

ANÁLISIS DE LA FRECUENCIA DE OCURRENCIA DE VALORES EXTREMOS: UNA APLICACIÓN AL RAMO DE AUTOMÓVILES EN ESPAÑA

María José Pérez Fructuoso¹ y Almudena García Pérez²

Profesora Visitante de Matemática Actuarial, Departamento de Economía de la Empresa. Universidad Carlos III de Madrid.
Profesora Asociada, Departamento de Ciencias Empresariales. Universidad de Alcalá de Henares.

RESUMEN

El estudio de los riesgos extremos, igual que los riesgos de masa, implica la medición tanto de la magnitud de su siniestralidad como de la frecuencia de ocurrencia de los mismos. Centrándonos en el recuento de extremos en escala nominal, este artículo analiza, desde una perspectiva teórica, el modelo no paramétrico de los excedentes de Gumbel para medir la frecuencia en su acaecimiento y desarrolla una aplicación de dicho modelo a los datos de siniestralidad de responsabilidad civil en el ramo de automóviles para dos compañías de seguros que operan en el mercado asegurador español.

PALABRAS CLAVE

Teoría del Valor Extremo, Ordenación de Riesgos, Distribución Binomial, Distribución Hipergeométrica, Tiempo Medio de Espera entre Extremos.

¹ Avenida de la Universidad Carlos III, 22. 28270. Colmenarejo, Madrid. Teléfono: +34 91 856 13 08 Fax: +34 91 856 12 20 e-mail: mjpcruz@emp.uc3m.es

² Plaza de la Victoria, 2. 28802, Alcalá de Henares, Madrid. Teléfono: +34 91 885 52 19 Fax: +34 91 885 42 94. e-mail: ajmu.garcia@uah.es

1. INTRODUCCIÓN

La nueva normativa sobre solvencia, Solvencia II, establece tres componentes del riesgo que deben afrontar las compañías de seguros: volatilidad, incertidumbre y eventos extremos (IAA, 2004). Dichos componentes, en especial los dos primeros, son la base de toda investigación actuarial interesada en calcular la prima pura de la siniestralidad total de una cartera de pólizas y en estimar, en función de la volatilidad y la incertidumbre de los riesgos asegurados, los recargos de seguridad que le son aplicables. Este proceso de tarificación clásico de la siniestralidad, se fundamenta en la distribución central de los datos y en la ley de los grandes números. Pero los acontecimientos extremos, caracterizados por tener una baja probabilidad de ocurrencia y una elevada severidad, sobre todo en lo que a consecuencias económicas se refiere, provienen de distribuciones de cola gruesa (Pérez-Fructuoso et al. 2004), que muestran una mayor masa de probabilidad en la cola que las distribuciones tradicionalmente utilizadas como la normal. En estos casos, suponer normalidad en la distribución del fenómeno aleatorio, minimiza el efecto derivado de la ocurrencia de uno o más sucesos extremos que puede incrementar las fluctuaciones derivadas de siniestros de elevada cuantía muy por encima de las que cabría esperar como consecuencia de la volatilidad y la incertidumbre.

La Teoría del Valor Extremo proporciona a las aseguradoras un útil instrumento de gestión del riesgo (Embrechts, et al., 1999), ya que les permite inferir estadísticamente sobre los valores extremos de una población o proceso aleatorio y, con ello, estimar la probabilidad de que ocurran sucesos más extremos que los producidos hasta el momento del análisis. Mediante esta teoría se pueden predecir valores elevados en las colas y situaciones que superen todos los niveles históricos observados modelando los extremos separadamente de la muestra global, sin recurrir a la distribución del conjunto de datos observados. A través del estudio del riesgo de eventos extremos es posible asegurar la estabilidad y la solvencia de las entidades aseguradoras, a pesar de la ocurrencia de pérdidas extremas, aplicando modelos estadísticos que midan los riesgos y permitan establecer, óptimamente, el volumen de capital necesario, el nivel adecuado de

provisiones, la correcta tarificación, o las posibilidades de cesión al reaseguro.

En este artículo se realiza una aplicación de la Teoría del Valor Extremo al recuento, o frecuencia de ocurrencia, de los valores extremos a través de procedimientos estadísticos no paramétricos o de distribución libre. Los modelos no paramétricos permiten realizar inferencia basándose en los estadísticos de orden u ordenación de las variables aleatorias. En el campo actuarial, el método de recuento de los excedentes desarrollado por Gumbel (1958) posibilita la estimación de probabilidades y periodos de retorno asociados a la frecuencia de los extremos a partir de la ordenación de los siniestros. De esta forma, las compañías de seguros y reaseguros pueden calibrar sus necesidades financieras o de contratación de capacidad a través de reaseguro a partir de la probabilidad de ocurrencia de la siniestralidad grave determinada mediante la distribución de frecuencias de Gumbel.

La organización del artículo es la siguiente: en el epígrafe 2 se realiza una explicación teórica de todos los conceptos relativos al estudio de la frecuencia de los valores extremos que se van a aplicar en el análisis de los datos de siniestralidad por responsabilidad civil en el ramo de automóviles de dos compañías de seguros españolas. La sección 3 describe las series históricas analizadas y presenta los resultados obtenidos del ajuste del método de los excedentes de Gumbel sobre las mismas. Finalmente, en el apartado 4 se resumen las principales conclusiones alcanzadas en la aplicación práctica.

2. LA FRECUENCIA DE LOS VALORES EXTREMOS SEGÚN EL MÉTODO DE LOS EXCEDENTES DE GUMBEL

Uno de los principales problemas derivado del estudio de los eventos extremos es la medición de la frecuencia que se produce en este tipo de siniestralidad. En la estadística clásica, el recuento o frecuencia de las observaciones se modela a través de distribuciones discretas como la distribución Binomial, de Poisson, Geométrica o Hipergeométrica. Sin embargo, en el campo de los valores extremos, asumiendo la hipótesis de independencia entre las observaciones, la estadística que mide su frecuencia es la no paramétrica o de distribución libre.

El método de los excedentes de Gumbel es uno de estos modelos no paramétricos utilizados para la medición de la frecuencia de los valores extremos, el cual, a partir de la ordenación de variables aleatorias, permite obtener, de forma relativamente sencilla, información analítica acerca de los datos de siniestralidad estudiados. La aplicación de este método da respuesta a cuestiones fundamentales tales como el número de valores que, de entre las observaciones futuras, excederán una determinada posición en la jerarquía de los datos históricos.

En este apartado se analizan, desde una perspectiva teórica, los procedimientos estadísticos utilizados para estimar el número medio de valores extremos que ocurrirán en los próximos periodos.

2.1. La distribución de la frecuencia de los valores extremos

Sea $X_{1,n} \geq X_{2,n} \geq \dots \geq X_{k,n} \geq \dots \geq X_{n,n}$ un conjunto ordenado de observaciones independientes e idénticamente distribuidas, donde n es el número de datos históricos o tamaño de la muestra de referencia y k es la posición k -ésima u orden en una jerarquía decreciente³ y representa, por tanto, la k -ésima observación más alta.

Además consideramos:

- N : número de ensayos futuros o tamaño de la muestra para la que se quiere predecir (que no tiene por qué coincidir con n).
- h : número de observaciones iguales o superiores al valor k -ésimo de la muestra original, en los próximos N ensayos o periodos de tiempo, es decir, el número de valores extremos esperado.

Para determinar la distribución de h , de parámetros k , n y N , con el objetivo de analizar en cuántas ocasiones, de los N ensayos futuros, se igualará o excederá el valor k -ésimo de las n observaciones pasadas, definimos la variable aleatoria distribuida hipergeométricamente,

³ $k = 1$ es la mayor observación y $k = n$ es la menor observación.

$$S_N^n(k) = \sum_{h=1}^N I(X_h > X_{k,n})$$

que indica el número de valores h que se espera que sean superiores a $X_{k,n}$ en las próximas N observaciones (Embrechts, et al. 1997).

