del dominio de parámetros. Por lo tanto, es posible calcular los momentos de todos los órdenes. Además, también es posible calcular los momentos de la distribución de los excesos sobre un umbral, por la Proposición 4.1.2.

Proposición 4.2.1. *Para $\nu \in \mathbb{R}, \theta > 0, \rho > 0$, la función generatriz de momentos de la distribución FTG, (4.1), existe y viene dada por*

$$M(t) = (1 - t/\theta)^{-\nu} \exp(-\rho t/\theta) \Gamma(\nu, \rho(1 - t/\theta)) / \Gamma(\nu, \rho), \quad t < \theta. \quad (4.11)$$

Para $\nu > 0$, se extiende para $\rho = 0$ y coincide con la función generatriz de momentos de la distribución gamma $M_g(t) = (1 - t/\theta)^{-\nu}$

Demostración. Por definición $M(t) = E[\exp(tX)]$ viene dada por

$$M(t) = \int_0^\infty \frac{\theta(\rho + \theta x)^{\nu-1} \exp[-(\rho + (\theta - t)x)]}{\Gamma(\nu, \rho)} dx$$

y usando $\rho + \theta dx - t\rho/\theta - tdx = dy$, obtenemos

$$M(t) = (1 - t/\theta)^{-\nu} \exp(-\rho t/\theta) \Gamma(\nu, \rho(1 - t/\theta)) / \Gamma(\nu, \rho)$$

para $t < \theta$. □

La función generatriz de cumulantes viene dada por

$$K(t) = \log(M(t)) = -t\rho/\theta - \nu \log(1 - t/\theta) - \log \Gamma(\nu, \rho) + \log \Gamma(\nu, (1 - t/\theta)\rho)$$

así, los primeros momentos son,

$$\begin{aligned} E[X] &= K'(0) = (\nu - \rho + \mu)/\theta \\ \text{Var}[X] &= K''(0) = (\nu + (1 + \rho - \nu)\mu - \mu^2)/\theta^2 \end{aligned}$$

dónde $\mu = e^{-\rho}\rho^\nu/\Gamma(\nu, \rho)$. Notamos que usando la Proposición 4.1.2, para calcular la esperanza condicionada a cualquier umbral fijado $u > 0$ es la misma que calcular la esperanza con la correspondiente modificación de los parámetros

$$E[X | X > u] = (\nu - \rho + \mu')/\theta \quad (4.12)$$

donde $\mu' = e^{-(\rho+\theta u)}(\rho + \theta u)^\nu/\Gamma(\nu, \rho + \theta u)$.

4.2.2. Generación de variables aleatorias con distribución FTG

Los métodos de simulación para las distribuciones de Pareto y gamma son bien conocidas. También se ha estudiado bien la simulación de la distribución gamma truncada (4.4), ver Philippe (1997). Por lo tanto, sólo trataremos los casos del conjunto de parámetros $\{\nu < 0, \theta > 0, \rho > 0\}$ para la distribución FTG.

Una manera simple para simular la distribución es el método de inversión, ya que la función de distribución tiene una expresión fácil, ver Proposición 4.1.6, sin embargo, se tiene que utilizar complejos procesos numéricos usando la función gamma incompleta que lo hacen un método poco práctico.

Un método simple y eficiente desde el punto de vista numérico es obtenido de una idea de Devroye (1986) sobre una generalización del método de rechazo. Hacemos hincapié en la simplicidad de este algoritmo, ya que no requiere el uso de la gamma incompleta función. Vamos a dar algunos detalles.

Primero de todo, dado que $1/\theta$ es el parámetro de escala, es suficiente considerar simulaciones para $\theta = \rho$. Es decir, para simular $FTG(\nu, \theta, \rho)$, podemos simular $FTG(\nu, \rho, \rho)$ y finalmente aplicar el cambio de escala a la muestra obtenida.

Para $\theta = \rho$, la función de densidad (4.1) se divide en tres términos

$$f(x; \nu, \rho, \rho) = (\rho^{\nu-1} e^{-\rho} / \Gamma(\nu, \rho)) (\rho e^{-\rho x}) (1+x)^{\nu-1} = cg(x)\psi(x) \quad (4.13)$$

donde la función $\psi(x) = (1+x)^{\nu-1}$ es $[0, 1]$ -valorada, $g(x) = \rho e^{-\rho x}$ es una función de densidad sencilla de simular y c es una constante de normalización menor o igual a 1.

El método de rechazo para este caso puede escribirse como sigue. Generamos variables aleatorias independientes (X, U) dónde X tiene función de densidad $g(x)$ y U es uniformemente distribuida en $[0, 1]$ hasta lograr $U \leq \psi(X)$. Este método produce una v.a. X con función de densidad $f(x)$, (Devroye, 1986).

El siguiente código, en lenguaje R, aplica el método de Devroye a nuestro caso, ver R Development Core Team (2010).

```
#to generate a sample of size n of FTG(g,t,r)
rFTG<-function(n,g,t,r) {
  sample<-c(); m<-0
  while (m<n) {
    x<-rexp(1,rate=r);u<-runif(1)
    if (u<=(1+x)^(g-1)) sample[m+1]<-x
    m<-length(sample) }
  sample*r/t }
```

4.2.3. Estimador de máxima verosimilitud de la FTG

En (4.1), la distribución FTG ha sido introducida con parámetros (ν, θ, ρ) , dado que caa uno separadamente tiene una clara interpretación. Para la esti-

mación de los parámetros es mejor usar (ν, σ, ρ) , con parámetro de dispersión $\sigma = \rho/\theta$, dado que para σ fijado la familia FTG es un modelo exponencial y por lo tanto, de Barndorff-Nielsen (1978), sabemos que el MLE existe y es único. El resumen del procedimiento para calcular el MLE ya lo hemos introducido en la Sección 4.1.1 y consiste en hallar el parámetro de dispersión y entonces optimizar el problema resultante que sabremos que tiene solución.

Sea $\{x_1, \dots, x_n\}$ una muestra de tamaño n , la función de log-verosimilitud para la distribución FTG es

$$l(\nu, \sigma, \rho) = \tag{4.14}$$

$$-n \left(\log \Gamma(\nu, \rho) + \log(\sigma \rho^{-\nu}) - \frac{\nu-1}{n} \sum_{i=1}^n \log\left(1 + \frac{x_i}{\sigma}\right) + \frac{\rho}{n} \sum_{i=1}^n \left(1 + \frac{x_i}{\sigma}\right) \right)$$

Para simplificar, denotamos

$$d = d(\nu, \rho) = \log \Gamma(\nu, \rho) \tag{4.15}$$

y consideramos (r, s) el estadístico suficiente del modelo exponencial para σ fijado como

$$r(x; \sigma) = (1 + x/\sigma) \quad \text{and} \quad s(x; \sigma) = \log(1 + x/\sigma)$$

entonces denotamos las correspondientes medias muestrales como

$$\bar{r}(\mathbf{x}; \sigma) = \frac{1}{n} \sum_{i=1}^n (1 + x_i/\sigma) \quad \text{and} \quad \bar{s}(\mathbf{x}; \sigma) = \frac{1}{n} \sum_{i=1}^n \log(1 + x_i/\sigma)$$

Para simplificar, usaremos los parámetros como subíndices para denotar las derivadas parciales y omitiremos la dependencia de los parámetros en la expresión de estas derivadas. Así, la *scoring* es $(l_\nu, l_\sigma, l_\rho)$ y viene dada por

$$l_\nu = -n \{d_\nu - \log(\rho) - \bar{s}(\mathbf{x}; \sigma)\} \tag{4.16}$$

$$l_\sigma = -n \{\sigma^{-1} - (\nu - 1)\bar{s}_\sigma + \rho\bar{r}_\sigma\} \tag{4.17}$$

$$l_\rho = -n \{d_\rho - \nu\rho^{-1} + \bar{r}(\mathbf{x}; \sigma)\} \tag{4.18}$$

la matriz de información observada viene dada por

$$I_O(\nu, \sigma, \rho) = -n \begin{pmatrix} d_{\nu\nu} & -\bar{s}_\sigma & d_{\nu\rho} - \rho^{-1} \\ -\bar{s}_\sigma & -\sigma^{-2} - (\nu - 1)\bar{s}_{\sigma\sigma} + \rho\bar{r}_{\sigma\sigma} & \bar{r}_\sigma \\ d_{\nu\rho} - \rho^{-1} & \bar{r}_\sigma & d_{\rho\rho} + \nu\rho^{-2} \end{pmatrix}$$

y puede ser usada para calcular el intervalo de confianza para el MLE $\hat{\nu}$, $\hat{\sigma}$ and $\hat{\rho}$ de los parámetros ν , σ and ρ , respectivamente. De hecho, la matriz de información de Fisher esperada puede calcularse de una forma cerrada como veremos en esta Sección.

