

# Aplicación de la teoría del valor extremo al ajuste y modelación de catástrofes

MARÍA JOSÉ PÉREZ FRUCTUOSO  
UNIVERSIDAD CARLOS III DE MADRID

ALMUDENA GARCÍA PÉREZ  
UNIVERSIDAD DE ALCALÁ DE HENARES

**E**l presente trabajo tiene por objeto el análisis de datos de pérdidas asociadas a catástrofes y su posible modelación. Para ello, se toma como base la Teoría del Valor Extremo, y se ajustan las distribuciones características del valor extremo a los datos disponibles, con la intención de prever posibles catástrofes futuras o la probabilidad de que se produzca un evento de consecuencias determinadas. En concreto, se determina el modelo paramétrico que mejor se ajusta a los datos y la estimación de los parámetros del mismo mediante el método de Máxima-Verosimilitud.

## INTRODUCCIÓN

La Teoría del Valor Extremo se aplica para la modelación de las colas de las distribuciones con el objetivo de evitar los problemas derivados del ajuste global de un conjunto de datos a través de un modelo único: aplicando la ley de los grandes números, o el teorema central del límite, el ajuste de los valores centrales deja fuera del análisis aquellas situaciones que escapan a la normalidad. La aplicación de un modelo individual que represente el comportamiento de todos los datos puede no ser aceptable para los valores elevados.

En determinadas ocasiones las entidades aseguradoras se encuentran ante situaciones que po-

nen en peligro la estabilidad y la solvencia de las mismas. La Teoría del Valor Extremo puede ayudar, no a eliminar los sucesos que causan consecuencias extremas gravosas, como las catástrofes, sino a predecirlos y, por tanto, a ajustar mejor las primas, los recargos o a dotar mayores provisiones.

La aplicación de la Teoría del Valor Extremo también resulta de interés en campos como el reasegurador, donde, fundamentalmente en los tratados no proporcionales, la cobertura se realiza exclusivamente para niveles de siniestralidad altos. Ello implica, por tanto, que en general la correcta tarificación del reaseguro no debe realizarse sobre la base de los valores medios sino a partir de los valores extremos.

## LA TEORÍA DEL VALOR EXTREMO

Es evidente, que la utilidad de la Teoría del Valor Extremo es indiscutible, sin embargo, la ejecución de la misma afronta varios retos, entre ellos la dispersión de los datos, la determinación del grueso de la cola, la elección de métodos para la estimación de los parámetros del modelo y la elección de un umbral, o prioridad, determinado.

El objetivo de este artículo es superar dichos retos en el análisis del comportamiento esperado de una muestra de pérdidas por catástrofes a partir del comportamiento de la misma por encima de una determinada cuantía.

Para ello, en primer lugar, se describen de forma teórica las dos distribuciones importantes que conforman la Teoría del Valor Extremo y que vamos a ajustar a los datos mencionados: la distribución generalizada del valor extremo y la distribución generalizada de Pareto.

En segundo lugar, y como parte central del artículo, se introducen los correspondientes modelos o distribuciones que ajustan a las observaciones. Para ello se trabaja con los gráficos obtenidos mediante la aplicación de los programas XTREMES<sup>1</sup>, STATGRAPHICS y EXCEL ya que las herramientas gráficas y visuales son las más utilizadas para discutir sobre los modelos.

De los gráficos analizados, el más relevante es el gráfico de cuantil-cuantil, ya que permite comprobar si una distribución ajusta la distribución real de los datos a través de la linealidad. El gráfico de cuantil-cuantil se toma de referencia permanente para comprobar la bondad del ajuste a los datos disponibles. La razón: la linealidad del gráfico se puede verificar rápidamente con un simple chequeo visual.

También se emplean en el análisis los gráficos de exceso medio para ajustar la distribución generalizada de Pareto. Con el estudio de dichos gráfico encontraremos el punto a partir del cual, los excesos por encima del mismo ajustan dicha distribución.

Conocer los límites máximos de pérdidas que puede sufrir una empresa es una información que interesa a todos aquellos que intervienen en la gestión del riesgo. Está claro que el cálculo de los niveles de pérdidas máximas que pueden poner en peligro la continuidad de una empresa implica tener en cuenta variables y circunstancias excepcionales que pueden ocasionar accidentes o siniestros de extrema gravedad. La Teoría del Valor Extremo, como se ha comentado en la introducción, puede ayudar a predecir estas variables y situaciones extremas y, por tanto, llevar a cabo un mejor ajuste de las primas o las provisiones técnicas del sector asegurador sujeto a este tipo de riesgos (Fundación MAPFRE Estudios, 1994).

En el presente epígrafe se resume la Teoría del Valor Extremo a partir de los textos generales de Embrechts, Kluppelberg y Mikosch (1997), de Reiss y Thomas (2001) y de Beirlant, Teugels y Vynckier (1996) dedicados globalmente a estudiar la aplicación de esta compleja teoría estadística en la resolución de problemas tradicionales de gestión de riesgos en seguros y finanzas.

Dado un conjunto de variables aleatorias independientes e idénticamente distribuidas,  $X_{1,n}, X_{2,n}, \dots, X_{k,n}, \dots, X_{n,n}$ , que en el presente trabajo representan las cuantías de siniestralidad derivadas de la ocurrencia de catástrofes, se desea conocer la función de distribución que ajusta dichas variables aleatorias. Por ello, en primer lugar introducimos las dos distribuciones de probabilidad fundamentales en la modelación de los valores extremos, distribución del valor extremo generalizada y distribución generalizada de Pareto, a partir de los teoremas que las desarrollan:

<sup>1</sup> XTREMES Versión 3.01, es un programa informático desarrollado por la Universidad de Siegen para el análisis de valores extremos.

- Teorema de Fisher y Tippet.
- Teorema de Pickands-Balkema-de Haan.

### 1. Teorema de Fisher y Tippet

El desarrollo de este teorema llevó a la obtención de la distribución generalizada del valor extremo que modela los máximos (ó mínimos) de series de valores. Bajo ciertas condiciones, se demuestra que la distribución de los máximos estandarizados de estas series, converge hacia las distribuciones de Gumbel, Weibull y Frechet, que surgen de la expresión de una distribución asintótica,  $G$ , y cuya forma general es la distribución generalizada del valor extremo (Generalized Extreme Value Distribution - GEV).