Denotamos por $w(h;n,k,N)$ la función de densidad de h , que determina, bajo la hipótesis de que las características de la muestra analizada se mantienen o no cambian sustancialmente en el tiempo, la probabilidad de que haya exactamente h valores en las próximas N observaciones que superan el k -ésimo valor más grande de las n observaciones pasadas.

Si P es la probabilidad de que un nuevo dato sea menor que el k -ésimo valor histórico, estamos ante una situación binomial (que supere o que no supere) con probabilidad de éxito desconocida. Entonces, integrando esta distribución Binomial para todos los posibles valores de P , la expresión de la función de densidad de la frecuencia de los valores extremos queda definida como sigue (Kinnison, 1985):

$$w(h;n,k,N) = P(S_N^n(k) = h) = \frac{\binom{N+n-k-h}{n-k} \binom{h+k-1}{k-1}}{\binom{N+n}{n}}$$

o bien,

$$w(h;n,k,N) = \frac{k \binom{n}{k} \binom{N}{h}}{(N+n) \binom{N+n-1}{h+k-1}}, \sum_{h=0}^N w = 1$$

donde $0 \leq h \leq N$ y $1 \leq k \leq n$.

Denotamos por $W(x; n, k, N)$ la función de distribución de n que indica la probabilidad de que haya como máximo x valores, en N observaciones futuras, que superen el valor k -ésimo de referencia de la muestra,

$$W(x; n, k, N) = \sum_{h=0}^x w(h; n, k, N)$$

A partir de estas funciones de densidad y de distribución definidas, resulta interesante destacar las siguientes probabilidades:

- La probabilidad de que el mayor valor ($k=1$) de las n observaciones pasadas no se sobrepase ($h=0$) en las N observaciones futuras, con $N=n$, es $w(0; N, 1, N) = 0,5$.
- La probabilidad de que el k -ésimo valor mayor de las n observaciones iniciales sea excedido al menos una vez ($x \geq 1$) en las N nuevas observaciones, con $N=n$, es $P(x \geq 1) = 1 - P(x = 0) = 1 - w(0; N, 1, N) = 0,5$.
- La probabilidad de que el mayor valor de las observaciones iniciales ($k=1$), se supere en todas las nuevas observaciones (N) viene dada por la siguiente expresión:

$$w(N; n, 1, N) = \frac{n!N!}{(N+n)!}$$

cuyo valor es prácticamente cero incluso para valores pequeños de n y N .

- Si la muestra de referencia tiene un número impar de observaciones y consideramos la mediana de la misma, $k = \frac{n+1}{2}$, entonces es equiprobable que dicha mediana sea superada x o $(N-x)$ veces en N ensayos futuros. Esto supone que la densidad del número de valores extremos sobre la mediana es simétrico.

En cuanto a los principales momentos de la distribución de la frecuencia de los valores extremos, pueden obtenerse a partir de las propiedades de las distribuciones Hipergeométrica y Binomial (Gumbel, 1958):

- El número medio de los N valores futuros, que excederán el k -ésimo valor más alto de los n valores pasados, viene dado por:

$$\bar{x}_k = k \frac{N}{n+1}$$

y coincide con la media de la correspondiente distribución Binomial.

Algunos casos particulares de \bar{x}_k , de especial de interés para el análisis de la frecuencia de los valores extremos son los siguientes:

- Si $N = n + 1$, el número medio es $\bar{x}_k = k$.
- El número medio de valores excedentes sobre el valor más pequeño de la muestra original ($\bar{x}_{k=n}$), es n veces el número medio de excedentes sobre el valor mayor ($\bar{x}_{k=1} = n\bar{x}_{k=1}$) de la misma muestra.
- Si el número de observaciones de la muestra original, n , es impar y la mediana es $k = \frac{n+1}{2}$, el número medio de excedentes sobre la mediana es $\frac{N}{2}$.
- Si n y N son elevados (tienden a infinito), el número medio de excedentes sobre el valor más alto ($k=1$) de la muestra original es aproximadamente 1.
- La varianza del número de valores que superan el k -ésimo valor más alto de la muestra original viene dada por,

$$V_k = \frac{kN(n-k+1)(n+N+1)}{(n+2)(n+1)^2}$$

Esta ecuación muestra un aumento de la varianza a medida que lo hace el número de observaciones futuras, N , y una disminución de la misma cuando incrementa el número de observaciones pasadas, n .

Las características principales de este momento central de orden dos son las siguientes:

- El valor máximo de la varianza se alcanza para la mediana de la muestra original, es decir, para $k = \frac{n+1}{2}$.
- El cociente entre la varianza respecto a la mediana y la varianza para el mayor valor extremo ($k=1$) de la muestra original es,

$$\frac{V_{\frac{n+1}{2}}}{V_1} = \frac{(n+1)^2}{4n}$$

Como la varianza del número de excedentes toma su mayor valor en la mediana de la distribución inicial y el menor valor en su mayor valor extremo, se deduce que la predicción sobre el extremo merece mayor confianza que la predicción sobre la mediana. Esta ventaja de los extremos incrementa con el tamaño de la muestra.

- El coeficiente de variación, definido a partir de la media y la varianza de la frecuencia de los valores extremos, resulta:

$$CV^2 = \frac{V_k}{\bar{x}_k^2} = \frac{(N+n+1)(n-k+1)}{N(n+2)k}$$

o bien,

$$CV = \frac{\sqrt{V_k}}{\bar{x}_k}$$

2.2. La ley de los excedentes raros

Decimos que un valor extremo es “raro” cuando k toma un valor pequeño respecto a las n observaciones pasadas. En este caso, la función de densidad de la frecuencia de los valores extremos viene dada por (Gumbel, 1958):

$$w(h; n, k, n) = w(h; k) \cong \binom{h+k-1}{h} \left(\frac{1}{2}\right)^{k+h}$$

A continuación se destacan algunas probabilidades importantes para el caso en que N y n son valores elevados y, además, k y h toman valores pequeños (Gumbel, 1958):

- La probabilidad de que h valores de las N observaciones nuevas, sobrepasen el valor k -ésimo de las n observaciones pasadas, viene dada por la siguiente expresión asintótica:

$$w(h; n, k, N) \cong \binom{h+k-1}{h} \frac{n^k N^h}{(N+n)^{k+h}}$$

- La probabilidad de que en las N observaciones futuras, el valor k -ésimo de los datos pasados nunca se exceda es,

$$w(0; n, k, N) \cong \left(\frac{n}{(N+n)}\right)^k$$

- La probabilidad de que el mayor valor de las n observaciones pasadas ($k=1$), se exceda h veces en las N observaciones nuevas es:

$$w(h; n, k, N) \cong \frac{n}{N+n} \left(\frac{N}{N+n} \right)^h$$

- La probabilidad de que h valores de las N observaciones nuevas, sobrepasen la observación de k -ésima de las n observaciones previas, cuando el tamaño de la muestra futura coincide con el de la muestra original ($N = n$), es independiente de n y tiene k como único parámetro.
- Finalmente, la probabilidad de que h valores de las N observaciones futuras, superen el mayor valor ($k = 1$) de las observaciones pasadas, cuando $N = n$, es:

$$w(h; 1) = \left(\frac{1}{2} \right)^{h+1}$$

Estas probabilidades asintóticas son muy útiles porque no dependen del tamaño muestral. Sin embargo no debe olvidarse que se han obtenido suponiendo muestras originales y futuras muy grandes y homogeneidad de la distribución subyacente a lo largo de tiempo.