Para calcular el MLE es conveniente resolver la ecuación (4.17) para obtener $\hat{\sigma}$ usando $(\hat{\nu}(\sigma), \hat{\rho}(\sigma))$ para los parámetros (ν, ρ) o, más general, para maximizar la profile-ecuación de log-verosimilitud

$$l_p(\sigma) = \tag{4.19}$$

$$-n \left(\log \Gamma(\hat{\nu}(\sigma), \hat{\rho}(\sigma)) + \log(\sigma \hat{\rho}(\sigma)^{-\hat{\nu}(\sigma)}) - (\hat{\nu}(\sigma) - 1)\bar{s}(x, \sigma) + \hat{\rho}(\sigma)\bar{r}(x, \sigma) \right)$$

donde $(\hat{\nu}(\sigma), \hat{\rho}(\sigma))$ es la única solución del sistema de variables (ν, ρ) que consiste de las ecuaciones (4.16) y (4.18) para σ fijado, dado que FTG es un modelo exponencial en este caso. Observamos que desde un punto de vista práctico, es conveniente considerar esta pareja de ecuaciones para simplificar (4.2.3) (or (4.17)) en una ecuación que haga referencia lo menos posible a la muestra explícitamente. Para hacer esto, necesitaremos considerar las relaciones (4.22) encontradas en la Sección 4.2.4. Por ejemplo, la ecuación (4.2.3) puede ser simplificada por

$$l_p(\sigma) = \tag{4.20}$$

$$-n \left(\log \Gamma(\hat{\nu}(\sigma), \hat{\rho}(\sigma)) - \log(\hat{\rho}(\sigma)\sigma^{-1}) - (\hat{\nu}(\sigma) - 1)d_\nu - \hat{\rho}(\sigma)d_\rho + \hat{\nu}(\sigma) \right)$$

observamos que es una expresión en la que no aparece de forma explícita la muestra.

Un procedimiento para obtener el MLE en R es calcular el MLE de la muestra estandarizada $y = \{x_i/\bar{x}\}_{1 \leq i \leq n}$, considerando el valor inicial como detallamos a continuación. Tenemos dos opciones para la semilla, como $(\dot{\nu}, 1, \dot{\theta})$ dónde $(\dot{\nu}, \dot{\theta})$ es el MLE del modelo gamma o tomarla como $(\dot{\nu}, \dot{\sigma}, \dot{\rho})$ dónde $(-\dot{\nu}, \dot{\sigma})$ es el MLE del modelo Pareto y $\dot{\rho}$ se obtiene por la relación, de la ecuación (4.18),

$$d_\rho - \dot{\nu}\dot{\rho}^{-1} + 1 + \dot{\sigma}^{-1} = 0.$$

Finalmente, $(\hat{\nu}, \hat{\sigma}, \hat{\rho})$ (el MLE para la muestra x) se obtiene usando la Proposición 4.1.2, en efecto, obtenemos $\hat{\nu} = \hat{\nu}'$, $\hat{\sigma} = \hat{\sigma}'/\bar{x}$ and $\hat{\rho} = \hat{\rho}'$.

Algunos comentarios adicionales deben ser considerados. por ejemplo, podría ser conveniente considerar la escala logarítmica para el parámetro σ y ρ . R tiene un paquete para optimizar el cual simplifica el algoritmo gratamente. Si usamos Maple, entonces debemos ser cuidadosos con las muestras de tamaño muy grande.

4.2.4. Matriz de información de Fisher esperada

La matriz de información de Fisher

$$I_E(\nu, \sigma, \rho) = E(I_O(\nu, \sigma, \rho)) \quad (4.21)$$

puede ser calculada por una forma cerrada usando el valor esperado de las relaciones

$$\begin{aligned} r_\sigma &= \sigma^{-1}(1 - r) \\ r_{\sigma\sigma} &= -\sigma^{-2}(1 - r) \\ s_\sigma &= -\sigma^{-1}(1 - r^{-1}) \\ s_{\sigma\sigma} &= \sigma^{-2}(1 - r^{-2}) \end{aligned} \quad (4.22)$$

dado que para todo $k \in \mathbb{R}$,

$$E(r^k) = \rho^{-k}\Gamma(\nu + k, \rho)/\Gamma(\nu, \rho)$$

Complementarias expresiones para I_E pueden ser obtenidas usando otras relaciones, por ejemplo,

$$\rho E(r) + (\nu - 1)E(r^{-1}) = \rho + \nu$$

la cual es obtenida de $\Gamma(\nu + 1, \rho)/\Gamma(\nu, \rho) = \nu - \rho d_\rho$ o equivalentemente, $\Gamma(\nu, \rho) = (\nu - 1)\Gamma(\nu - 1, \rho) + \rho^{\nu-1} \exp(-\rho)$, esto es una consecuencia de aplicar integración por partes a la expresión integral (4.2).

Observamos que algunos términos de la matriz I_E tienen una expresión explícita, dado que son exactamente los términos en la matriz de información observada. Denotamos la integral integral (4.2) por

$$\Gamma = \Gamma(\nu, \rho) = \rho^\nu E_{1-\nu}(\rho), E_n(\rho) = \int_1^\infty t^{-n} e^{-\rho t} dt,$$

y las primeras derivadas pueden expresarse por

$$d_\rho = \Gamma_\rho/\Gamma \quad \text{and} \quad d_\nu = \Gamma_\nu/\Gamma$$

y las segundas por

$$\begin{aligned} d_{\rho\rho} &= \Gamma_{\rho\rho}/\Gamma - (\Gamma_\rho/\Gamma)^2 \\ d_{\rho\nu} &= \Gamma_{\rho\nu}/\Gamma - (\Gamma_\nu/\Gamma)(\Gamma_\rho/\Gamma) \\ \Gamma_{\nu\nu} &= \Gamma_{\nu\nu}/\Gamma - (\Gamma_\nu/\Gamma)^2 \end{aligned}$$

Finalmente, por convergencia dominada y/o el teorema de derivación bajo el signo integral, las primeras derivadas de la función gamma incompleta son

$$\Gamma_\rho = \Gamma_\rho(\nu, \rho) = -\rho^{\nu-1} e^{-\rho} \quad \text{and} \quad \Gamma_\nu = \Gamma_\nu(\nu, \rho) = \int_\rho^\infty \log(t) t^{\nu-1} e^{-t} dt$$

y las segundas

$$\begin{aligned} \Gamma_{\rho\rho} = \Gamma_{\rho\rho}(\nu, \rho) &= (\rho^{\nu-1} - (\nu - 1)\rho^{\nu-2}) e^{-\rho} \\ \Gamma_{\rho\nu} = \Gamma_{\rho\nu}(\nu, \rho) &= -\log(\rho) \rho^{\nu-1} e^{-\rho} \\ \Gamma_{\nu\nu} = \Gamma_{\nu\nu}(\nu, \rho) &= \int_\rho^\infty \log^2(t) t^{\nu-1} e^{-t} dt \end{aligned}$$

4.3. Estudios numéricos

Ciertos fenómenos que pueden ajustarse por la distribución power-law o la distribución Pareto presentan una desviación de estas leyes para valores grandes. A menudo es debido a la interferencia que produce un límite global (la cuenca finita del océano o una pérdida limitada al total valor de la economía). La principal motivación de hallar la FTG era encontrar un modelo para explicar este hecho en varios casos, tales como la energía de ciclones tropicales o al cálculo de capital regulador por pérdidas derivadas del riesgo operacional. Nuestro principal logro es que la distribución de Pareto se halla dentro de este modelo, por lo tanto, la inferencia por verosimilitud está disponible. Tenga en cuenta que incluso para la comparación de la distribución gamma truncada con la distribución de Pareto es necesaria la extensión que viene dada por la FTG.

4.3.1. Análisis de los ciclones tropicales

Corral, *et al.* (2010) analiza la influencia de la variabilidad climática y calentamiento global a través del análisis de la ocurrencia de los ciclones tropicales. Su enfoque se basa en la aplicación de una estimación de la energía liberada por los ciclones tropicales individuales. Vamos a comparar nuestro modelo con el análisis estadístico de la distribución power-law de 494 ciclones tropicales en el Norte Atlántico entre 1966 y 2009.

Para medir la importancia de los ciclones tropicales se utiliza una estimación de la energía liberada, el índice de disipación de energía (PDI), definido por

$$PDI = \sum_t v_t^3 \Delta t$$

dónde t denota el tiempo y recorre toda la vida de la tormenta y v_t es el máximo velocidad del viento superficial sostenida en el instante t (la unidades de la PDI son m^3/s^2). La PDI de los datos originales se hallan entre $5,38 \cdot 10^8$

y $2,5410^{11}$. Las desviaciones de la power-law para valores de PDI pequeños se atribuyeron a que los datos son incompletos para casos de menor importancia. Las estimaciones sólo consideran ciclones tropicales con PDI mayor que 310^9 , que corresponde a un tamaño de muestra de 372 (75 % de los datos originales).