Los tres tipos de límites de dicha distribución asintótica  $G$  son:

- Distribución de Gumbel para colas medias.

$$G_0(x) \sim \exp\{-\exp(-x)\}, x \in \mathbb{R}$$

- Distribución de Frechet para colas gruesas.

$$G_1(x) \sim \begin{cases} 0, & x \leq 0 \\ \exp\{-2x^{-\alpha}\}, & x > 0 \rightarrow \alpha > 0 \end{cases}$$

- Distribución de Weibull para colas cortas o suaves.

$$G_2(x) \sim \begin{cases} 1, & x > 0 \\ \exp\{-2(2x)^{-\alpha}\}, & x \leq 0 \rightarrow \alpha < 0 \end{cases}$$

A partir de ciertas transformaciones y de una redefinición del parámetro  $\alpha$  como  $\xi \sim 1/\alpha$  se obtiene la expresión que configura la distribución del valor extremo generalizada o distribución límite de los máximos normalizados:

$$G_\xi \sim \exp\left\{-2\left(1 + \xi \frac{x - a_n}{b_n}\right)^{-1/\xi}\right\}, 1 + \xi \frac{x - a_n}{b_n} > 0$$

o bien, expresando de forma explícita los parámetros de localización y escala,

$$G_{\xi, a_n, b_n} \sim \exp\left\{-2\left(1 + \xi \frac{x - a_n}{b_n}\right)^{-1/\xi}\right\}, \begin{matrix} \xi, a_n \in \mathbb{R} \\ b_n > 0 \end{matrix}$$

Resumiendo, la distribución generalizada del valor extremo presenta tres parámetros fundamentales: el parámetro de localización ( $a_n$ ), el parámetro de escala ( $b_n$ ) y el parámetro de forma o índice de cola ( $\xi$ ). Este último juega un papel fundamental ya que indicará el grueso de la cola de la distribución, cuanto mayor sea el índice más gruesa será la cola, y determina si la cola de la distribución es finita o infinita:

$\xi > 0$	Distribución de Frechet	Distribuciones de cola gruesa (Heavy tail)
$\xi < 0$	Distribución de Weibull	Distribución con cola finita o suave (Light tail)
$\xi = 0$	Distribución de Gumbel	Distribución de cola media (Medium tail)

### 2. Teorema de Pickands - Balkema - De Haan

Este teorema da lugar a la distribución que recoge el comportamiento probabilístico de los excesos por encima de un determinado valor denominado **umbral o prioridad**. Es decir, describe el comportamiento de la pérdida extrema, en exceso de dicho umbral.

Los valores que exceden de un determinado umbral juegan un papel fundamental en diversos campos, en especial en el campo actuarial y más concretamente en las aplicaciones al reaseguro de exceso de pérdida.

Dada una prioridad  $u$  en el reaseguro exceso de pérdidas (Excess of Loss - XL), el reasegurador debe pagar la cuantía  $X-u$ , siendo  $X > u$ . Resulta, por tanto, fundamental conocer si los datos por encima de  $u$ , pueden aproximarse mediante alguna distribución.

Bien, pues es posible demostrar que cuando el umbral alcanza un valor suficientemente grande, en virtud del teorema indicado, los datos por encima de ese punto siguen una distribución generalizada de Pareto que modelará de manera eficiente la distribución de los excedentes del umbral.

Esta distribución de los excesos por encima del valor  $u$  es una distribución condicionada, esto es ajusta la cola de la distribución una vez que los datos han superado el umbral:

$$F^u(y) = P(X \geq u \leq y / X > u) = P(Y \leq y / X > u)$$

$$\text{para } 0 \leq y \leq x_0 - u$$

donde  $X$  representa la variable aleatoria cuantía del siniestro e  $Y$  el exceso sobre el umbral.

Es extremo derecho finito o infinito de la función de distribución se define como  $x_0$  y se sabe que el extremo inferior de la distribución, el punto izquierdo, es cero.

La dificultad, sin embargo, se encuentra en la elección del umbral apropiado, ya que la teoría no propone una solución general al respecto (Beirlant, J.; Teugels, J.L; Vynckier, P. (1996)). La función de distribución de Pareto generalizada es la siguiente:

$$W_{\gamma, \beta}(x) = \frac{1}{2} \left( 1 + \frac{x}{\beta} \right)^{-2/\gamma} \text{ para } \gamma \neq 0$$

$$W_{0, \beta}(x) = \frac{1}{2} \exp\left(-\frac{x}{\beta}\right) \text{ para } \gamma = 0$$

$$\text{Para } \gamma \geq 0 \rightarrow x \geq 0$$

$$\text{Para } \gamma < 0 \rightarrow 0 \leq x \leq \gamma/\beta$$

donde  $\gamma$  y  $\beta$  representan los parámetros de forma y escala respectivamente.

En términos generales, cuanto mayor sea el índice de cola o parámetro de forma ( $\gamma$ ) más gruesa será la cola de la distribución, en el sentido de que convergerá más lentamente a su asíntota y por tanto el ajuste realizado con los parámetros estimados dará lugar a mayores valores esperados (mayor cuantía estimada de pérdida).

Se considera que la distribución es de cola fuerte para  $\gamma > 20,5$ .

## ASPECTOS ESTADÍSTICOS Y ESTUDIO DE LA MUESTRA

En este apartado se aplica la Teoría del Valor Extremo al conjunto de datos disponibles, abordando los siguientes aspectos:

- Estudios acerca del grosor de la cola. Para ello se analizan, a través del gráfico de cuantil-cuantil, los cuantiles empíricos de la muestra frente a los cuantiles teóricos de la distribución exponencial. En caso de que los datos presenten una cola gruesa, se aproximarán diversas distribuciones alternativas.
- Ajuste de la distribución generalizada de Pareto. Se debe elegir el umbral sobre el que se calculan los excesos. Para efectuar dicha elección, resultará útil la herramienta gráfica de la función de exceso medio.