2.3. El comportamiento asintótico de la distribución del número de extremos

Hasta este punto se ha analizado la distribución del número de valores extremos como una generalización de la distribución Binomial. Vamos a estudiar ahora el comportamiento asintótico de esta distribución, cuando las muestras analizadas tienen un tamaño elevado y los valores extremos se expresan en términos relativos, es decir, cuando se quiere inferir sobre un número N elevado de observaciones futuras y el valor de k es pequeño.

Denotamos por q la proporción entre el número de futuros valores extremos (h) sobre el número de observaciones nuevas (N),

$$q = \frac{h}{N}, \quad 0 \leq q \leq 1$$

Si N toma un valor elevado, se puede demostrar que este cociente puede aproximarse a través de variables continuas (Gumbel, 1958). Para ello, definimos en primer lugar la densidad de q como:

$$f(q; n, k, N) = Nk \binom{n}{k} \frac{N!(qN + k - 1)!(n - qN + n - k)!}{(N + n)!(qN)!(N - qn)!}$$

Expresión que cuando $N \rightarrow \infty$, se aproxima por la fórmula de Stirling:

$$f(q; n, k) = \binom{n}{k} kq^{k-1} (1-q)^{n-k}$$

La función de distribución de q se obtiene integrando la función de densidad y la solución es una integral recursiva de k y n cuya solución viene dada por la distribución Beta incompleta:

$$F(q; n, k) = \int_0^q f(t; n, k) dt = k \binom{n}{k} \int_0^q t^{k-1} (1-t)^{n-k} dt = BI(q; k, n - k + 1)$$

Las principales probabilidades a destacar en esta distribución son:

- La probabilidad de que, como máximo, una proporción q de las N observaciones nuevas, supere el valor más pequeño ($k = n$) de las n observaciones previas es,

$$F(q; n, n) = q^n$$

- La probabilidad de que en la futura muestra de tamaño N , como máximo, una proporción q de observaciones supere todas las observaciones previas ($k = 1$) es,

$$F(q; n, 1) = 1 - (1 - q)^n$$

Lógicamente, esta probabilidad también puede interpretarse como la fracción q de los nuevos datos muestrales que son menores que el valor más pequeño de la muestra original.

2.4. El periodo de retorno

Además del número de siniestros extremos y del volumen esperado de los mismos, es conveniente analizar el momento estimado de ocurrencia del suceso estudiado o el tiempo medio de espera hasta el acontecimiento del siguiente evento extremo, también denominado periodo de retorno.

Para estudiar dicho periodo, en primer lugar definimos la variable aleatoria momento de ocurrencia de un nuevo excedente por encima de un valor alto especificado, Z , en un horizonte temporal indefinido y realizamos observaciones por intervalos regulares de tiempo interrumpiendo el experimento cuando el valor Z se exceda una vez (Gumbel, 1958). Esto supone considerar una secuencia infinita de variables aleatorias $[X_1, X_2, \dots]$ con función de distribución común F .

Entonces, el momento r -ésimo en el que el valor X_t excede el valor Z se define como:

$$\tau_r = \min\{t > \tau_{r-1} / X_t > Z\} \quad r \geq 1$$

y la probabilidad de que el excedente ocurra en el ensayo t y se supere, en ese momento t , el valor Z por primera vez es:

$$w(t) = P(\tau_1 = t) = P\{X_1 \leq Z, X_2 \leq Z, \dots, X_{t-1} \leq Z, X_t > Z\} = p(1-p)^{t-1} \quad \text{para } t = 1, 2, \dots$$

con $p = 1 - F(Z)$.

Si el suceso ocurre en el ensayo t , el experimento debe haber fallado en $t-1$ ocasiones. Por tanto, el momento en que se supera Z por primera vez se distribuye geométricamente con parámetro p (Embrechts, et al. 1997) y cuya función generatriz de momentos es (Gumbel, 1958):

$$G_t(m) = p(e^{-m} - (1-p))^{-1}$$

A partir de dicha expresión, la media del primer momento en que se superará Z , también llamado periodo medio de retorno en Z , es,

$$E(\tau_1, Z) = \frac{1}{p} = \frac{1}{1-F(Z)} = T$$

e indica que si un suceso tiene una probabilidad de ocurrencia p , entonces tienen que producirse $\frac{1}{p}$ ensayos para que el suceso se produzca una vez. Esta ecuación puede invertirse para obtener el valor Z tal que el número medio de excedentes por encima de dicho valor en un periodo de amplitud T sea igual a 1, es decir, para cada unidad de tiempo (año, mes,...) existe una observación que cumple $E\left[\sum_{i \leq T} I(X_i \geq Z)\right] = 1$.

El valor Z que se producirá con un tiempo medio de espera T viene dado entonces por la siguiente expresión:

$$Z(T) = F^{-1}\left(1 - \frac{1}{T}\right)$$

La varianza de Z a partir de su función generatriz de momentos es,

$$V(\tau_1, Z) = (T-1)T = T^2 - T$$

Normalmente, para valores elevados de la media, la desviación estándar tiende a aproximarse a la media, $T - 0,5$. Por ello, desde un

punto de vista práctico, el periodo medio de retorno coincide con la desviación típica de la variable.

En cuanto a la mediana de la distribución, resulta:

$$\frac{\text{Ln}(2)}{-\text{Ln}\left(1 - \frac{1}{T}\right)} \approx \text{Ln}(2)T - \text{Ln}(\sqrt{2})$$

un 44% menor que la media y su valor incrementa con el periodo medio de retorno, T . Además, la probabilidad de que el evento se produzca antes del momento 0,693 ($\text{Ln}(2)$) del periodo de retorno, es la misma que la probabilidad de que ocurra después.

Finalmente, si la función de distribución se define como $W(t) = 1 - (1 - p)^t$ entonces, la probabilidad de que el suceso ocurra antes o en el momento exacto del periodo de retorno es:

$$W(t) = 1 - \left(1 - \frac{1}{T}\right)^T \sim 1 - \frac{1}{e} = 0,63212$$

El análisis del periodo de retorno resulta interesante cuando las observaciones se realizan en intervalos equidistantes, por ejemplo, cuando las mediciones son mensuales y el máximo mensual es el objetivo del estudio. En ese caso, el periodo de retorno es el número de meses que en media, habría que esperar antes de que un siniestro de una magnitud específica ocurra otra vez.

3. ANÁLISIS DE LA FRECUENCIA DE LOS VALORES EXTREMOS EN EL RAMO DE AUTOMÓVILES ESPAÑOL

En este apartado se realiza un estudio empírico sobre dos bases de datos de siniestralidad por responsabilidad civil en el ramo de automóviles, pertenecientes a dos entidades aseguradoras suficientemente representativas del mercado de seguros español.

El objetivo es aplicar los conceptos de la teoría del valor extremo, analizados en el epígrafe anterior, al estudio de la siniestralidad más elevada tanto para el reasegurador de casos extremos como para el propio asegurador directo.

La relevancia de esta investigación reside en la posibilidad de inferir sobre el acaecimiento de eventos extremos basándonos en la información histórica disponible. La incertidumbre ante la ocurrencia de grandes siniestros se reduce si se consigue obtener una distribución límite que modele la siniestralidad más elevada y que permita conocer su probabilidad y periodos de retorno. Por tanto, la inferencia que se consigue modelando los valores extremos es un *input* adicional al sistema de información total que facilita la toma de decisiones óptimas dentro de las entidades aseguradoras.