La Figura 4.2 muestra el ajuste de la distribución power-law con una aproximación empírica de la función de densidad de la muestra. Corral, *et al.* (2010) aproximan la función de densidad en los puntos $p_r = 10^{8+(r-1)/5}$ para $1 < r < m$ dónde $m = 21$, para los valores del histograma

$$h_r = \frac{\#\{x_i : l_r < x_i \leq l_{r+1}\}}{n(l_{r+1} - l_r)}$$

para los intervalos dados por $l_s = 0,5 \cdot 10^{8+s/5} \cdot 11^{1/5}$ con $1 < s < m + 1$. El punto clave de su método es representar en escala logarítmica (base 10) los dos ejes, dado que la densidad de la power-law en esta situación corresponde a una línea recta. El ajuste se realiza por el método de mínimos cuadrados para el conjunto de puntos $\{(u_r, v_r)\}$, dónde $u_r = \log_{10} p_r$ y $v_r = \log_{10} h_r$.

Nuestra primera contribución consiste en ajustar la distribución FTG por MLE para la muestra entera. La distribución FTG muestra a realmente mejor ajuste, especialmente en la cola de los datos, ver Figura 4.2. El rápido decaimiento para valores grandes del PDI está asociado con el tamaño finito de la cuenca del océano. Es decir, las tormentas con mayor PDI no tienen espacio suficiente para durar más tiempo. Lo relevante es que la distribución de FTG ajusta los datos, incluso en esta situación. El Teorema 4.1.5 muestra que la distribución de Pareto se halla dentro del modelo FTG, por lo tanto tenemos disponible inferencia basada en la verosimilitud para la elección del modelo.

El MLE de los parámetros y las desviaciones estándar se muestran en la Tabla 4.1 para la distribución FTG y Pareto (con dos parámetros). Los valores de la función de log-verosimilitud son $-667,58$ Para el caso de FTG (distribución gamma truncada) y $-680,06$ para el caso de Pareto.

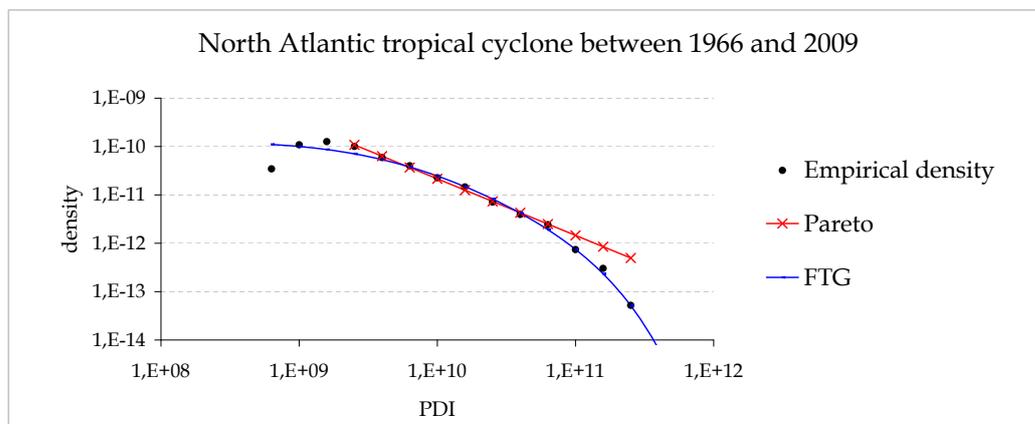


Figura 4.2: La distribución FTG ajusta mejor que la distribución Pareto la PDI de los ciclones tropicales, especialmente en la cola de las observaciones. El gráfico usa escala logarítmica para los dos ejes.

Primero de todo, la prueba de bondad de ajuste para la distribución Pareto dada por Choulakian&Stephens (2001) rechaza el ajuste Pareto con p-valor menor de 0.001 para ambos estadísticos, $W^2 = 0,28$ y $A^2 = 2,4$, pero este método no ofrece alternativas al modelo. En cambio, la prueba de razón de verosimilitud (LRT) puede ser usado para hallar una región de confianza alrededor de los parámetros de la FTG, concluyendo que la diferencia entre la distribución FTG y la Pareto es altamente significativa. El p-valor es $5,810^{-7}$.

4.3.2. Análisis de la distribución de pérdidas agregadas

Las instituciones financieras utilizan datos de pérdidas internas y externas con el fin comparar los diferentes enfoques para el modelado de la *distribución de pérdidas agregadas*, asociadas a la modelización cuantitativa del riesgo operacional, ver Dutta&Perry (2006), Degen *et al.* (2006) y Moscadelli (2004). Los datos utilizados para el análisis fueron recogidos por varios bancos participando en la encuesta para proporcionar las pérdidas por encima de un umbral, a partir del 2002. Los datos se agrupan en ocho estandarizados eventos *líneas de negocio* y siete *tipos de evento*.

Cuadro 4.1: MLE para las distribuciones FTG y Pareto para ciclones tropicales ocurridos en el Norte Atlántico entre 1966 y 2009. Los datos usados corresponden a los PDI por encima de $3 \cdot 10^9$, con el origen trasladado a cero (las unidades son $10^{10} m^3/s^2$). Este cambio no afecta al LRT ni al parámetro ν .

	Distribución Pareto			Distribución FTG				LRT
	$\alpha = -\nu$	σ	l	ν	σ	ρ	l	
MLE	-1.63	2.01	-680.06	0.28	0.09	0.02	-667.58	24.96
s.e.	0.22	0.41		0.15	0.11	0.02		

Capital riesgo se mide como el percentil 99,9% de la estimación del capital simulado para distribución en pérdidas agregadas en un periodo de un año. Recordamos que un evento de pérdida L_i , o severidad de la pérdida, es un incidente por el cual una entidad sufre daños que pueden medirse con un valor monetario. Una pérdida agregada sobre un periodo de tiempo se expresa como la suma

$$S = \sum_{i=1}^N L_i \quad (4.23)$$

dónde N es una v.a. que representa la frecuencia de pérdidas que ocurren en un periodo. Como es usual, aquí se asume que las L_i son i.i.d., y que cada L_i es independiente de N , que tiene distribución de Poisson, con parámetro λ .

El conjunto de datos usado aquí corresponde a las 40 mayores pérdidas asociadas con la línea de negocios *finanzas corporativas* y el tipo de evento *fraude externo*, observado por encima de un umbral, $u = \$20,000$. Para simplificar, los data $\{x_j\}$ has sido escalado al umbral cero y media 100, según

$$y_j = 100 \left(\frac{x_j - u}{\bar{x} - u} \right) \quad (4.24)$$

dónde $u = 20,000$ y $\bar{x} = 942211,3$.

Los 40 excesos, redondeados a dos decimales son: 0,07, 0,11, 0,26, 0,40, 0,46, 0,62, 0,70, 0,75, 0,89, 1,08, 1,52, 1,64, 1,69, 2,04, 2,19, 2,52, 2,73, 3,16, 3,74, 4,04, 4,63, 5,44, 5,86, 6,02, 10,32, 19,63, 29,13, 30,36, 30,88, 35,78, 40,07, 46,12, 137,52, 237,05, 311,14, 314,19, 396,29, 552,48, 864,88, 891,62.

Pérdidas agregadas son determinadas principalmente por los valores extremos de la distribución de eventos de pérdidas. En este caso, el capital riesgo depende de 40 excesos, pero, para calcular el cuantil 99,9%, se requiere un modelo. Bajo el enfoque POT se modelan con distribuciones de Pareto, ver Degen, *et al.* (2007) y Moscadelli (2004). Sin embargo, este enfoque puede dar lugar a estimaciones del capital poco realistas, especialmente cuando el ajuste en el modelo Pareto corresponde a una distribución con esperanza infinita.

Dado que el conjunto de datos tiene solamente excesos por encima de un umbral, el método PoT es el más apropiado. Cuando todas pérdidas son recogidas, Dutta&Perry (2006) utiliza una cuatro-paramétrica distribución, llamada g-and-h, para modelar los datos. Si nos centramos en los eventos extremos de returns de activos financieros, tanto al alza como a la baja, metodologías estándares también incluyen la clásica t de Student y distribuciones Pareto estables, ver Rachev, *et al.* (2010). Pero esto no debería ser considerado un modelo para colas, como veremos en el Capítulo 5, y por lo tanto, adecuada para combinar con la metodología PoT.