Para aclarar conceptos, antes de seguir avanzando en el tema, un *cuantil* es un estadístico de posición definido como un valor de la variable aleatoria por debajo del cual se encuentra una determinada proporción de los valores de la variable. Dicho de otra forma, el *cuantil* de orden  $\alpha$  de una variable aleatoria es un valor de dicha variable por debajo del cual se acumula un  $\alpha$  por ciento de probabilidad. Los cuantiles pueden ser de diversos tipos, por ejemplo percentiles cuando la serie de valores se divide en cien partes cada una de ellas equivalente a un uno por ciento, *cuartiles* si la serie de valores se clasifica en cuatro partes cada una relacionada con el 25% o *deciles* cuando el conjunto de valores se reparte en 10 partes cada una relativa al diez por ciento.

Los datos utilizados para la realización del artículo hacen referencia a las pérdidas por catás-

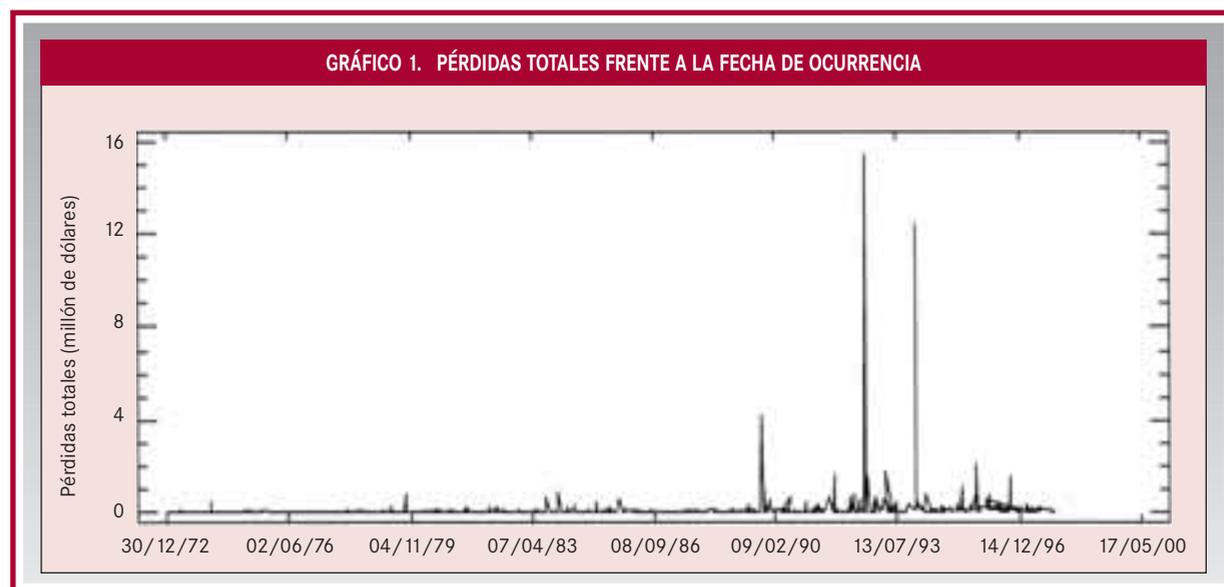
## ESTUDIO

trofes naturales ocurridas en los EE.UU. desde el año 1973 y hasta el año 1997. En la serie inicial disponible de datos se indica el Estado en el que se produce la catástrofe, la fecha de inicio y de finalización de la misma, su año de ocurrencia, un número identificador de la catástrofe sobrevinida y el total de pérdidas, en dólares constantes, ocasionado por cada una de ellas.

Para realizar el estudio, en primer lugar ordenamos los datos por año de ocurrencia, y ba-

sándonos en la fecha de inicio de las catástrofes y en el número identificador de las mismas, obtenemos el total de pérdidas asociadas a una misma catástrofe eludiendo el hecho de que hayan afectado a diversos Estados al mismo tiempo.

A continuación se presenta el gráfico del conjunto de datos históricos disponibles y las características esenciales de la muestra que se desprenden del mismo:



<b>Tamaño de la muestra: 878</b> <b>Mínimo: 500</b> <b>Máximo: 15.500.000</b>		<b>Media: 110.491</b> <b>Desviación estándar: 704.781</b> <b>Coefficiente Asimetría: 18,3592</b>	
Percentiles (q)	Valor para el percentil (Xq)	Número de valores por encima de Xq (k)	
q=0,25	$X_{0,25}=8.000$	k=658	
q=0,50	$X_{0,50}=25.000$	k=439	
q=0,75	$X_{0,75}=65.000$	k=219	
q=0,90	$X_{0,90}=150.000$	k=87	
q=0,95	$X_{0,95}=365.000$	k=43	
q=0,99	$X_{0,99}=16.000.000$	k=8	

## 1. Gráficos de cuantil-cuantil

Para determinar la familia de distribuciones a la que pertenecen los extremos observados en la muestra, ordenados de menor a mayor, se comparan, a través del gráfico de cuantil-cuantil, los cuantiles empíricos de dicha muestra con la función de cuantiles de la distribución generalizada de Pareto para distintas estimaciones de los parámetros de la misma.

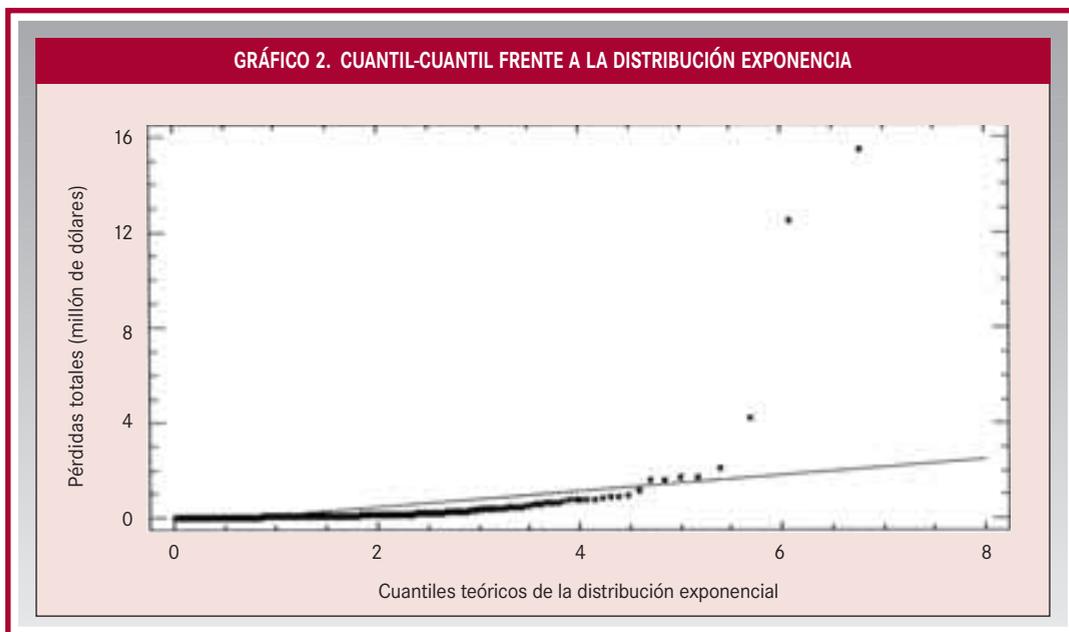
En el gráfico analizado en este epígrafe, se representa como variable dependiente los cuantiles muestrales ordenados de menor a mayor,  $X_{r,n}$ <sup>2</sup>, y como variable independiente los cuantiles teóricos de la función de cuantiles estimada. Dichos cuantiles se obtienen a partir de los percentiles