3.1. Descripción de las bases de datos e hipótesis de modelación

Las bases de datos utilizadas para llevar a cabo el presente artículo están formadas por dos carteras de pólizas de responsabilidad civil en automóviles, que recogen 10 años de experiencia siniestral para dos compañías de seguros, A y B, que operan en el mercado español. Sin embargo, hay que mencionar que los resultados estadísticos obtenidos en cada entidad no son comparables entre sí, por tres razones fundamentalmente: en primer lugar, los 10 años de historia no son los mismos en las dos entidades analizadas; en segundo lugar, cada compañía tiene una posición diferente en el mercado, una cuenta con una larga trayectoria empresarial lo que da estabilidad a la muestra analizada, mientras que la otra tiene una historia reciente con un fuerte crecimiento en el sector durante los años estudiados lo que hace despreciables, en algunos análisis, la consideración de los primeros años de la muestra. Finalmente, se ha introducido una distorsión distinta en cada compañía para mantener oculta su identidad lo que impide la comparación en valores absolutos y en determinados valores relativos.

Para llevar a cabo el estudio, se considera que el coste de la siniestralidad está formado por dos conceptos:

- El coste de los siniestros ocurridos dentro del periodo de observación y ya finalizados, cuyo importe está formado por todos los pagos realizados, netos de recobros efectivos.
- El coste de los siniestros ocurridos dentro del periodo de observación que todavía están pendientes de finalizar, cuyo importe viene dado por los pagos (netos de recobros) que ya se hayan realizado y/o las reservas de los pagos futuros estimados por la parte pendiente de liquidar.

Además para cada siniestro, se dispone de información acerca del coste total, expresado en unidades monetarias (u.m.), de la fecha de ocurrencia, de la fecha de declaración y de la fecha de su liquidación total.

El efecto de la inflación sobre la muestra, únicamente se ha tenido en cuenta en aquellos siniestros que ya han finalizado y su indexación se realiza desde la fecha de su liquidación total hasta la fecha de realización del análisis estadístico. En los siniestros pendientes de liquidación o pago, la constante actualización de las provisiones y la continuidad en la realización de los pagos, genera desconocimiento acerca de la inflación que afecta a cada siniestro lo que lleva a asumir que su coste está actualizado y expresado en u.m. presentes.

Finalmente, respecto a las hipótesis de modelación se supone que la distribución subyacente de los datos no ha variado sustancialmente a lo largo del tiempo, por lo que cualquier cambio real en la distribución de la siniestralidad extrema provocaría una actualización del estudio.

En las tablas 1 y 2 a continuación, se presentan, para cada una de las compañías de seguros analizadas y para cada uno de los 10 años de experiencia, el número de siniestros registrados, el coste total de siniestralidad y el coste medio individual, ambos en u.m. nominales:

TABLA 1: Número de siniestros, Coste total y coste medio de la cartera de la Entidad A

Año	Nº siniestros	Coste total (u.m. nominales)	Variación interanual	Coste medio (u.m. nominales)	Variación interanual
1	3.103	24.702,2		7,96	
2	3.530	30.702,5	24,290%	8,70	9,247%
3	3.948	34.616,6	12,749%	8,77	0,799%
4	4.336	40.367,7	16,614%	9,31	6,187%
5	4.845	40.389,7	0,054%	8,34	-10,447%
6	5.046	42.707,5	5,739%	8,46	1,511%
7	5.618	48.359,7	13,235%	8,61	1,717%
8	5.894	52.649,0	8,870%	8,93	3,773%
9	5.959	57.097,8	8,450%	9,58	7,262%
10	6.025	61.580,8	7,851%	10,22	6,673%

TABLA 2: Número de siniestros, Coste total y Coste medio de la cartera de la Entidad B

Año	Nº siniestros	Coste total (u.m. nominales)	Variación interanual	Coste medio (u.m. nominales)	Variación interanual
1	2.099	654.865		311,9	
2	5.697	1.827.451	179%	320,8	2,84%
3	9.400	2.847.242	56%	302,9	-5,58%
4	11.569	3.754.142	32%	324,5	7,13%
5	13.717	4.443.483	18%	323,9	-0,17%
6	16.397	5.858.425	32%	357,3	10,29%
7	20.923	8.313.259	42%	397,3	11,21%
8	28.039	11.775.184	42%	420,0	5,69%
9	34.415	14.086.821	20%	409,3	-2,53%
10	39.501	17.743.476	26%	449,2	9,74%

Resulta evidente al observar los datos, que no existe, en ninguna de las dos entidades, un comportamiento estable en la evolución de la cuantía media de la siniestralidad. Ello se debe a diversos factores entre los que cabe mencionar la descomposición del siniestro en distintas garantías, inflaciones dispares en los distintos conceptos de gasto, diferentes tiempos de liquidación, siniestros punta, etc.

Analizando la siniestralidad respecto al número total de pólizas en las carteras, recogida en las tablas 3 y 4 siguientes, se observa que, debido a la consolidación de la entidad A en el mercado asegurador español, la frecuencia de la siniestralidad sobre las pólizas está acotada entre el 14% y el 16% y la media ponderada por el número de siniestros es del 14,98% no existiendo fuertes desviación respecto a dicha media. Sin embargo, la historia menos consolidada de la entidad B provoca que la frecuencia de siniestralidad sobre pólizas sea más elevada (la media ponderada de los últimos cuatro años es 45,13%), si bien se observa una propensión decreciente en esta ratio, tendente a estabilizarse a medida que el volumen de pólizas de la entidad es mayor.

TABLA 3: Siniestralidad sobre total de pólizas en cartera Entidad A

Año	Nº siniestros	Variación interanual	Nº de pólizas	Variación interanual	Relación Siniestros/Pólizas
1	3.103				
2	3.530	13,770%	26.029		13,562%
3	3.948	11,855%	28.861	10,880%	13,681%
4	4.336	9,819%	30.420	5,404%	14,254%
5	4.845	11,726%	31.724	4,287%	15,271%
6	5.046	4,165%	32.919	3,766%	15,330%
7	5.618	11,323%	34.887	5,978%	16,103%
8	5.894	4,912%	36.962	5,946%	15,945%
9	5.959	1,107%	39.482	6,820%	15,093%
10	6.025	1,104%	41.476	5,049%	14,526%

TABLA 4: Siniestralidad sobre total de pólizas en cartera Entidad B

Año	Nº siniestros	Variación interanual	Nº de pólizas	Variación interanual	Relación Siniestros/Pólizas
1	2.099		0		
2	5.697	171,34%	7.913		72,00%
3	9.400	65,01%	14.207	79,54%	66,16%
4	11.569	23,08%	18.454	29,90%	62,69%
5	13.717	18,56%	22.300	20,84%	61,51%
6	16.397	19,54%	28.564	28,09%	57,40%
7	20.923	27,60%	41.027	43,63%	51,00%
8	28.039	34,01%	59.467	44,95%	47,15%
9	34.415	22,74%	78.297	31,67%	43,95%
10	39.501	14,78%	93.444	19,35%	42,27%

A partir de esta información bruta se deduce que un estudio agrupado de siniestralidad no permite extraer conclusiones particularizadas para cada grupo de riesgo o para cada tramo de coste. Un análisis pormenorizado de la siniestralidad requiere segmentación por tramos de coste, que permita identificar, aislar y modelar el comportamiento de los siniestros más graves.

3.2. Resultados de la aplicación práctica

Como ya se ha comentado anteriormente, a partir del método de los excedentes de Gumbel, es posible obtener una serie de resultados analíticos bastante sencillos, basados en la ordenación de las variables aleatorias, que proporcionan información sobre los datos analizados de utilidad para la gestión del riesgo en el sector asegurador.

Para llevar a cabo el estudio, se han observado 10 años de experiencia de siniestralidad en las dos compañías de seguros, A y B, considerando los máximos valores mensuales, lo que da lugar a una muestra de 120 valores extremos mensuales en cada compañía ($n = 120$). A partir de estas series históricas, se pueden obtener las probabilidades de frecuencia de ocurrencia de valores extremos para los próximos 3 años ($N = 12$, $N = 24$ y $N = 36$).