La Tabla 4.2 da el MLE de los parámetros para las distribuciones exponencial, gamma, Pareto y FTG, así como la correspondiente desviación estándar y los valores de la función de log-verosimilitud. Primero de todo observamos que para la distribución de Pareto los parámetros están en el rango $0 < \nu = -\alpha < 1$, esto es, una distribución con esperanza infinita. Esto no puede ser rechazado con la prueba de bondad de ajuste dado por Choulakian&Stephens (2001), puesto que el parámetro está fuera del rango de parámetros proporcionados en sus tablas. No obstante, la distribución de Pareto se halla anidada en el modelo FTG ver Teorema 4.1.5, y la prueba de

Cuadro 4.2: MLE y el valor de la función de log-verosimilitud para las distribuciones exponencial, gamma, Pareto y FTG ajustando las pérdidas por fraude externo.

Distribución Exponencial		λ	l		
	MLE	0.010	-224.207		
	s.e.	0.002			
Distribución Gamma		ν	θ	l	
	MLE	0.271	0.003	-181.938	
	s.e.	0.048	0.001		
Distribución Pareto		$\alpha = -\nu$	σ	l	
	MLE	-0.447	1.382	-174.440	
	s.e.	0.102	0.732		
Distribución FTG		ν	σ	ρ	l
	MLE	-0.197	0.651	4.3e-4	-172.369
	s.e.	0.152	0.586	6.2e-4	

razón de verosimilitud da 4,142, con p-valor 0,042. Por lo tanto, la distribución FTG es un modelo más verosímil para el conjunto de datos, dado que la distribución Pareto está fuera de la región de confianza del 95 % para los parámetros de la FTG.

La Figura 4.3 muestra la función de supervivencia empírica y el ajuste dado por la distribución de Pareto y la FTG. la probabilidad de exceder el máximo de la muestra se estima en 5,52 % usando el modelo Pareto y 2,65 % con el modelo FTG, esta diferencia no parece crucial. No obstante, la estimación de altos cuantiles dependen fuertemente del modelo. El cuantil 0,999 es $6,95 \cdot 10^6$ con el modelo Pareto y $3,93 \cdot 10^3$ con el de la FTG. Más aún, la diferencia es aún mayor si calculamos el valor esperado para la pérdida

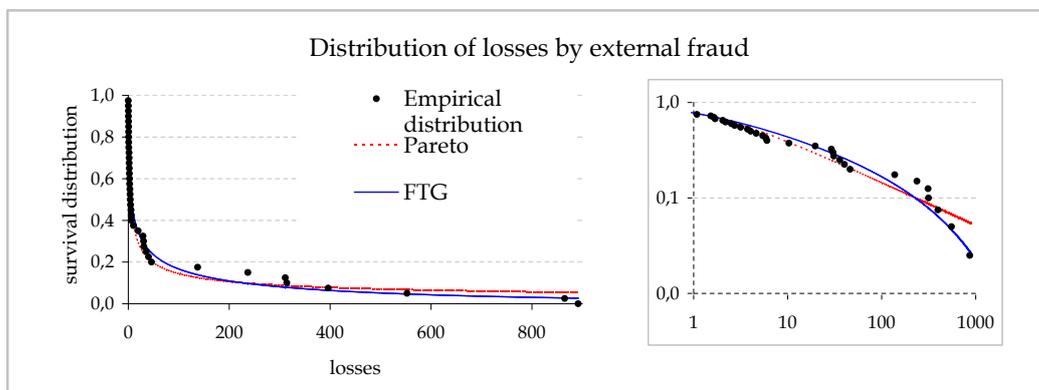


Figura 4.3: La FTG y la distribución de Pareto ajusta la función de supervivencia empírica de un modo similar en el rango de los datos observados. No obstante, la estimación de altos cuantiles difieren enormemente.

que supere este cuantil, que es el valor esperado de una pérdida si se produce un evento de la cola; esto es 12970,6 para el modelo FTG, dado que (4.12), y infinito para la distribución de Pareto. Notamos que estas cantidades están medidas en una unidad monetaria (como dólares) para calcular el capital de riesgo, por lo tanto un factor de 10^3 es realmente importante.

El capital de riesgo ha sido calculado como el cuantil 0,999 de las pérdidas totales, calculadas a partir de (4.23), mediante la simulación de 10^5 para N siniestros, donde N es Poisson distribuido con el parámetro $\lambda = 20$ y los eventos de pérdida, L_i , son simulaciones de la distribución ajustada de Pareto y FTG. Usando la distribución FTG, el capital de riesgo es 10,820, 4, utilizando la distribución de Pareto es $5,78 \cdot 10^9$, ver la última fila de la Tabla 4.3. En escala monetaria, ver (4.24), estos valores corresponden a 99.806.952 dólares para la FTG y $5,33 \cdot 10^{13}$ para la distribución de Pareto. Los producto nacional bruto de los EE.UU. en el primer trimestre de 2013 fue de $16,2 \cdot 10^{12}$, por lo tanto, el cálculo Pareto del capital de riesgo es poco realista. Por otro lado la estimación de la FTG es 12 veces el máximo de la muestra, 8242664 dólares en el escala monetaria, que es un resultado razonable.

Cuadro 4.3: Estimaciones de los parámetros y el capital de riesgo con la distribución de Pareto y la distribución FTG de 10 muestras *bootstrap* y del conjunto de datos original.

muestra	Distribución Pareto			Distribución FTG			
	$\alpha = -\nu$	σ	CapRisk	ν	$\log \theta$	$\log \rho$	CapRisk
1	-0,310	0,367	2,47E+13	-0,038	-7,093	-9,250	10832,98
2	-0,373	1,122	3,50E+11	0,003	-6,771	-8,251	9292,72
3	-0,410	1,719	5,36E+10	-0,106	-7,341	-7,792	13407,05
4	-0,423	1,351	1,87E+10	-0,057	-6,543	-7,460	6934,26
5	-0,441	2,195	1,23E+10	-0,006	-6,520	-7,217	7603,78
6	-0,460	1,205	2,63E+09	-0,298	-8,039	-8,287	16860,11
7	-0,486	1,097	6,78E+08	-0,276	-7,313	-7,828	8921,12
8	-0,538	1,769	1,78E+08	-0,360	-7,613	-7,444	10997,30
9	-0,612	3,923	3,86E+07	-0,257	-6,723	-6,141	6503,94
10	-0,763	3,916	1,66E+06	-0,371	-6,113	-5,461	3276,98
original	-0,448	1,382	5,78E+09	-0,197	-7,325	-7,754	10820,37

Para ver la dependencia de la muestra de la estimación del capital de riesgo, hemos generado varias muestras *bootstrap* del mismo tamaño que el conjunto de datos original. Se observa inmediatamente, con un pequeño número de muestras, la inestabilidad del capital de riesgo estimado obtenido con la distribución de Pareto. Sin embargo, las estimaciones obtenidas con la distribución FTG son mucho más estables.

La Tabla 4.3 muestra los parámetros estimados y capital de riesgo con la distribución de Pareto y la distribución FTG de 10 muestras *bootstrap* y para el conjunto de datos original. En todos los casos el riesgo capital de se ha calculado de la misma manera. Las muestras fueron seleccionadas entre 100 muestras *bootstrap* y ordenadas por el parámetro ν , eligiendo uno de cada 10, para más diversidad. Tenga en cuenta que sólo la segunda muestra se

corresponde con la distribución gamma truncada (4.4) y su comportamiento no es diferente del resto. El hecho más prominente es que, además del capital de riesgo poco realista estimado con la distribución de Pareto, su estimación es altamente inestable, con un factor de 10^7 . Por lo tanto, la distribución FTG puede ser una valiosa alternativa a la distribución de Pareto sobre el riesgo operacional. Observamos que el método de *bootstrap* sólo se utiliza con el fin de construir otros escenarios que nos permitan evaluar la estabilidad del valor estimado para el capital de riesgo y el cálculo de este valor no está sujeto a las debilidades del método de *bootstrap* para colas pesadas.

Por último debemos recordar que al igual que los niveles extremos de la energía para el Los ciclones tropicales se ven afectados por los límites de la Tierra, la economía también es finita. Por lo tanto, FTG distribución puede ser una interesante alternativa a la distribución de Pareto en estos ámbitos.

Capítulo 5

Propuestas finales y conclusiones

A continuación vamos a proponer unas condiciones para modelos para colas que son coherentes con lo que hemos observado en nuestra investigación. Además haremos un repaso del estado de la modelización estadística de los valores extremos a fin de este trabajo, dando lugar a posibles líneas de investigación futuras. Finalmente veremos las conclusiones generales de este trabajo.

5.1. Propuestas finales en modelización

Antes de pasar a las conclusiones resumimos nuestro punto de vista sobre la metodología más apropiada en la modelización estadística de valores extremos. Los problemas básicos de la teoría que consideramos son el cálculo del índice de la cola, la selección del umbral y la elección del modelo.