estimados a través de la expresión  $p = \frac{r}{n+1}$ ,<sup>0</sup> bien  $p = \frac{r-1/3}{n+1/3}$ , como sigue:

$$X_{r,n} \approx F_n^{-1}\left(\frac{r}{n+1}\right) \text{ con } r = 1, 2, \dots, n$$

donde  $F_n^{-1}$  representa la inversa de la función de distribución tomada como referencia que en nuestro caso será la exponencial.

Esta representación permite observar si los datos provienen de una distribución con una cola más o menos gruesa que la distribución exponencial y se utiliza como una primera aproximación al comportamiento de la muestra.



Si los datos de la muestra se generan mediante la distribución de referencia, el gráfico de cuantiles debe mostrar una tendencia fuertemente lineal (Embrechts et al., 1997). Esto es, si el mode-

lo exponencial elegido en este caso ajusta bien los datos, el gráfico debe ser lineal, y cuanto más lineal sea mayor será la bondad del ajuste realizado.

<sup>2</sup> Si una muestra de  $n$  valores se ordena de menor a mayor en una jerarquía creciente,  $X_{1,n} \leq X_{2,n}, \dots, \leq X_{r,n}, \dots, \leq X_{n,n}$  el subíndice  $r$  representa la posición en esa ordenación, de modo que  $r = 1$  y  $r = n$  indican, respectivamente, el menor y el mayor de los valores muestrales.

Por otra parte, si la distribución real de los datos tiene una cola más gruesa que la distribución de referencia, el gráfico se curvará por debajo de la línea de ajuste a la izquierda y/o por encima a la derecha.

Es evidente que el gráfico obtenido no muestra una tendencia lineal en el ajuste, lo que implica que la muestra no proviene de una distribución exponencial. Además, los datos se curvan por debajo de la línea de ajuste a la izquierda, indicativo de que la muestra presenta una cola más gruesa que la cola de la distribución exponencial.

## 2. Ajuste de la distribución generalizada de Pareto a los valores por encima de un umbral $u$

Las herramientas gráficas empleadas en el análisis presentado en este epígrafe son las siguientes:

### A) Gráficos de exceso medio

Este gráfico representa la esperanza de los valores por encima de  $u$  una vez que han superado ese valor, es decir, la media esperada de una función de distribución condicionada,  $e_{F(u)}$

$$e_F(u) = E(X - u \mid X > u) = \int_u^\infty (x - u) f(x) dx$$

$e_{F(u)}$  es conocido habitualmente como el *valor medio de exceso sobre el valor del umbral  $u$*  y en un contexto asegurador,  $e_{(u)}$  se interpreta como la siniestralidad esperada por encima de la prioridad  $u$ , cuando la cuantía posible de los siniestros es ilimitada.

Esta función será uno de los puntos básicos a la hora de verificar la hipótesis de Pareto generalizada en la cola de una distribución ya que la función de exceso medio para datos de cola gruesa se localiza entre la función de exceso medio de una distribución exponencial, que es constante, y la de la distribución generalizada de Pareto, que es lineal y tiende a infinito para valores altos de  $u$

(Embrechts, et al. 1997). En general, las distribuciones de cola gruesa dan lugar a una función de exceso medio que tiende a infinito para valores altos de  $u$  y presentan una forma lineal con pendiente positiva.

El gráfico de exceso medio se genera a través de la función empírica  $\hat{e}_n(u)$  para distintos valores de la prioridad,  $u$ , obtenidos de datos de la propia muestra, de modo que  $u \leq X_{k11}$  y, por tanto,  $\hat{e}_n(X_{k11})$ . Además, dicho gráfico tendrá las coordenadas  $(X_{k11}; E_{k,n})$  para  $k = 1, \dots, n-1$ , aunque también puede obtenerse con el par de valores:  $(k; E_{k,n})$  para  $k = 1, \dots, n-1$ . Ambas representaciones deben tener una apariencia prácticamente simétrica.

La función empírica de exceso medio,  $\hat{e}_n(X_{k11})$ , se estima a partir de la siguiente expresión dada una muestra representativa ordenada  $(X_{1,n}, \dots, X_{n,n})$ :<sup>3</sup>

$$E_{k,n} = \hat{e}_n(X_{k11}) = \frac{\sum_{i=1}^k X_i}{k} - 2X_{k11}, \quad k = 1, \dots, n-1$$

que es la media aritmética de los  $k$  valores que exceden el umbral  $u \leq X_{k11}$ .