Hay que mencionar en este punto que al tratarse de un método independiente de la cuantía de los siniestros máximos, las frecuencias que se estimen serán idénticas para ambas entidades. La diferencia se encuentra en el nivel de los máximos mensuales históricos registrados en cada entidad.

En los 10 años estudiados, los cinco siniestros máximos registrados para ambas entidades, aparecen indicados en la tabla 5 a continuación:

TABLA 5: Coste del siniestro (u.m.)		
Siniestro de orden k	ENTIDAD A	ENTIDAD B
1	388,598	71.851
2	282,769	71.541
3	195,971	60.152
4	171,105	54.792
5	169,675	54.420

Aplicando el método de Gumbel y a partir de la función de densidad $w(h; n, k, N)$, la información obtenida es la siguiente:

A. Frecuencia esperada de siniestralidad para el próximo año (N=12)

En el siguiente gráfico (Gráfico 1), realizado a partir de la información disponible en la tabla 6, se muestran las frecuencias esperadas para los próximos 12 máximos mensuales. En el eje de abscisas se representa el número, h , de siniestros esperados y en el eje de ordenadas la frecuencia esperada de que ese suceso ocurra ($w(h)$):

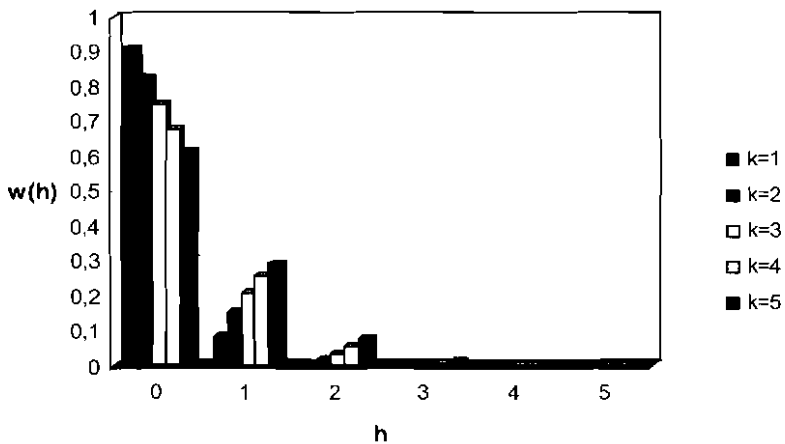


Gráfico 1: frecuencia esperada de valores extremos para el próximo año

TABLA 6: Probabilidades Individuales y Acumuladas de que el siniestro de orden k , sea superado h veces en los próximos 12 máximos mensuales

h	k=1		k=2		k=3		k=4		k=5	
	w(h)	W(h)	w(h)	W(h)	w(h)	W(h)	w(h)	W(h)	w(h)	W(h)
0	0.90909091	0.909090909	0,825815406	0,825815406	0,749586292	0,749586292	0,679857334	0,679857334	0,616120709	0,616120709
1	0,0832755	0,92366412	0,152458229	0,978273635	0,209186872	0,958773164	0,2549465	0,934803835	0,29108065	0,907201359
2	0,00704639	0,999412801	0,019500471	0,997774106	0,035953994	0,994727157	0,055204951	0,990008785	0,076235408	0,983436767
3	0,00054623	0,999959033	0,002031299	0,999805405	0,004718372	0,999445529	0,008762691	0,998771476	0,01423061	0,997667377
4	3,8407E-05	0,99999744	0,000179938	0,999985343	0,00050554	0,999951069	0,001104099	0,999875575	0,002065734	0,999733111
5	2,4193E-06	0,999999859	1,37096E-05	0,999999052	4,52964E-05	0,999996365	0,000113972	0,999989547	0,000241842	0,999974953
6	1,3441E-07	0,999999993	8,95691E-07	0,999999948	3,4094E-06	0,999999775	9,72928E-06	0,999999276	2,3127E-05	0,99999808
7	6,4516E-09	1	4,95313E-08	0,999999998	2,1383E-07	0,999999989	6,83556E-07	0,999999959	1,8021E-06	0,999999882
8	2,6014E-10	1	2,26515E-09	1	1,09544E-08	1	3,88384E-08	0,999999998	1,12631E-07	0,999999994
9	8,4599E-12	1	8,25192E-11	1	4,42603E-10	1	1,72615E-09	1	5,46856E-09	1
10	2,0803E-13	1	2,25052E-12	1	1,32781E-11	1	5,65713E-11	1	1,94644E-10	1
11	3,4385E-15	1	4,09186E-14	1	2,63736E-13	1	1,22034E-12	1	4,53715E-12	1
12	2,8654E-17	1	3,72508E-16	1	2,60756E-15	1	1,30378E-14	1	5,21511E-14	1

Analizando los valores recogidos en la tabla 6, se estima que en las próximas 12 observaciones máximas mensuales, el máximo valor mensual histórico ($k = 1$) no será superado en el año siguiente ($h = 0$) con una probabilidad del 90,90%. Sin embargo, a medida que aumenta el valor de k (es decir, a medida que se reduce el importe del valor extremo considerado), la probabilidad de que dicho máximo no sea superado se va reduciendo.

La probabilidad de que el mayor máximo histórico ($k = 1$) sea superado una vez ($h = 1$) es del 8,32% y a medida que vamos descendiendo posiciones en la jerarquía siniestral, la probabilidad va incrementando, alcanzando su máximo valor para el quinto valor máximo con un 29,10%.

Considerando los 5 valores máximos mensuales históricos ($k = 1,2,3,4,5$), las mayores frecuencia estimadas se producen para $h = 0$, y por tanto, en conjunto, resulta más probable que ninguna de las 12 observaciones máximas futuras supere en ningún caso los 5 valores extremos históricos.

B. Frecuencia esperada de siniestralidad para los próximos dos años (N=24)

De igual forma que en el apartado anterior, estimamos la probabilidad de que los máximos mensuales históricos sean superados en alguna de las 24 siguientes máximas observaciones. Los datos obtenidos de la estimación se presentan en el gráfico 2 y la tabla 7, a continuación:

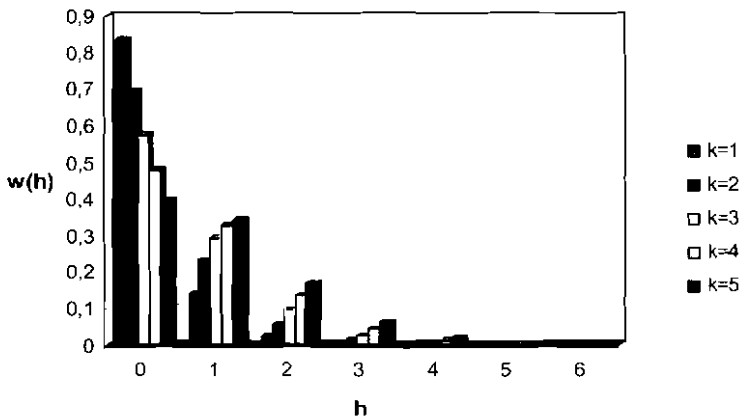


Gráfico 2: frecuencia esperada de valores extremos para los próximos

TABLA 7: Probabilidades Individuales y Acumuladas de que el siniestro de orden k , sea superado h veces en los próximos 24 máximos mensuales