5.1.1. Cálculo del índice de la cola

El cv -estimador del parámetro de forma, ξ , que se deduce de (2.9), es decir,

$$\xi = \frac{cv_n^2 - 1}{2cv_n^2} \quad (5.1)$$

dónde el cv_n es el coeficiente de variación de una muestra de tamaño n , resulta un candidato idóneo para estimar el parámetro de forma en el caso de colas ligeras por su simplicidad. Podemos valorar el sesgo y la eficiencia de dicho estimador a partir de la Figura 5.1. Observamos que el resultado asintótico (3.21) ya se aproxima para muestras de tamaño 100 así como la simetría del estimador, a pesar que este estimador comete errores de subestimación.

En la Tabla 5.1 podemos ver los márgenes de error de este estimador para muestras pequeñas. Tanto para los gráficos de la Figura 5.1 como los datos obtenidos de la Tabla 5.1 son el resultado de la simulación de 100000 muestras de GPD con el parámetro de forma correspondiente y $\psi = 1$, puesto que el sesgo y la eficiencia de este estimador no dependen del parámetro de escala.

Este estimador para el caso de cola pesada, $\xi > 0$, no tiene buenas propiedades. No obstante, este estimador debería ser analizado y extendido debidamente en un futuro, puesto que para el caso $\xi = 0$ este estimador ha dado muy buenos resultados, no sólo para estimar el parámetro de forma, también permite hallar el umbral óptimo y permite realizar pruebas de exponencialidad, ver Castillo *et al.* (2013).

5.1. Propuestas finales en modelización

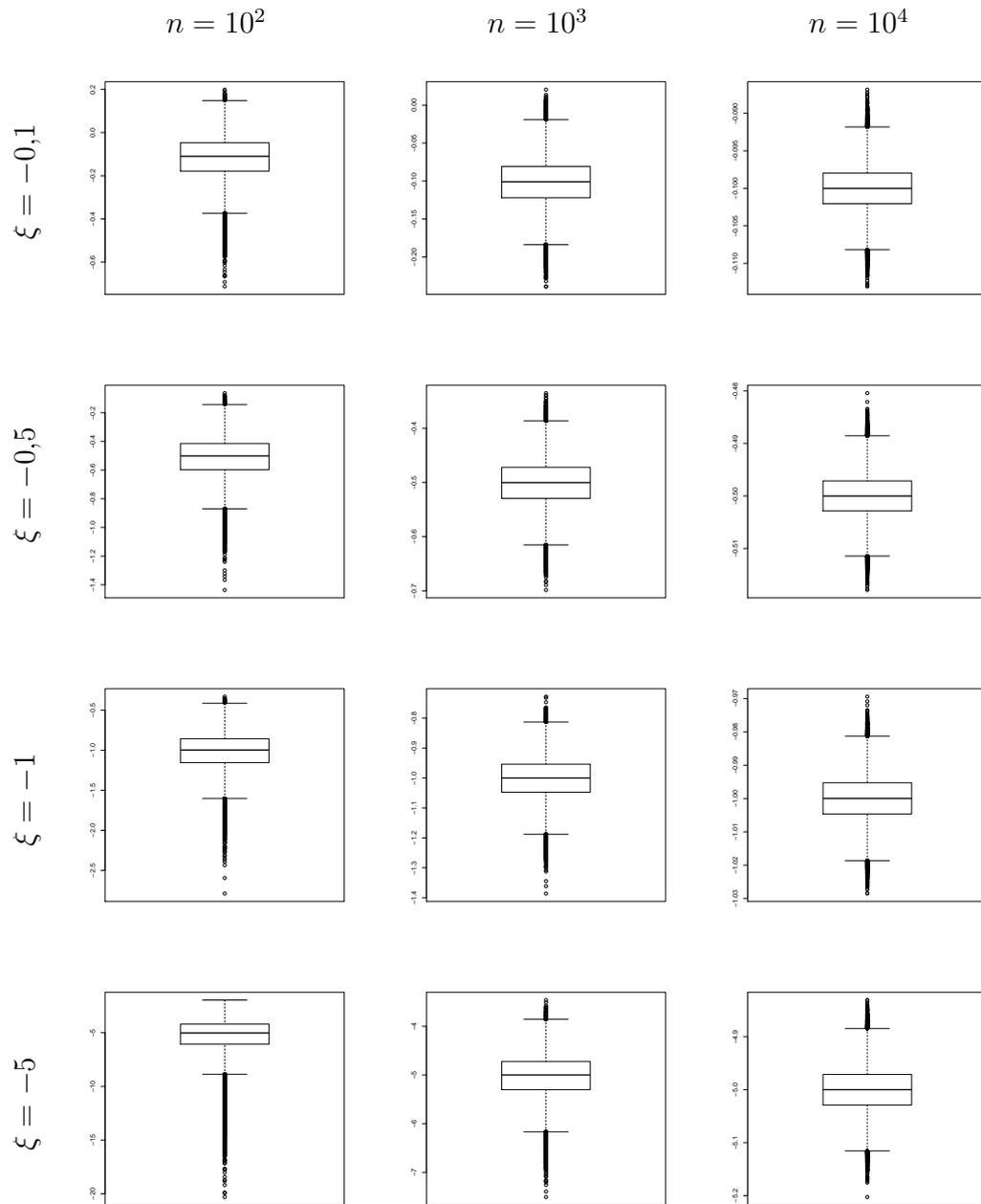


Figura 5.1: Para los índices de la GPD $\xi = -0,1, -0,5, -1, -5$ y tamaños de muestra $n = 10^2, 10^3, 10^5$ se han generado 100000 muestras y estimado el parámetro ξ con el cv-estimador. Los *boxplot* representan los valores de dicha estimación obtenidos.

Cuadro 5.1: Para los índices de cola de la GPD $\xi = -0,1, -0,5, -1, -5$ y tamaños de muestra pequeños $n = 10, 20, 50, 100$ se han generado 100000 muestras y estimado el parámetro ξ con el cv-estimador. Observamos el intervalo de confianza exacto del 95 % obtenido.

n	ξ	$media$	$desv.$	i.c. 955 %	
10	-0.1	-0.29	0.44	-1.06	0.18
	-0.5	-0.67	0.65	-1.81	-0.01
	-1.0	-1.27	1.22	-3.27	-0.21
	-5.0	-44.62	1176.06	-73.80	-1.39
20	-0.1	-0.19	0.25	-0.65	0.13
	-0.5	-0.57	0.35	-1.22	-0.12
	-1.0	-1.10	0.60	-2.21	-0.39
	-5.0	-8.05	19.35	-21.21	-2.03
50	-0.1	-0.13	0.14	-0.39	0.07
	-0.5	-0.52	0.20	-0.89	-0.24
	-1.0	-1.04	0.33	-1.64	-0.58
	-5.0	-5.68	2.73	-10.56	-2.80
100	-0.1	-0.12	0.10	-0.29	0.03
	-0.5	-0.51	0.14	-0.75	-0.30
	-1.0	-1.02	0.23	-1.42	-0.68
	-5.0	-5.29	1.59	-8.24	-3.29

5.1.2. Selección del umbral óptimo

Si consideramos el caso general, GPD, y después de haber hallado una metodología satisfactoria desde un punto de vista práctico para el cálculo del MLE, debemos considerar como referencia el método no paramétrico para hallar el umbral óptimo que proviene de este método de estimación, se puede

denominar ML-plot, y a poder ser con la corrección propuesta por Coles, para el parámetro de escala, ver Figura 5.2.

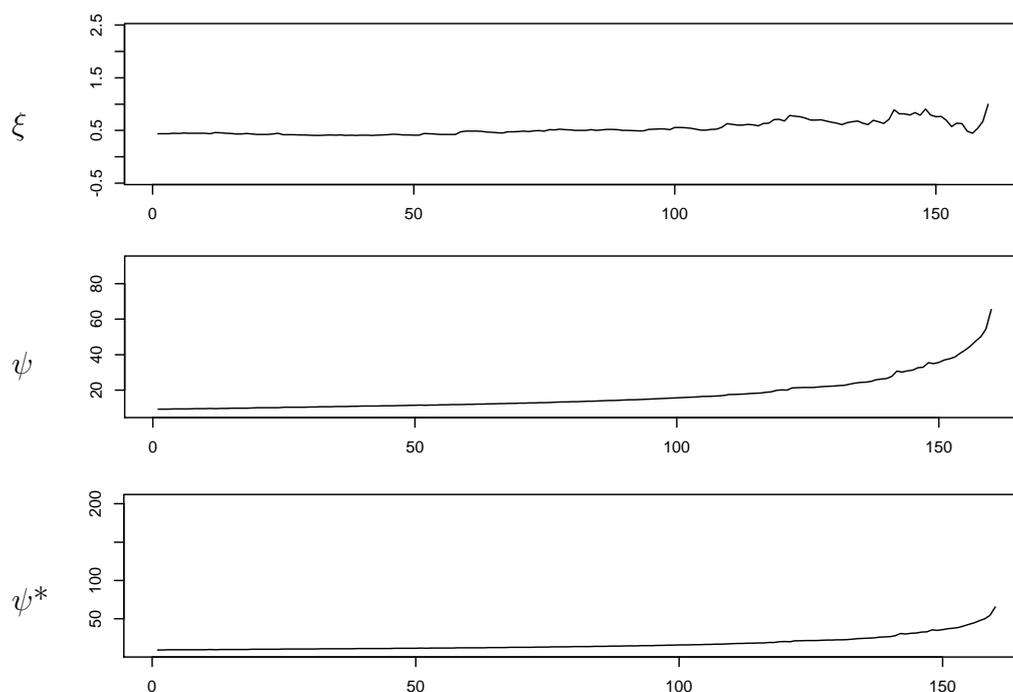


Figura 5.2: Gráfico ML-plot para los 150 mayores excesos de las pérdidas de incendios en Dinamarca.