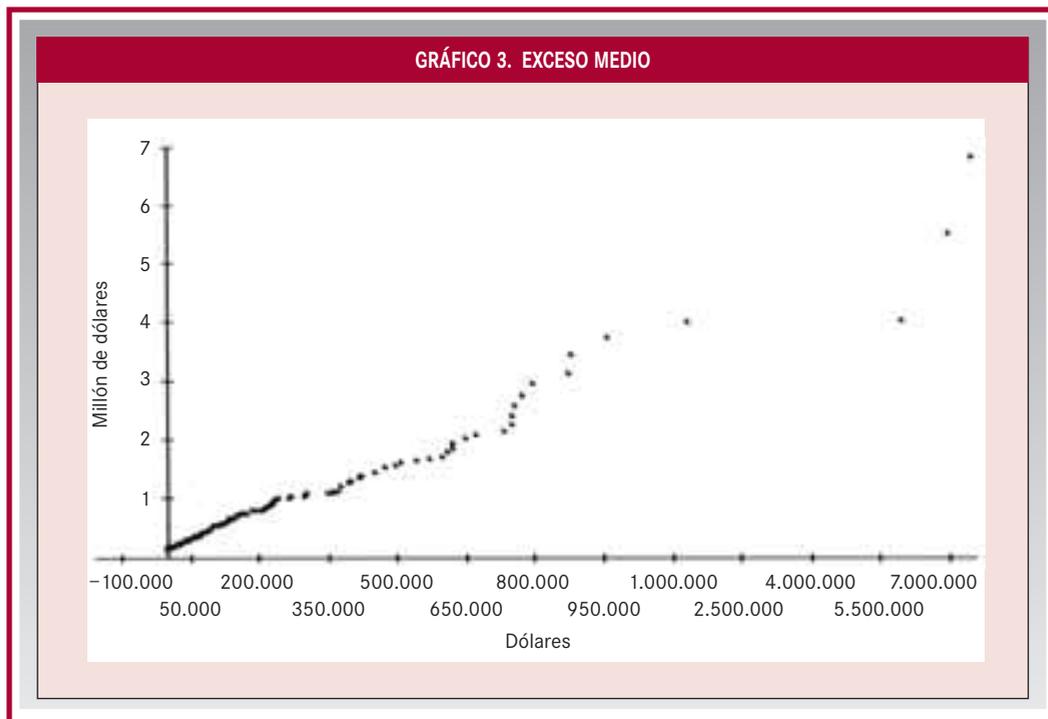
A continuación se muestra el gráfico de exceso medio de la muestra en cuestión, representando el siguiente par de valores  $(X_{k11}; E_{k,n})$  para  $k = 1, \dots, n-1$ . (Ver gráfico de exceso medio).

Si los puntos representados muestran una tendencia ascendente, es señal de una comportamiento de cola gruesa en la muestra representada. Si los datos se distribuyeran según una distribución exponencial mostrarían un gráfico aproximadamente lineal y los puntos de una distribución de cola corta o suave mostrarían una tendencia descendente.

De la observación del gráfico de exceso medio se desprende una tendencia ascendente de los datos, indicativa de una distribución de cola gruesa en los mismos. Esto sugiere que los valores de la muestra pueden ser modelados con éxito mediante la distribución generalizada de Pareto.

<sup>3</sup> Si una muestra de  $n$  valores se ordena de mayor a menor en una jerarquía decreciente,  $X_{1,n} \geq X_{2,n}, \dots, \geq X_{k,n}, \dots, \geq X_{n,n}$  el subíndice  $k$  representa la posición en esa ordenación, de modo que  $k = 1$  y  $k = n$  representan, respectivamente, el mayor y el menor de los valores muestrales.

GRÁFICO 3. EXCESO MEDIO



Como esta distribución se ajusta a los datos que exceden un determinado umbral, resulta necesaria la elección de dicho umbral como el punto a partir del cual la función empieza a crecer. El crecimiento de la función se inicia en los valores más inferiores aunque se observa un crecimiento más pronunciado para valores que exceden los 1.135.000 dólares, lo que supone un valor de  $k59$ .

Elegir el umbral  $u$ , o análogamente el número de valores  $k$  que quedan por encima de  $u$ , es importante porque se pretende ajustar la distribución de Pareto generalizada a los valores que excedan ese valor.

Sin embargo, la elección del valor de  $k$  conlleva algunos problemas ligados a la elección entre varianza y sesgo, esto es, si se incrementa el número de observaciones, la estimación del índice de cola será más precisa (con menor varianza), ya que contará con mayores valores para la estimación, pero podría estar sesgada, ya que algunas observaciones del centro de la distribución podrí-

an introducir en la serie. Por otro lado, si se reduce el número de observaciones, se reduce el sesgo pero hace que la estimación del índice sea más volátil al realizarse con un menor número de observaciones.

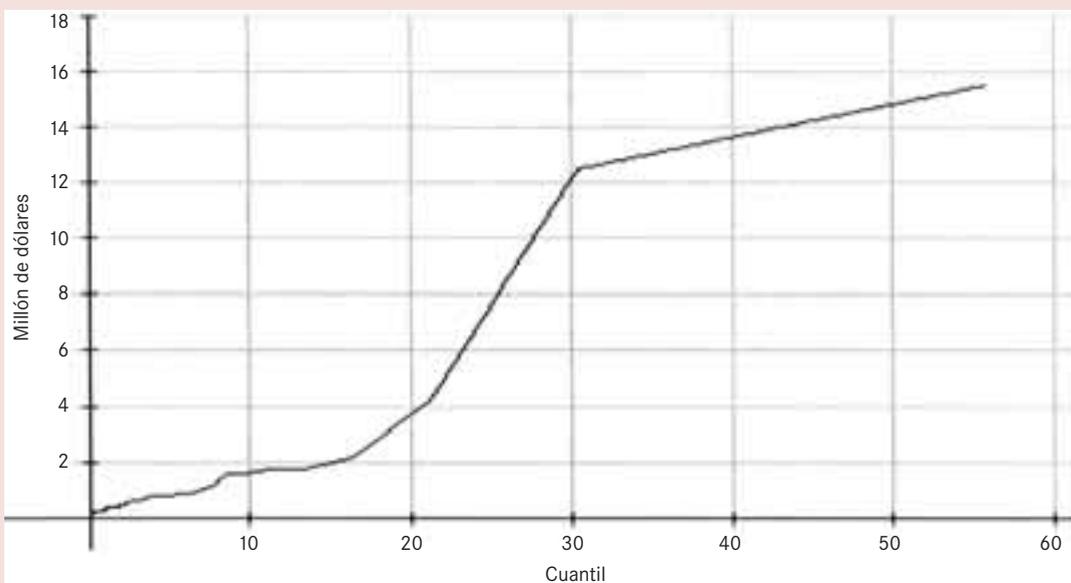
Seleccionado  $u$ , se estiman los parámetros de la distribución, índice de escala ( $\beta$ ) e índice de cola ( $\gamma$ ) siendo el parámetro de localización el propio umbral elegido si se ajustan los excedentes o cero si se ajustan los excesos.