h	k=1		k=2		k=3		k=4		k=5	
	w(h)	W(h)	w(h)	W(h)	w(h)	W(h)	w(h)	W(h)	w(h)	W(h)
0	0.833333333	0.833333333	0.693473193	0.693473193	0.576266457	0.576266457	0.478178549	0.478178549	0.396205083	0.396205083
1	0.13986014	0.973193473	0.234413474	0.927886667	0.294263722	0.870530179	0.327893862	0.806072411	0.342047554	0.738252638
2	0.0226534	0.995846876	0.057356488	0.985243156	0.096686652	0.967216831	0.135639547	0.941711959	0.171023777	0.909276415
3	0.00353457	0.99991636	0.01201755	0.997260706	0.025504872	0.992721703	0.043247392	0.98495935	0.064081902	0.973358317
4	0.00053019	0.99991636	0.002269501	0.999530207	0.005821764	0.998343467	0.011601034	0.996560384	0.019789999	0.999148316
5	7.6286E-05	0.999987921	0.000394696	0.999924903	0.00118985	0.999733317	0.002729655	0.999290039	0.005277333	0.998425649
6	1.0503E-05	0.999998425	6.3862E-05	0.999988765	0.000221639	0.999954955	0.000576261	0.9998663	0.001247131	0.99967278
7	1.38E-06	0.999999805	9.6598E-06	0.999998425	3.79952E-05	0.999992951	0.000110583	0.999976883	0.000265233	0.999938013
8	1.725E-07	0.999999977	1.36847E-06	0.999999793	6.02536E-06	0.999998976	1.94352E-05	0.999996318	5.12382E-05	0.999989251
9	2.0444E-08	0.999999997	1.81555E-07	0.999999975	8.85934E-07	0.999999862	3.14104E-06	0.999999459	9.03948E-06	0.999999829
10	2.2885E-09	1	2.25237E-08	0.999999997	1.20809E-07	0.999999983	4.67559E-07	0.999999927	1.46022E-06	0.999999751
11	2.409E-10	1	2.60605E-09	1	1.52583E-08	0.999999998	6.4085E-08	0.999999991	2.16101E-07	0.999999967
12	2.3725E-11	1	2.80167E-10	1	1.78014E-09	1	8.07273E-09	0.999999999	2.92636E-08	0.999999996
13	2.1732E-12	1	2.7851E-11	1	1.9107E-10	1	9.31469E-10	1	3.61586E-09	1
14	1.8389E-13	1	2.54452E-12	1	1.87659E-11	1	9.79666E-11	1	4.05862E-10	1
15	1.4255E-14	1	2.12044E-13	1	1.67464E-12	1	9.33015E-12	1	4.11273E-11	1
16	1.0023E-15	1	1.59659E-14	1	1.3457E-13	1	7.97728E-13	1	3.73131E-12	1
17	6.3137E-17	1	1.07334E-15	1	9.62568E-15	1	6.05486E-14	1	2.9979E-13	1
18	3.5076E-18	1	6.34461E-17	1	6.03761E-16	1	4.02017E-15	1	2.10235E-14	1
19	1.6837E-19	1	3.23155E-18	1	3.25519E-17	1	2.28931E-16	1	1.26196E-15	1
20	6.789E-21	1	1.37932E-19	1	1.46751E-18	1	1.08789E-17	1	6.30979E-17	1
21	2.2078E-22	1	4.73771E-21	1	5.31328E-20	1	4.14436E-19	1	2.52493E-18	1
22	5.429E-24	1	1.22803E-22	1	1.44908E-21	1	1.18727E-20	1	7.58645E-20	1
23	8.9736E-26	1	2.13571E-24	1	2.6472E-23	1	2.2748E-22	1	1.52236E-21	1
24	7.478E-28	1	1.86949E-26	1	2.43034E-25	1	2.18731E-24	1	1.53111E-23	1

Analizando los valores obtenidos, la probabilidad de no superar el mayor valor de los máximos mensuales históricos ($h = 0$) es del 83,3%, por tanto la probabilidad de que dicho máximo sea superado alguna vez, por alguno de los próximos 24 máximos mensuales, es del 16,7%.

Para el valor extremo mensual histórico más elevado ($k = 1$), en $h = 2$ se acumula una probabilidad del 99,58% lo que implica que existe una probabilidad del 0,42% de que el máximo histórico sea superado 3 o más veces en los próximos 2 años.

Las probabilidades individuales más altas se concentran en $h = 0$ para los 5 valores de k , lo que indica que es más probable no superar ninguno de los 5 máximos mensuales históricos que superarlos en alguna ocasión, en los 24 máximos mensuales próximos.

Finalmente, destacar que para el menor máximo mensual histórico, $k = 5$, la probabilidad de que no se supere su valor en ninguno de los máximos mensuales de los próximos dos años y la probabilidad de que se supere una única vez son muy similares ($w(h = 0) = 39,6\%$ y $w(h = 1) = 34,2\%$).

C. Frecuencia esperada de siniestralidad para los próximos tres (N=36)

Como se deriva de los resultados de la estimación (ver gráfico 3 y tabla 8), a medida que aumenta el número de observaciones futuras (N), disminuyen las probabilidades de que los máximos históricos no sean superados y aumentan las probabilidades de superar los valores mensuales extremos históricos.

La probabilidad de que no se superen los tres siniestros históricos mayores en los próximos 36 máximos mensuales, sigue siendo mucho mayor que para el resto de situaciones, pero no ocurre así para el cuarto mayor valor, $k = 4$, que presenta una probabilidad de ser superado en una ocasión (32,8%) muy similar a la de no ser alcanzado en ningún caso (34,6%). En el quinto mayor siniestro mensual histórico (169,675 u. m. en la entidad A y 54.420 u. m. en la entidad

B), la probabilidad de no ser superado en los próximos tres años se reduce hasta el 24,6% lo que indica que existe una probabilidad acumulada del 83,6% de que sea superado una o más veces en el periodo considerado.

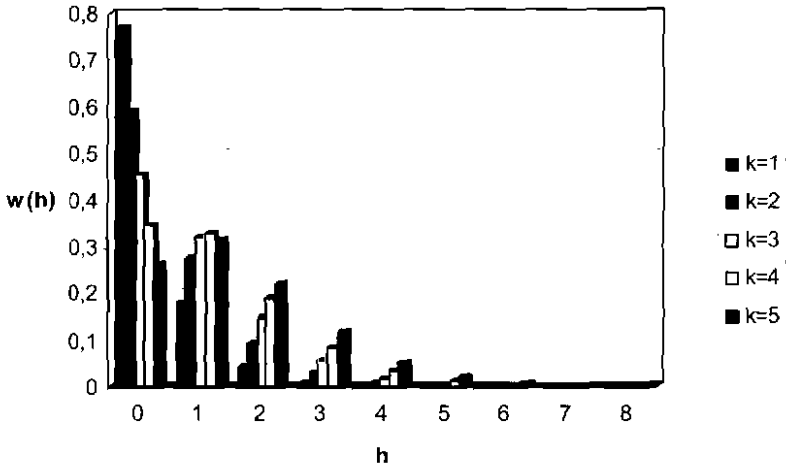


Gráfico 3: frecuencia esperada de valores extremos para los próximos tres años

TABLA 8: Probabilidades Individuales y Acumuladas de que el siniestro de orden k , sea superado h veces en los próximos 36 máximos mensuales