La Figura 5.2 muestra el ML-plot para los 150 mayores excesos de las pérdidas de incendios en Dinamarca considerados en el Capítulo 2. El primero gráfico corresponde a la estimación por MLE del parámetro ξ , el segundo del ψ y el tercero la corrección propuesta por Coles para buscar zonas constantes del parámetro ψ . En el Capítulo 2 vimos que estos datos tenían un buen ajuste con la distribución LogNormal, la cual tiene cola exponencial. El ML-plot muestra indicios de considerar el modelo de cola pesada para parámetro de forma alrededor de 0,5 y no parece muy clara la aceptación de $\xi = 0$. No obstante, el gráfico ML para el valor de ψ corregido no muestra

ninguna señal de constancia, lo que nos lleva a concluir acorde a lo analizado en el Capítulo 2. Además, observemos la discordancia entre el ML-plot y Hill-plot, Figura 2.2 que proponía valores muy diferentes.

En cualquier caso, para el caso de colas no ligeras, este método es más satisfactorio que el Hill-plot. Para el caso de colas ligeras, así como el cv -estimador tiene potencial para ser un buen estimador para el índice de la cola, el CV-plot debería ser considerado para hallar el umbral óptimo, puesto que muestra mejores resultados a nivel empírico que el ME-plot. Insistimos que este método ya se utiliza para hallar el umbral óptimo de colas exponenciales.

5.1.3. Modelos para colas

Podemos concluir que un modelo para colas debe ser cerrado por escala, por truncamiento y debe contener distribuciones del modelo GPD para cierto rango del índice de la cola. Nosotros hemos considerado modelos para colas no ligeras y por tanto, hemos considerado un modelo que contenía la distribución Pareto y la distribución exponencial, la FTG.

Nuestro objetivo es establecer un concepto genérico para modelo de colas (no ligeras) que nos dé ciertas garantías de éxito en los análisis de valores extremos.

Del mismo modo que la distribución de Pareto y la power-law no son la misma, dado un modelo estadístico, $\mathcal{M} = \{f(x; \theta)\}$, como una familia de funciones de densidad sobre un espacio muestral $S \subset \mathbb{R}$, hay también una clara diferencia entre la distribución truncada por la izquierda de \mathcal{M} en un umbral c y las distribuciones de las colas de \mathcal{M} . Por ahora, sea $[c, \infty) \subset S$, las distribuciones de \mathcal{M} truncadas en el umbral c corresponde a la familia de funciones de densidad definidas en $[c, \infty)$, donde c está fijado. Así, para un umbral $u > c$, la densidad de probabilidad de los excesos a u definidos por (1.2.7) no necesariamente pertenecen a la anterior familia, aunque la consideremos trasladada a c . Si denotamos por \mathcal{M}_c la familia de distribu-

ciones truncadas por la izquierda en el umbral c y trasladadas a 0, entonces el modelo para colas asociado a \mathcal{M} y c es $\cup_{u \geq c} \mathcal{M}_u$. En algunos casos los dos modelos concuerdan, por ejemplo si el modelo corresponde a la familia de distribuciones Pareto o a las distribuciones exponenciales, los cuales son modelos de escala, y más aún, el truncamiento es un cambio de escala.

Definición 5.1.1. Sea \mathcal{M} un modelo estadístico sobre un espacio muestral, $S \subset \mathbb{R}$, y $c \in S$. El modelo de colas asociado a \mathcal{M} y c es el conjunto de funciones de densidad de los excesos de u definidos por (1.2.7) para todo $u > c$ y para cada $f \in \mathcal{M}$.

Como conclusión final al trabajo realizado analizando diferentes modelos para modelización de colas, hemos podido llegar a establecer algunos criterios deberían cumplir los modelos para colas. Una primera exigencia es la de ser cerrado por cambios de escala, puesto que el cambio de escala de nuestros datos no puede afectar a la calidad del ajuste del modelo, por ejemplo, cambio de divisa, etc. Una segunda exigencia es que debe ser un modelo que contenga todas sus colas, en el sentido de que el modelo debe ser válido para modelar cualquier cola de los datos originales. Y por último, debe contener la familia GPD, puesto que desde un punto de vista teórico, es el modelo que le corresponde a una cola para cierto umbral. Vamos a formalizar esta definición para colas no ligeras, es decir, pesadas o exponenciales. Insistimos que para proceder a una modelización de colas, así como hallar el índice de una cola, se debe hacer, en la medida de lo posible, posteriormente a fijar el tipo de colas.

Definición 5.1.2. Diremos que \mathcal{M} es un modelo para colas no ligeras si las siguientes condiciones se satisfacen:

1. \mathcal{M} es igual al modelo de colas asociado a algún modelo \mathcal{M}' y algún $c \in \mathbb{R}$.
2. \mathcal{M} es un modelo de escala.

3. Las distribuciones exponenciales y de Pareto están contenidas en \mathcal{M} .

Si consideramos el modelo que contenga solamente la familia de distribuciones Pareto y exponencial, obtenemos el ejemplo más sencillo de modelo para colas no ligeras, lo cual concuerda perfectamente con el Segundo Teorema Fundamental en EVT. Veamos que la distribución FTG también puede ser usada para modelar colas en el sentido de la Definición 5.1.2.

Corolario 5.1.3. *El modelo FTG es un modelo para colas.*

Demostración. Sea \mathcal{M} el modelo FTG. Por la Proposición 4.1.2, el modelo FTG es modelo de escala y $f_u \in \mathcal{M}$, para todo $f \in \mathcal{M}$ y para todo $u > 0$. Por otro lado, para cualquier parámetros (α, θ, ρ) de la FTG y para cualquier $\epsilon > 0$ tal que $0 < \epsilon < \rho/\theta$, si $u = \rho/\theta - \epsilon$ y g es la distribución FTG con parámetros $(\alpha, \theta, \rho - \epsilon\theta)$, entonces g_u es la distribución FTG con parámetros (α, θ, ρ) , por la Proposición 4.1.2. Entonces, $\mathcal{M} = \cup_{u \geq 0} \mathcal{M}_u$ donde $\mathcal{M}_u = \{f_u | f \in \mathcal{M}\}$, es decir, el modelo para colas asociado al modelo FTG y $0 \in \mathbb{R}$ es exactamente el modelo FTG. Finalmente, por el Teorema 4.1.5, la distribución exponencial y Pareto pertenecen al modelo \mathcal{M} . \square

Si observamos las buenas propiedades de la FTG, vemos que no quedan aquí, la FTG con el parámetro σ fijado es un modelo exponencial y esto ya hemos visto que era de mucha utilidad sobre todo para su calibración. En general, si a un modelo exponencial natural, es decir, con estadístico suficiente $T(x) = x$, le añadimos un parámetro de escala se denomina modelo exponencial de dispersión, ver Barndorff-Nielsen (1978). Si generalizamos este concepto a modelos no naturales, obtenemos que la FTG es un modelo exponencial de dispersión. Veamos como podemos generalizar esta propiedad con la finalidad de refinar un poco más los modelos para colas que deben considerarse a nivel práctico.

En general, sea \mathcal{M} un modelo exponencial, generado por la medida de Lebesgue en $[0, \infty)$, con estadístico suficiente $T(x)$. Dado $\sigma > 0$ un umbral fijado, entonces podemos considerar un nuevo modelo exponencial con

estadístico suficiente $T(x+u)$, o equivalentemente $T(1+x/u)$, que denotaremos por \mathcal{M}'_σ . Si ahora consideramos el modelo $\mathcal{M} = \cup_{\sigma \geq 0} \mathcal{M}'_\sigma$, obtenemos que \mathcal{M} es un modelo exponencial de dispersión y es candidato a modelo para colas, puesto que es el modelo de colas asociado a \mathcal{M}' y 0, lo llamaremos modelo de colas exponencial de dispersión. Si realizamos este proceso con \mathcal{M}' el modelo bi-paramétrico gamma, obtenemos que \mathcal{M} corresponde al modelo FTG. Y este sí es un modelo para colas.