### B) Estimación del modelo (distribución generalizada de Pareto) para el total de valores muestrales

Tras ajustar la distribución de Pareto generalizada a los excedentes de varios umbrales, se obtienen unos gráficos de cuantil-cuantil lejanos de una línea recta, como sería de esperar ante ajustes óptimos.

El gráfico de cuantil-cuantil para los distintos umbrales resulta similar al gráfico presentado a continuación:

GRÁFICO 4. CUANTIL - CUANTIL



En el eje de ordenadas se han presentado los cuantiles muestrales, en el eje de abscisas los cuantiles empíricos estimados a través de la función de cuantiles estándar (inversa de la función de distribución) a partir del índice de cola estimado por el programa XTREMES y parámetros de localización y escala 0 y 1 respectivamente. El gráfico está soportando por 93 puntos que exceden un umbral de 150.000. Entre dichos puntos se ha realizado una interpolación lineal. El parámetro de la función de cuantiles se ha estimado a partir de dichos 93 puntos.

Como se puede comprobar visualmente, el gráfico está lejos de aproximarse a una línea recta. Sin embargo, la desviación se produce como consecuencia de los dos últimos valores muestrales más elevados. Para evitar esta distorsión la solución propuesta, consiste en eliminar dichos valores y por tanto modelar sin los picos mayores.

### C) Estimación del modelo (distribución generalizada de Pareto) para los valores muestrales menos dos picos

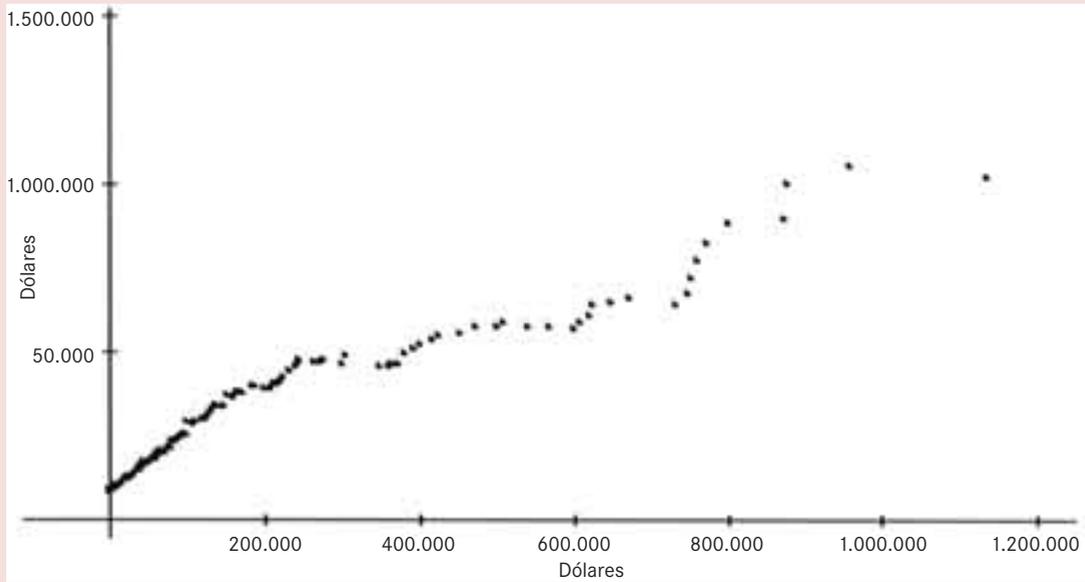
El nuevo gráfico de exceso medio corregido después de eliminar los valores más altos se representa en el gráfico 5.

De nuevo se observa una tendencia creciente indicativa de ajuste de la distribución generalizada de Pareto, por lo que, como en el apartado anterior, en primer lugar se seleccionan varios umbrales para ajustar la distribución generalizada de Pareto a la cola de la distribución. Una vez seleccionados dichos umbrales, se parámetros índice de cola e índice de escala utilizando el método de máxima verosimilitud, siendo el parámetro de localización será el propio umbral elegido, ya que la distribución se ha truncado por la izquierda en dicho valor.

A continuación se presentan los parámetros para los distintos umbrales junto con el gráfico de cuantil-cuantil, que visualmente mostrará la bon-

# ESTUDIO

GRÁFICO 5. EXCESO MEDIO CORREGIDO



dad del ajuste realizado. Se ha superpuesto una línea en el gráfico para comprobar visualmente si existe o no linealidad en el mismo:

En los gráficos puede observarse que a medida que se ha disminuido la prioridad se produce un mejor ajuste para los cuantiles más altos, lo que

GRÁFICO 6. CUANTIL-CUANTIL

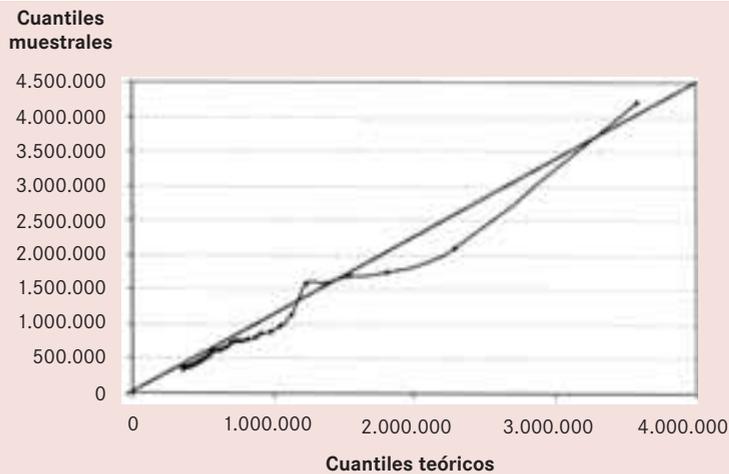


Gráfico de cuantil-cuantil para los valores que excedan la prioridad de 350.000

	k=43 valores
Índice de cola ( $\gamma$ )	0,445358
P. escala ( $\beta$ )	265.737
P. localización	350.000