N=36	k=1		k=2		k=3		k=4		k=5	
	h	w(h)	Probabil. Acumulada	w(h)	Probabil. Acumulada	w(h)	Probabil. Acumulada	w(h)	Probabil. Acumulada	w(h)
0	0,76923077	0,769230769	0,59057072	0,59057072	0,452515227	0,452515227	0,346041056	0,346041056	0,264083964	0,264083964
1	0,17866005	0,947890819	0,276110986	0,866681705	0,319422513	0,77193774	0,327828369	0,673869424	0,314802076	0,578886039
2	0,04060456	0,988495376	0,094743966	0,961425671	0,147102473	0,919040213	0,18996677	0,863836194	0,220361453	0,799247492
3	0,00902323	0,99751861	0,028256972	0,989682643	0,055204019	0,974244231	0,086118269	0,949954463	0,117328917	0,916576409
4	0,00195899	0,999477602	0,007719206	0,997401849	0,018217326	0,992461557	0,033378054	0,983332517	0,052322355	0,968898764
5	0,00041515	0,999892753	0,001976117	0,999377966	0,005477424	0,997938982	0,011547002	0,99487952	0,020501821	0,989400585
6	8,5798E-05	0,999978551	0,000479661	0,999857628	0,001529731	0,999468713	0,003652623	0,998532143	0,007255211	0,996655795
7	1,7275E-05	0,999995825	0,000111118	0,999968746	0,000401387	0,9998701	0,001072199	0,999604341	0,002358837	0,999014632
8	3,3849E-06	0,99999921	2,46615E-05	0,999993407	9,96595E-05	0,999969759	0,000294855	0,999899196	0,000712565	0,999727197
9	6,4475E-07	0,999999855	5,25511E-06	0,999998663	2,35212E-05	0,999993281	7,64438E-05	0,99997564	0,000201534	0,999928731
10	1,1923E-07	0,999999974	1,07639E-06	0,999999739	5,29226E-06	0,999998573	1,87635E-05	0,999994403	5,36477E-05	0,999982379
11	2,138E-08	0,999999996	2,12017E-07	0,999999951	1,13718E-06	0,99999971	4,37254E-06	0,999998776	1,34897E-05	0,999995869
12	3,7118E-09	0,999999999	4,01547E-08	0,999999991	2,33576E-07	0,999999944	9,69091E-07	0,999999745	3,21184E-06	0,999999908
13	6,2296E-10	1	7,30876E-09	0,999999998	4,58741E-08	0,99999999	2,04468E-07	0,999999949	7,25198E-07	0,999999806
14	1,009E-10	1	1,27737E-09	1	8,6131E-09	0,999999998	4,10826E-08	0,99999999	1,554E-07	0,999999961
15	1,5743E-11	1	2,14111E-10	1	1,54499E-09	1	7,85929E-09	0,999999998	3,16092E-08	0,999999993
16	2,3615E-12	1	3,43694E-11	1	2,64495E-10	1	1,43059E-09	1	6,10105E-09	0,999999999
17	3,3979E-13	1	5,27408E-12	1	4,31551E-11	1	2,47507E-10	1	1,11653E-09	1
18	4,6782E-14	1	7,72078E-13	1	6,69891E-12	1	4,06401E-11	1	1,93495E-10	1
19	6,1466E-15	1	1,07565E-13	1	9,87208E-13	1	6,32108E-12	1	3,17004E-11	1
20	7,6832E-16	1	1,42225E-14	1	1,37767E-13	1	9,29151E-13	1	4,89916E-12	1

TABLA 8: Probabilidades Individuales y Acumuladas de que el siniestro de orden k , sea superado h veces en los próximos 36 máximos mensuales

N=36	k=1		k=2		k=3		k=4		k=5	
	h	w(h) Probabil. Acumulada	w(h) Probabil. Acumulada	w(h) Probabil. Acumulada	w(h) Probabil. Acumulada	w(h) Probabil. Acumulada	w(h) Probabil. Acumulada	w(h) Probabil. Acumulada	w(h) Probabil. Acumulada	
21	9,106E-17	1	1,77907E-15	1	1,81519E-14	1	1,28714E-13	1	7,12346E-13	1
22	1,0193E-17	1	2,09768E-16	1	2,25024E-15	1	1,67479E-14	1	9,71381E-14	1
23	1,073E-18	1	2,32154E-17	1	2,61395E-16	1	2,03888E-15	1	1,23755E-14	1
24	1,0567E-19	1	2,39981E-18	1	2,83178E-17	1	2,31152E-16	1	1,46637E-15	1
25	9,6799E-21	1	2,30382E-19	1	2,84495E-18	1	2,4271E-17	1	1,60724E-16	1
26	8,1907E-22	1	2,04005E-20	1	2,63295E-19	1	2,34477E-18	1	1,61901E-17	1
27	6,3494E-23	1	1,65282E-21	1	2,22675E-20	1	2,0677E-19	1	1,48709E-18	1
28	4,4644E-24	1	1,21312E-22	1	1,70415E-21	1	1,64825E-20	1	1,23353E-19	1
29	2,8122E-25	1	7,96797E-24	1	1,16587E-22	1	1,17339E-21	1	9,12958E-21	1
30	1,5623E-26	1	4,6108E-25	1	7,02031E-24	1	7,34564E-23	1	5,93672E-22	1
31	7,4993E-28	1	2,303E-26	1	3,64548E-25	1	3,96222E-24	1	3,32368E-23	1
32	3,0239E-29	1	9,65434E-28	1	1,58743E-26	1	1,79078E-25	1	1,55797E-24	1
33	9,8338E-31	1	3,26128E-29	1	5,56574E-28	1	6,51191E-27	1	5,87166E-26	1
34	2,4181E-32	1	8,32363E-31	1	1,47328E-29	1	1,78651E-28	1	1,66842E-27	1
35	3,9969E-34	1	1,42691E-32	1	2,6176E-31	1	3,28752E-30	1	3,17794E-29	1
36	3,3308E-36	1	1,23239E-34	1	2,34154E-33	1	3,044E-32	1	3,044E-31	1

4. CONCLUSIONES

La preocupación internacional por la solvencia de las entidades aseguradoras, derivada de los continuos avances en tecnología y desarrollo económico, exige el tratamiento actuarial de los grandes riesgos asegurados por ser aquellos que, a pesar de su baja frecuencia, pueden ocasionar consecuencias devastadoras.

Hasta ahora, la búsqueda del equilibrio financiero-actuarial se ha basado en conseguir prestaciones cobradas a priori adecuadas a las contraprestaciones que se han de afrontar a posteriori. Pero actualmente, tanto el seguro y como el reaseguro, encargados de proporcionar cobertura ante la ocurrencia de sucesos económicamente desfavorables, aceptan el riesgo de situaciones, como la ocurrencia de catástrofes, que pueden dañar su estabilidad. Esto conlleva una preocupación por la solvencia ante quien deposita su confianza en ellos y la subsiguiente necesidad de coberturas que garanticen la supervivencia de los mismos. Es aquí donde la Teoría del Valor Extremo comienza a jugar un importante rol, ya que, a partir de los modelos que cuantifican todos los aspectos relativos a los valores extremos se puede medir su frecuencia y consecuencias de un modo estadísticamente óptimo.

La Teoría del Valor Extremo complementa a la teoría clásica del Riesgo basada en la ley de los grandes números y el teorema central del límite y es medio adecuado para adoptar decisiones financieras óptimas relacionadas con los siniestros esperados más graves. Esta teoría permite obtener resultados y conclusiones válidos para la toma de decisiones, de manera que se garantice la solvencia y la capacidad de afrontar situaciones extremas futuras, bien a través de capitales propios, bien a través del Reaseguro.

Las hipótesis en las que se basa la teoría del riesgo desprecian la información acerca de los sucesos que presentan una probabilidad de acaecimiento extremadamente pequeña con consecuencias graves en términos monetarios, puesto que contaminan el comportamiento central e incrementan la dispersión. El análisis de una cartera de siniestros no se debe limitar a realizar una transformación logarítmica

que permita estandarizar el estudio. Las distribuciones de cola gruesa muestran mayor masa de probabilidad en la cola que distribuciones como la normal. Esto significa que situaciones extremas ocurren con mayor frecuencia que bajo hipótesis de normalidad y que requieren de un estudio propio que tome las 'puntas' como base del análisis.

En este artículo se pone de manifiesto la utilidad de la Teoría del Valores Extremo en la gestión de riesgos de las entidades aseguradoras, a través de un estudio empírico de la cartera de siniestros de responsabilidad civil en automóviles de dos compañías representativas del sector asegurador español. La aplicación de la Teoría del Valor Extremo en este caso, se ha centrado en el análisis de la frecuencia de ocurrencia de dichos extremos con el fin de determinar el número de valores, de entre las futuras observaciones, que excederán una determinada posición en la jerarquía de las series de datos disponibles para de esa forma establecer provisiones adecuadas y poder afrontarlos con garantías de solvencia.