Teorema 5.1.4. *Dado $\sigma > 0$ la FTG es el más pequeño modelo exponencial completo que contiene la distribución de Pareto con σ fijado y la distribución exponencial. Es decir, la FTG es el modelo exponencial de dispersión completo más pequeño que contiene la distribución de Pareto y la exponencial.*

Demostración. Es consecuencia directa de la Proposición 4.1.3 y del Teorema 4.1.5 □

Las ventajas de modelar con modelos exponenciales son, entre otras, su posible extensión a modelos lineales generalizados y la estimación de los parámetros. Veamos esto último en el caso general.

Sea \mathcal{M}' un modelo exponencial, generado por la medida de *Lebesgue* en $[0, \infty)$, con estadístico suficiente $T(x)$ y lo describimos por la familia de funciones de densidad $\{f(x; \theta) | \theta \in \Theta\}$, el modelo de colas exponencial de dispersión asociado a $\mathcal{M}', \mathcal{M}$, es la familia de funciones de densidad

$$f(x; \theta, \sigma) = \exp(\theta \cdot T(x/\sigma) - k(\theta)) / \sigma.$$

con $\sigma > 0$ y $\theta \in \Theta$. Sea $\mathbf{x} = \{x_i\}$ una muestra de tamaño n , la función de log-verosimilitud viene dada por

$$l(\mathbf{x}; \theta, \sigma) = \theta \cdot t(\mathbf{x}/\sigma) - k(\theta) - \log(\sigma). \tag{5.2}$$

Tomamos las derivadas con respecto a θ y σ , y podemos describir las ecuaciones de verosimilitud por

$$t(\mathbf{x}/\sigma) - \nabla k(\theta) = 0 \tag{5.3}$$

$$\theta \cdot \nabla t(\mathbf{x}/\sigma) \cdot \mathbf{x}/\sigma - 1 = 0 \quad (5.4)$$

Definimos $\psi = (\nabla k)^{-1}$ y usamos (5.3) para simplificar (5.4) obteniendo

$$\psi(t(\mathbf{x}/\sigma)) \cdot \nabla t(\mathbf{x}/\sigma) \cdot \mathbf{x}/\sigma = 1 \quad (5.5)$$

Y esta es la ecuación que deberemos resolver y la que puede que no tenga solución o que no sea única deberá ser estudiado en cada caso. A modo más general, este primer paso para la calibración corresponde a maximizar la función profile-verosimilitud que obtenemos de simplificar (5.2) usando la solución de (5.3).

5.2. Conclusiones

A lo largo de este trabajo hemos intentado dar al lector una visión panorámica del estado del arte en modelización estadística de los valores extremos. A continuación vamos a enumerar las conclusiones de este trabajo.

- Desde un punto de vista matemático, o probabilista, la teoría de los valores extremos es satisfactoria, pero necesita de herramientas más pragmáticas como, por ejemplo, nuevos modelos para colas y resultados para muestras de tamaño pequeño.
- Desde un punto de vista estadístico, las herramientas son insuficientes y, por popularidad, se han aceptado metodologías con fundamento teórico que no alcanzan su realización a nivel práctico. En general, los resultados en modelización de valores extremos son escasos y en EVT no son satisfactorios, ver Diebold *et al.* (1998). La teoría muestra algunas sorpresas en aplicaciones prácticas. Por ejemplo, Dutta&Perry (2006) observaron resultados no realistas en un análisis estadístico de modelos para estimar las pérdidas por riesgo operacional.
- A nivel práctico, el enfoque basado en la GPD que caracteriza la distribución de una cola utiliza más información que el basado en la GEV. Además la metodología POT debería ser el método de referencia para enlazar modelos para colas. Los modelos para colas deben ser seleccionados a posteriori de hallar un rango que contenga el índice de la cola. Los principales retos en modelización estadística de valores extremos son: estimar el índice de la cola, hallar el umbral óptimo donde enlazar y presentar nuevos modelos para colas.
- Hay que ampliar la clasificación clásica del tipo de colas en ligeras, exponenciales y pesadas. El trato especial que requieren las colas exponenciales debe ser también considerado para colas uniformes. Además, sobretodo si trabajamos con pocos datos, debemos seleccionar a priori

al análisis a través de indicios del contexto de los datos un rango para el índice de la cola.

- En modelización estadística de valores extremos el MLE es viable, en contra de algunas opiniones. Aclaremos el comportamiento anómalo del MLE, analizando cuidadosamente sus sub-modelos. Esto permite usar herramientas de inferencia como AIC, BIC, LRT.
- Los resultados anteriores se complementan con el uso del CV-plot que debe ser tomado como método de referencia, en lugar del ME-plot para colas con varianza finita, así como el ML-plot debería sustituir al Hill-plot, al menos para colas pesadas.
- El modelo FTG que hemos introducido tiene la particularidad de ser tan próximo a Pareto como se desee, pero con decrecimiento exponencial de las colas. Este modelo cumple todas las exigencias que nuestra experiencia nos lleva a considerar imprescindibles en un modelo para colas no ligeras. Estos requerimientos han sido estructurados desde un punto de vista teórico como fundamento para nuevos modelos para colas.
- Se debe avanzar hacia la axiomatización de modelo para colas. Nuestra propuesta se resume en que debe ser cerrado por escala, por truncamiento y debe contener distribuciones del modelo GPD para cierto rango del índice de la cola. En particular, un modelo para colas no ligeras debe contener la distribución de Pareto y la distribución exponencial. Si manejamos funciones analíticas, por ejemplo familias exponenciales, el conocimiento local implicará el conocimiento global. Por lo tanto, si además los modelos para colas están basados en familias exponenciales haremos más accesibles los modelos a nivel práctico, tanto por simplificar el proceso de calibración como por enlazar con la teoría de modelos lineales generalizados.

5.3. Conclusions

Throughout this paper we have tried to give the reader a overview of the state of the art in statistical modeling of extreme values. Below we list the conclusions of this work.

- *From a mathematical point of view, or probabilistic theory, the theory of extreme values is satisfactory, but needs more pragmatic tools as, for example, new models for tails and results for small sample size.*
- *From a statistical point of view, tools are insufficient and, popularity, have accepted theoretical methodologies that do not reach their conducting practical level. In general, the results for modeling extreme values are scarce and are not satisfactory, see Diebold et al. (1998). The theory shows some surprises in practical applications. For example, Dutta&Perry (2006) reported results unrealistic on analysis of statistical models to estimate risk losses operational.*
- *From a practical level, the GPD based approach that characterizes the tail distribution uses more information than the one based on the GEV. Furthermore POT methodology should be the reference method to link models for tails. the models for tails must be selected a posteriori of to have found a range that contains the index tail. The main challenges in statistical modeling of extreme values are: estimate the index tail, find the optimum threshold and search new models for tails.*
- *We must extend the traditional classification of the type of tails light, exponential and heavy. The special treatment that require the exponential tails should also be considered for uniform tails. Also, especially for small sample size, we have to select a priori, for instance from the context of the data, a range of data for the tail index.*

- *In statistical modelling of extreme values, the use of MLE is feasible, contrary to some opinions. We clarify the behavior MLE anomalous, carefully analyzing their sub-models. This allows using inference tools as AIC, BIC, LRT.*
- *The above results are supplemented by the use of CV-plot that should be taken as the reference method, instead of the ME-plot for tails with finite variance, and the ML-plot should replace the Hill-plot, at least for heavy tails.*
- *The model we have introduced FTG has the distinction being so close to Pareto as desired for some distributions with exponential tails. This model meets all requirements that our experience leads us to consider essential in a model for non-light tails. These requirements have been structured from a theoretical viewpoint as a basis for new models for tails.*
- *We should move towards the axiomatization of the tails model. Our proposal is summarized in that must be closed by scale, truncation and contain GPD model distributions for certain range of the index tail. In particular, a model for non-light tails must contain Pareto distribution and exponential distribution. If the model contains analytic functions, for instance, exponential families, local knowledge involves global knowledge. Therefore, if the models for tails are based on exponential families, they will be more accessible, for simplifying the process of calibration or for linking with the theory of generalized linear models.*

Bibliografía

- Aban, B. I., Meerschaert, M.M. & Panorska, K. A. (2006). Parameter Estimation for the Truncated Pareto Distribution. *J. Amer. Statist. Assoc*, 101, 270-277.
- Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover.
- Akinsete, A., Famoye, F & Lee, C. (2008). The beta-Pareto distribution. *Statistics*, 42, 547-563.
- Arnold, B. C. (1983). *Pareto Distributions*. Fairland, Maryland: International Cooperative Publishing House.
- Artzner, P., Delbaen, F., Eber, J. & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9, 203-228.
- Bank for international settlements (BIS). (2005). *Studies on the Validation of Internal Rating Systems*. Basel Committee on Banking Supervision. Working Paper No. 14.
- Balkema, A. & de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2, 792-804.
- Baricz, A. (2008). Mills ratio: Monotonicity patterns and functional inequalities. *Journal of Mathematical Analysis and Applications*, 340, 1362-1370.
- Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. Wiley Series in Probability and Mathematical Statistics. Chichester: John Wiley & Sons.