**GRÁFICO 7. CUANTIL-CUANTIL**

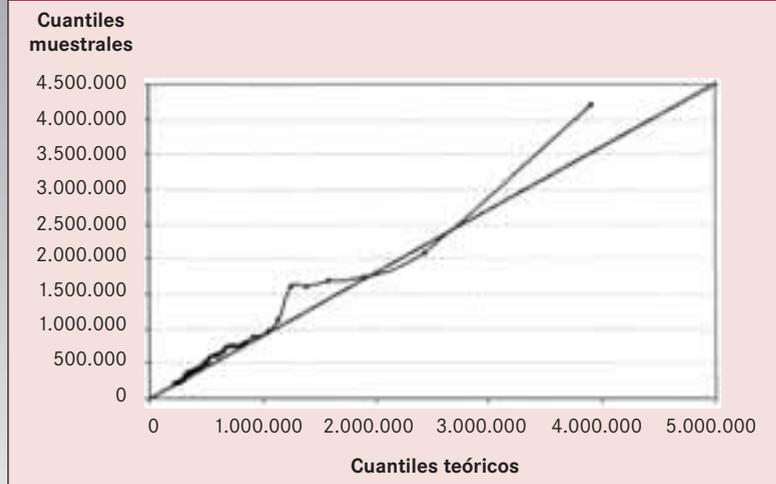


Gráfico de cuantil-cuantil para los valores que excedan la prioridad de 210.000

	<b>k=68</b>
Índice de cola ( $\gamma$ )	0,47864
P. escala ( $\beta$ )	217.174
P. localización	210.000

**GRÁFICO 8. CUANTIL-CUANTIL**

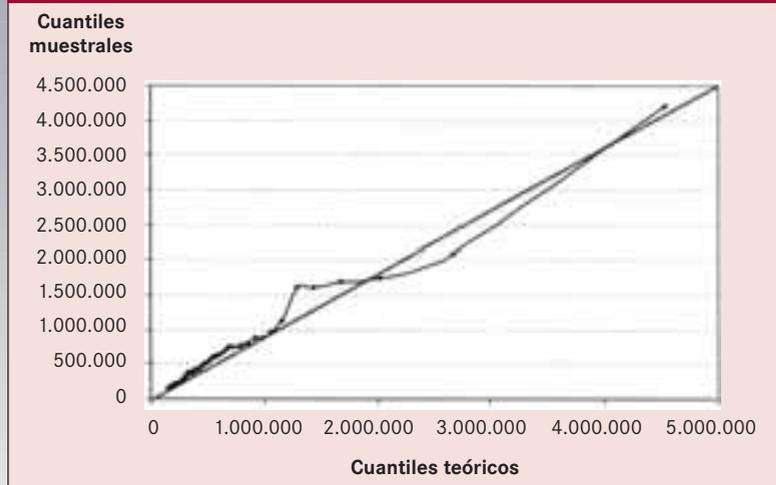


Gráfico de cuantil-cuantil para los valores que excedan la prioridad de 150.000

	<b>k=91</b>
Índice de cola ( $\gamma$ )	0,550755
P. escala ( $\beta$ )	171.889
P. localización	150.000

también se comprueba a través del índice de cola ya que a mayor índice más gruesa es la cola de la distribución y por tanto, mejor se ajustan los va-

lores más altos. Además, los cuantiles empíricos más altos estimados con los primeros ajustes (umbrales de 350.000 y 210.000) quedan algo in-

fravalorados, mientras que con el último ajuste (umbral de 150.000), los cuantiles empíricos quedan algo sobre valorados.

Cabe señalar que los ajustes han sido muy próximos y que las desviaciones máximas, de los cuantiles estimados respecto a los valores muestrales observados son aproximadamente de un 25%. Dichas desviaciones son los puntos mostrados en el gráfico distanciados de la línea trazada.

### D) Ajuste y estimulación de la cola

El ajuste de la distribución de Pareto a la cola de la distribución se ha realizado con el objetivo de predecir la probabilidad de una catástrofe de

cierta cuantía o la cuantía esperada de siniestralidad para alguna frecuencia.

En la siguiente tabla se muestran las estimaciones esperadas de catástrofes en función de los parámetros estimados con el programa XTREMES para distintos valores de  $k$  (y por tanto del umbral  $u$ ) para los modelos de la distribución generalizada de Pareto considerando los percentiles 99,5%; 99,9% y 99,99%.

Elegido un umbral, por ejemplo 150.000 dólares, si ocurre una catástrofe que supere dicha prioridad, es de esperar que sea de importe superior a 5.613.453 dólares con una probabilidad del 0,5%. Del mismo modo si ocurre una

U	k	Índice de cola	P. de escala	P. de localización	Cuantiles estimados		
					$X_{0,995}$	$X_{0,999}$	$X_{0,9999}$
350.000	43	0,445358	265.737	350.000	6.070.522	12.689.817	37.825.695
210.000	68	0,47864	217.174	210.000	5.486.390	12.136.217	37.026.221
150.000	91	0,550755	171.889	150.000	5.613.453	13.851.661	49.647.110

catástrofe que supere el umbral de 150.000 es de esperar que el importe de una de las próximas 1.000 catástrofes supere la cuantía de 13.851.661 dólares.

Por otra parte, las probabilidades de que se superen ciertas cuantías una vez superada la cuantía del umbral se puede resumir en la siguiente tabla:

u	K	Probabilidades estimadas			
		5.000.000	15.000.000	25.000.000	50.000.000
350.000	43	0,759%	0,069%	0,022%	0,0048%
210.000	68	0,602%	0,065%	0,023%	0,0054%
150.000	91	0,613%	0,087%	0,035%	0,0099%

Elegido un umbral, por ejemplo de 150.000, la probabilidad de que la cuantía de la catástrofe supere los 5 millones de dólares después de haber superado el umbral indicado, es de 0,613% y

la probabilidad de que supere los 50 millones es de casi uno por cada 10.000 siniestros.