Los principales resultados obtenidos con el estudio se presentan a continuación:

Tomando como referencia el máximo siniestro histórico $k = 1$, en las series de datos disponibles para las dos entidades, se observa que cuanto menor es el número de máximos futuros mensuales sobre los que se realiza la inferencia (es decir, cuanto menor es N), menor es la probabilidad de alcanzar la cuantía máxima histórica. Por otro lado, a medida que este número de máximos futuros (h) incrementa, dicha probabilidad tiende a disminuir:

Resumen			
	N=12	N=24	N=36
h=0	90,909%	83,33%	76,92%
h=1	8,328%	13,99%	17,87%
h=2	0,705%	2,26%	4,06%

La aplicación del método de los excedentes de Gumbel también permite calcular otro tipo de frecuencias como por ejemplo aquellas relacionadas con la mediana de las observaciones históricas. En este caso, la probabilidad de que la mediana de las 120 observaciones

pasadas (60), no se supere en ninguna de las N observaciones futuras, $w(h = 0; n = 120, k = 60, N)$, con $N = 12, N = 24, N = 36$:

$$w(0;120,60,12) \approx 0,44\%$$

$$w(0;120,60,24) \approx 0\%$$

$$w(0;120,60,36) \approx 0\%$$

cuyo resultado es prácticamente nulo en los tres casos analizados lo que indica que, casi con toda seguridad, en los próximos 12, 24 y 36 meses, el máximo de siniestralidad mensual superará al menos en una ocasión a la mediana de los 120 extremos históricos mensuales.

La probabilidad de que la mediana de las 120 observaciones pasadas se supere en todas las observaciones futuras, $w(h = N; n = 120, k = 60, N)$, con $N = 12, N = 24, N = 36$ es,

$$w(12;120,60,12) \approx 0,037\%$$

$$w(12;120,60,24) \approx 0\%$$

$$w(12;120,60,36) \approx 0\%$$

Lógicamente, resulta imposible que todas las observaciones futuras superen a la mediana de los máximos mensuales históricos.

Finalmente, la probabilidad de que la mediana de las observaciones pasadas sea excedida en la mitad de las observaciones futuras,

$w(h = \frac{N}{2}; n = 120, k = 60, N)$, resulta:

$$w(6;120,60,12) = 21,5047\%$$

$$w(12;120,60,24) = 14,71\%$$

$$w(18;120,60,36) = 11,577\%$$

Respecto al número medio de valores de las N observaciones futuras que se espera que excedan el k -ésimo valor más alto de los datos históricos, $\bar{x}_k = k \frac{N}{n+1}$, los resultados obtenidos se muestran en la siguiente tabla:

TABLA 9: Número medio de siniestros esperado por encima de los máximos históricos en los próximos años

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$N=12$	0,099	0,198	0,298	0,397	0,496
$N=24$	0,198	0,397	0,595	0,793	0,992
$N=36$	0,298	0,595	0,893	1,190	1,488

Esta tabla indica, por ejemplo para $N = 12$, que si el número medio de siniestros que se espera que superen el máximo histórico en el próximo año es 0,0992, entonces, el tiempo medio esperado para que el máximo histórico sea superado una vez es de 10 años ($0,0992 \times 10$ años = 0,992 \approx 1 siniestro) mientras que el tiempo de espera para que el quinto máximo mensual histórico sea superado una vez es de sólo 2 años ($0,459 \times 2$ años = 0,992 \approx 1 siniestro).

La dispersión o desviación típica respecto a la media viene dada por la siguiente expresión:

$$S_k = \sqrt{\frac{kN(n-k+1)(n+N+1)}{(n+2)(n+1)^2}}$$

La tabla 10 muestra las desviaciones respecto a la media de los próximos 12, 24 y 36 máximos mensuales:

TABLA 10: Desviación respecto a la media de los próximos máximos mensuales

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$N=12$	0,327	0,461	0,562	0,647	0,720
$N=24$	0,484	0,681	0,830	0,955	1,063
$N=36$	0,616	0,868	1,058	1,217	1,355

Asumiendo que $\bar{x}_k + S_k$ es una aproximación prudente al número máximo de siniestros probable que se espera que superen el k -ésimo valor máximo histórico, las estimaciones para dicho número se muestran en la siguiente tabla,

TABLA 11: Número máximo de siniestros probable

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$N=12$	0,426	0,659	0,86	1,044	1,216
$N=24$	0,682	1,078	1,425	1,748	2,055
$N=36$	0,914	1,463	1,951	2,407	2,843

Para relativizar la dispersión respecto a la media y conocer cual es la media con mayor desviación, se emplea el coeficiente de variación definido como,

$$CV = \frac{S_k}{\bar{x}_k}$$

y cuyos resultados (mostrados en la tabla 12) confirman que existe una mayor dispersión en la estimación de la media cuanto menores son k y N . Así por ejemplo, mientras que para $k = 5$ y $N = 36$ la dispersión respecto a la media es del 91%, para $k = 1$ y $N = 12$ la dispersión es del 329,6%.

TABLA 12: Coeficiente de variación

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$N=12$	329,6%	232,5%	188,9%	163,1%	145,2%
$N=24$	244,0%	171,7%	139,5%	120,4%	107,2%
$N=36$	207,0%	145,9%	118,5%	102,3%	91,1%

Esto supone que existe un riesgo mayor en la estimación del número medio de siniestros que se espera que superen el valor mayor de los registrados en los últimos 120 meses de historia de la entidad aseguradora.

Otra importante conclusión que se puede obtener a través de la distribución de Gumbel, es la estimación del tamaño de la muestra necesario para obtener una determinada probabilidad de frecuencia.

Consideremos el siguiente ejemplo: se supone que el tamaño de la nueva muestra de datos (N) es desconocido aunque se sabe que es elevado. El objetivo es determinar un número suficiente de observaciones de referencia, n , para tener un 95% de probabilidad de que como máximo el 10% de los extremos futuros excedan el mayor valor de los datos de referencia ($k=1$). Esto implica que, la función de distribución es $F = 0,95$ y $\frac{h}{N} = q = 0,1$. Entonces, a partir de la aproximación de Gumbel,

$$F(0,1;n,1) = 1 - (1 - 0,1)^n = 0,95$$

despejamos n y obtenemos que el número de observaciones de referencia necesarias es de 28. Por tanto, si tomamos por ejemplo, los mayores siniestros de los últimos 28 años, se puede decir, con una probabilidad del 95%, que como máximo el 10% de la siniestralidad futura superará al mayor de esos 28 datos anuales.

BIBLIOGRAFÍA

1. Embrechts, P., C. Klüppelberg y T. Mikosch (1997), *Modelling extremal events for insurance and finance. Applications of Mathematics*. Springer-Verlag Berlin Heidelberg.
2. Embrechts, P., S. L. Resnick y G. Samorodnitsky (1999), Extreme value theory as a risk management tool. *North American Actuarial Journal*, volumen 3, nº 2, pp. 30-41.

3. Gumbel, E. J. (1958), *Statistics of extremes*. Columbia University Press, New York.
4. International Actuarial Association (2004), *A global framework for insurance solvency assessment*. Research Report of the Insurance solvency assessment working party. International Actuarial Association.
5. Kinnison, R. P. (1985), *Applied extreme value statistics*. Macmillan Publishing Company. New York. Battelle Press, Columbus.
6. Pérez-Fructuoso, M. J. y A. García Pérez (2004), Aplicación de la teoría del valor extremo al ajuste y modelación de catástrofes. *Gerencia de Riesgos y Seguros*. Número 88, 4.º Trimestre 2004, pp. 19-32.