-
- Barndorff-Nielsen, O. & Cox, D. (1994) *Inference and asymptotics*. Monographs on Statistics and Applied Probability, 52. Chapman & Hall, London.
- Beirlant, J., Goegebeur, Y., Segers, J. & Teugels Beirlant, J. (2004) *Statistics of Extremes: Theory and applications* Goegebeur, segers, Teugels.
- Beirlant, J., Dierckx, G., Guillou, A. & Staaricaa, C. (2006). On Exponential Representations of Log-Spacings of Extreme Order Statistics. *Extremes*, 5, 157-180.
- Bensalah, Y. (2000) Steps in Applying Extreme Value Theory to Finance: A review. *Working Paper 2000-20 Bank of Canada*
- Brent, R. (1973). *Algorithms for Minimization without Derivatives*. Englewood Cliffs N.J.: Prentice-Hall.
- Brown, L. (1986). *Fundamentals of statistical exponential families with applications in statistical decision theory*. Lecture Notes Monograph Series, 9. Hayward, CA: Institute of Mathematical Statistics.
- Carrillo, S. & Suárez. A. (2006). Medición efectiva del riesgo operacional. *Estabilidad Financiera*, 11, 61-90.
- Castillo, E. & Hadi, A. S. (1997). Fitting the generalized Pareto distribution to data. *J. Amer. Statist. Assoc*, 92, 1609-1620.
- Castillo, J. & Puig, P. (1999). Invariant Exponential Models Applied to Reliability Theory and Survival Analysis. *Journal of the American Statistical Association*, 94, 522-528.
- Castillo, J. del & Daoudi, J. (2009). Estimation of the generalized Pareto distribution. *Statistics & Probability Letters*, 79, 684-688.
- Castillo, J. del, Daoudi, J. & Serra, I. (2013) The full-tails gamma distribution applied to model extreme values. <http://arxiv.org/abs/1211.0130>

-
- Castillo, J. del, Daoudi, J. & Lockhart, R. (2013) Methods to distinguish between polynomial and exponential tails. Aparecerá en *Scandinavian Journal of Statistics*. <http://arxiv.org/abs/1112.0514>
- Chapman, D. G. (1956). Estimating the Parameters of a Truncated Gamma Distribution. *Ann. Math. Statist.*, 27, 498-506.
- Cheng, R. & Amin, N. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *J. Roy. Statist. Soc. Ser. B*, 45, 394-403.
- Cheng, R. & Iles, T. C. (1987). Corrected maximum likelihood in nonregular problems. *J. Roy. Statist. Soc. Ser. B*, 49, 95-101.
- Choulakian, V. & Stephens, M. A. (2001). Goodness-of-Fit for the Generalized Pareto Distribution. *Technometrics*, 43, 478-484.
- Clauset, A., Shalizi, C. R. & Newman, M. E. J. (2009). Power-law Distributions in Empirical Data. *SIAM Review*, 51, 661-703.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics.
- Coles, S. & Sparks (2006). Extreme value methods for modelling historical series of large volcanic magnitudes. *Statistics in Volcanology*, Spec. Publ. of the Int. Assoc. of Volcanol. and Chem. of the Earths Inter. Ch. 5.
- Corral, A., Osso, A. & Llebot, J.E. (2010). Scaling of tropical-cyclone dissipation. *Nature Physics*, 6, 693-696.
- Davis, H. T. & Michael L. F. (1979). The Generalized Pareto Law as a Model for Progressively Censored Survival Data. *Biometrika*, 66, 299-306.
- Davison, A. C. & Smith, R. L. (1990). Models for exceedances over high thresholds. With discussion and a reply by the authors. *J. Roy. Statist. Soc. Ser. B*, 52, 393-442.
- Degen, M., Embrechts, P. & Lambrigger, D. (2007). The Quantitative Modeling of Operational Risk: Between g-and-h and EVT. *Astin Bulletin*, 37, 265-291.

-
- Den Broeder, G. G. (1955) On parameter estimation for truncated Pearson type III distributions. *Ann. Math. Statist.*, 26, 659-663.
- Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.
- Dierckx, G., Beirlant, J., De Waal, D. & Guillou, A. (2008). A new estimation method for Weibull-type tails based on the mean excess function. *Journal of Statistical Planning and Inference*, 139, 1905-1920.
- Diebold, F., Schuermann, T. & Stroughair, J. (1998) Pitfalls and Opportunities in the Use of Extreme value Theory in Risk management. *Working papers series*, FIN-98-081.
- Dixit, U. J. & Phal, K. D. (2005). Estimating scale parameter of a truncated gamma distribution. *Soochow Journal of Mathematics*, 31, 515-523.
- Drees H., de Haan, L. & Resnick, S. (2000). How to make a hill plot. *Annals of Statistics*, 28, 254-274.
- Dutta, K. & Perry, J. (2006). *A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital*. Federal Reserve Bank of Boston. Working Paper 06-13.
- Embrechts, P. Klüppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin.
- Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London. Ser. A*, 222, 309-368.
- Furlan, C. (2010). Extreme value methods for modelling historical series of large volcanic magnitudes. *Statistical Modelling*, 10, 113-132.
- Grimshaw, S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics*, 35, 185-191.
- Guillou, A. & Hall, P. (2001). A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society, Series B*, 63, 293-305.

-
- Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32, 177-203.
- Hall, P. & Wang, J. Z. (2005). Bayesian likelihood methods for estimating the end point of a distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67, 717-729.
- Harter, H. L. (1967). Maximum-likelihood estimation of the parameters of a four-parameter generalized gamma population from complete and censored samples. *Technometrics*, 9, 159-165.
- Hegde, L. M. & Dahiya, R. C. (1989). Estimation of the parameters of a truncated gamma distribution. *Communications in Statistics: Theory and Methods*, 18, 561-577.
- Hosking, J.R.M. & Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29, 339-349.
- Jawitz, J. (2004). Moments of truncated continuous univariate distributions. *Advances in Water Resources*, 27, 269-281.
- Juarez, S. F. & Schucany, W. R. (2004). Robust and efficient estimation for the generalized Pareto distribution. *Extremes*, 7, 237-251.
- Kozubowski, T. J.; Panorska, A. K.; Qeadan, F.; Gershunov, A. & Rominger, D. (2009) Testing exponentiality versus Pareto distribution via likelihood ratio. *Comm. Statist. Simulation Comput.*, 38, 118-139.
- Letac, G. (1992). *Lectures on natural exponential families and their variance functions*. Monografias de Matemática, 50, IMPA, Rio de Janeiro.
- Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Comput. Statist. Data Anal.*, 51, 904-917.
- McNeil, A. J., Frey, R. & Embrechts P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

-
- Moscadelli, M. (2004). *The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee*. Economic working papers, 517, Bank of Italy, Economic Research Department.
- Papastathopoulos, I. & Tawn, J.A. (2012). Extended generalised Pareto models for tail estimation. *J. Statist. Plann. Inference*, 143, 131-143.
- Philippe, A. (1997). Simulation of right and left truncated gamma distributions by mixtures. *Statistics and Computing*, 7, 173-181.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 119-131.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rachev, S. T., Racheva-Iotova, B. & Stoyanov, S. (2010). Capturing fat tails in Risk. *Risk Management, Derivatives and Regulation*, 72-77.
- Resnick, S.I. (1997). Discussion of the Danish data on large fire insurance losses. *Astin Bulletin*, 27, 139-151.
- Finkenstädt, B. & Rootzén, H. (2003). *Extreme values in finance, telecommunication and the environment*, edited by B. , Chapman&Hall.
- Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, 72, 67-90.
- Song, J & Song, S. (2012). A quantile estimation for massive data with generalized Pareto distribution. *Comput. Statist. Data Anal.*, 56, 143-150.
- Sornette, D. (2006). *Critical phenomena in natural sciences*. Springer Berlin Heidelberg New York.
- Stacy, E. W. (1962). A generalization of the gamma distribution. *Ann. Math. Stat.*, 33, 1187-1192.
- Weiss, L. & Wolfowitz, J. (1973). Maximum likelihood estimation of a translation of a truncated distribution. *Ann. Statist*, 1, 944-947.

-
- Wong, T. & Li, W. (2006). A note on the estimation of extreme value distributions using maximum product of spacings. *Time series and related topics: in memory of Ching-Zong Wei. Beachwood*, 272 - 283.
- Zhang, J. (2010). Improving on estimation for the generalized Pareto distribution. *Technometrics*, 52, 335-339.
- Zhang, J. & Stephens, M. A. (2009). A new and efficient estimation method for the generalized Pareto distribution. *Technometrics*, 51, 316-325.