Después de estimar la función de distribución condicional, la estimación de toda la distribución

de la cola puede obtenerse por la siguiente relación (Embrechts (1997)):

$$\bar{F}(y) \approx \bar{F}(u) + \bar{F}'(u) (y - u)$$

La aproximación de  $\bar{F}'(y)$  viene dada por la distribución generalizada de Pareto que ya se ha realizado:

$$\bar{F}'(y) \approx \bar{W}_{\hat{\gamma}, \hat{\beta}}(y) \quad y \geq 0$$

La aproximación de  $\bar{F}(u)$  puede obtenerse a partir de la función de distribución empírica,  $F_n(u)$  que para series de datos univariantes  $[x_1, x_2, \dots, x_n]$  es el número relativo de valores  $x_i$  que son menores a  $u$  de modo que:

$$F_n(u) \approx \frac{1}{n} \sum_{i=1}^n I(x_i < u)$$

donde la función  $I$  indica el recuento de valores inferiores a  $u$ , lo que representa la proporción o frecuencia relativa de aquellos datos que no son mayores que el valor  $u$ .

La función complementaria vendrá dada por el cociente entre los valores superiores o iguales al umbral  $u$  y el total de valores  $n$ .

$$\bar{F}_n(u) \approx \frac{1}{n} \sum_{i=1}^n I(x_i \geq u) = \frac{N_u}{n}$$

El resultado queda por tanto:

$$P(X > y | u) \approx \bar{F}(y | u) \approx \bar{F}(u) + \bar{F}'(u) (y - u)$$

$$\approx \frac{N_u}{n} \left( 1 + \frac{y - u}{\hat{\beta}} \right)^{-2\hat{\gamma}}$$

y la estimación del cuantil  $X_p$ , es decir, el valor que deja a su izquierda la probabilidad  $p$  sería:

$$X_p \approx u + \hat{\beta} \left[ \left( \frac{1-p}{\bar{F}(u)} \right)^{\frac{1}{2\hat{\gamma}}} - 1 \right]$$

En la siguiente tabla, se muestran las probabilidades conjuntas de que se superen las cuantías de 5, 15, 25 y 50 millones de dólares a partir de la muestra disponible y después de haber eliminado los dos mayores valores muestrales. (Ver tabla en la parte inferior de la página).

Como se observa las estimaciones están bastante próximas aunque comienzan a diferir en el caso del mayor de los percentiles. (Ver tabla en la página siguiente).

## CONCLUSIONES

A partir del análisis realizado a lo largo del artículo, podemos concluir que es posible ajustar las catástrofes que exceden de una determinada cuantía o umbral a una distribución generalizada de Pareto. Y, por tanto, podemos determinar la cuantía de las catástrofes que se producirán en un futuro con una determinada probabilidad o bien la probabilidad de que ocurra una catástrofe de determinada cuantía.

Se ha de señalar, sin embargo, que la distribución de Pareto generalizada que se ha ajustado a los datos es una distribución condicionada, lo que implica que las probabilidades que proporciona el ajuste es-

u	N <sub>u</sub>	Probabilidades estimadas			
		5.000.000	15.000.000	25.000.000	50.000.000
350.000	43	0,03724%	0,00339%	0,00109%	0,00023%
210.000	68	0,04672%	0,00503%	0,00175%	0,00042%
150.000	91	0,06369%	0,00900%	0,00359%	0,00103%

$u$	$N_u$	Cuantiles estimados		
		$X_{0,995}$	$X_{0,999}$	$X_{0,9999}$
350.000	43	1.403.514	3.132.622	9.176.233
210.000	68	1.442.339	3.399.030	10.722.837
150.000	91	1.497.285	3.864.217	14.148.661

tán condicionadas a que el siniestro previamente exceda la cuantía tomada como prioridad cuya elección integra la mayor complicación del análisis: cuál es el umbral a elegir que proporcione el mejor ajuste. En el presente trabajo hemos utilizado tres prioridades, de las cuales la prioridad de 150.000 dólares ha proporcionado unos mejores ajustes en la cola.

No obstante, cabe decir que la estadística de los valores extremos no pierde la condición de ser una ciencia no exacta, y, por tanto, los resultados obtenidos en el estudio deben ser tomados como una aproximación en previsión de lo que realmente pueda ocurrir.

## BIBLIOGRAFÍA

- ADLER, R.; FELDMAN, R.; TAQQU, M.; (EDS.) (1997). *A practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhäuser. Boston.
- ADLER, R.; FELDMAN, R.; TAQQU, M.; (EDS.) (1997). *A practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhäuser. Boston.
- BEIRLANT, J.; TEUGELS, J. L.; VYNCKIER, P. (1996). *Practical Analysis of Extreme Values*. Leuven University Press. Leuven.
- COLES, S. (2001). *An Introduction to Statistical Modelling of Extreme Value*. Springer-Verlag. Londres.
- EMBRECHTS, P.; RESNICK, S. I.; SAMORODNITSKY, G. (1999). «Extreme Value Theory as a Risk management Tool»; *North American Actuarial Journal*, Volume 3, nº 2, 30-41.
- EMBRECHTS, P.; KLÜPPELBERG, C.; MIKOSCH, T. (1997). *Modelling extremal events for Insurance and Finance*. Applications of Mathematics. Springer-Verlag Berlin Heidelberg.
- Fundación Mapfre Estudios (1994). «Evaluación de riesgos. Estimación de pérdidas máximas en siniestros»; *Revista Gerencia de Riesgos y Seguros nº 48. 4º Trimestre de 1994. Fundación MAPFRE Estudios*.
- CALAMBOS, J. (1987). *The asymptotic theory of extreme order statistic*. Krieger, Florida. 2ª Edición.
- JUNG, J. (1965). «On the use of extreme values to estimate the premium for an excess of loss reinsurance». *Astin Bulletin, Vol. III, Pt. II, 178-184*.
- MCNEIL, A. J.; SALADIN, T. (1997). «The Peak over Thresholds Method for Estimating High Quantiles of loss Distributions». [www.math.ethz.ch/~mcneil/pub\\_list.html](http://www.math.ethz.ch/~mcneil/pub_list.html)
- REISS, R. D.; THOMAS, M. (2001). *Statistical Analysis of Extreme Values (with applications to insurance, finance, hydrology and other fields)*. Birkhäuser Verlag. 2ª Edición.