Aplicaciones económicas de la estadística de valores extremos

JOSÉ MARÍA SARABIA ALEGRÍA
Departamento de Economía
Escuela Universitaria de Estudios Empresariales
Universidad de Cantabria

INDICE

1. INTRODUCCION

En la actual dinámica de cambio, que afecta a la totalidad de las estructuras organizativas, y dentro de ellas de manera particular las de tipo económico, uno de los problemas que con mayor intensidad afecta a las mismas es el relativo a acotar la incertidumbre que acompaña, no sólo a la complejidad interna, sino también a la correspondiente al entorno (1). En efecto, el planteamiento estratégico dentro de las organizaciones ha adquirido relevancia importante; más aún, se ha situado en primer lugar, en cuanto que supone el mantenimiento de la posición competitiva empresarial que, en última instancia, no es sino mantener una supervivencia, mediante el establecimiento de equilibrios dinámicos intertemporales y que permitirán llevar a cabo la cumplimentación de sus, en definitiva, misiones sociales (8).

Dentro de este contexto es bien significativo abordar la incertidumbre de sucesos que pueden impactar muy fuertemente en tal supervivencia; tal puede ser el caso de siniestros derivados de accidentes o catástrofes que inciden o afectan a las compañías aseguradoras en particular.

Tales siniestros, o en general tales sucesos, reciben el nombre de sucesos «extremos».

A lo largo de todo el trabajo vamos a suponer observando un determinado fenómeno económico, dado por *n* observaciones cuantitativas $X_1, X_2, ..., X_n$ bajo hipótesis de equidistribución e independencia.

Consideremos ahora la muestra ordenada de menor a mayor:

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$$

y estamos interesados en obtener información en términos probabilísticos de esos valores extremos, máximo y mínimo, suponiendo, dada una hipótesis acerca de la distribución de probabilidad generadora de los datos, que representaremos por F (x).

Según lo visto anteriormente, podemos preguntarnos sobre el comportamiento de ese máximo suponiendo un tamaño grande de muestra, es decir, su distribución asintótica, y si tal distribución asintótica posee un comportamiento estable, importante a la hora de hacer predicciones.

2. CONCEPTO DE EXCEDENCIAS. PROBABILIDADES ASOCIADAS. PERIODOS DE RETORNO

Para empezar nuestro estudio sobre probabilidades de valores extremos, vamos a dar el concepto de excedencia.

Consideremos una variable aleatoria *x* (ventas de una empresa, cuantía de un determinado tipo de accidentes, etc.) y *n* realizaciones independientes de *x*.

Diremos que el suceso [w / X (w) = x] es una excedencia del nivel *v* si $x \geq v$.

El primer problema inmediato que se nos plantea es determinar la distribución del número de excedencias del valor *x* en los próximos *n* ensayos.

Si $E_n(x)$ = número de excedencias del valor *x* en *n* ensayos, es claro que si *F* es la c.d.f. de *X*, se tiene que $E_n(x) \rightarrow B[n, 1 - F(x)]$, de donde:

$$p[E_n(x) = x] = \begin{bmatrix} n \\ k \end{bmatrix} [1 - F(x)]^k F(x)^{n-k}$$

que nos da la distribución *k* excedencias del valor *x* en los próximos *n* ensayos.

El valor máximo de $E_n(x)$ verifica que

$$[1 - F(x)](n + 1) - 1 \leq \text{Máx } E_n(x) \leq [1 - F(x)](n + 1)$$

Por otra parte, para valores grandes de *n*, podemos usar la habitual aproximación de Poisson:

$$E_n(x) \rightarrow P(\lambda) \text{ con } \lambda = n[1 - F(x)]$$

siempre que $n \geq 50$ y $F(x) \geq 1 - 1/10n$.

La aproximación a la distribución binomial será tanto más óptima cuanto con mayor holgura los parámetros *n*, *x* verifiquen las inecuaciones propuestas.

Ejemplo 1: Un negocio de automóviles posee una venta anual media de $\mu = 50$ vehículos. Se ha observado que la cifra de ventas anual no depende nunca del número de automóviles vendidos hasta el momento. Con vistas a una posible ampliación del negocio, determinar la probabilidad de *k* excedencias del valor 60 vehículos vendidos en los próximos cinco años.

Debido a la propiedad de pérdida de memoria, podemos representar las ventas anuales por:

$$F(x) = 1 - \exp(-x / 50) \text{ si } x \geq 0$$

con una probabilidad de no excedencia de $F(60) = 1 - \exp(-1.2)$.

Entonces la probabilidad de *k* excedencias del valor 60 en los próximos cinco años vendrá dada por:

$$p[E_n(60) = k] = \begin{bmatrix} 5 \\ k \end{bmatrix} [\exp(-1.2)]^5 [1 - (\exp - 1.2)]^{5-k}$$

K	0	1	2	3	4	5
$p[E_n(60) = k]$	0,166	0,359	0,309	0,133	0,028	0,002

Otro problema interesante, relativo a excedencias, se plantea en los siguientes términos:

Determinar la probabilidad de *k* excedencias en los siguientes *N* ensayos de la m-ésima observación más grande de las *n* totales.

Si p (m) representa la probabilidad de excedencia de la m-ésima observación más grande de las *n* totales, la probabilidad de *k* excedencias en los siguientes *N* ensayos $P^*[n, m, N, k/p(m)]$, será una binomial

$$P^* [n, m, N, k/p(m)] \rightarrow B[p(m), N]$$

$$\text{con } p(m) = 1 - F(X_{n-m+1:n}) = F(X_{m:n}) = U_{m:n}$$

donde $U_{m:n}$ es el m-ésimo estadístico ordenado de una ley uniforme en (0, 1), cuya p.d.f. viene dada por:

$$f_{p(m)} = \frac{1}{\beta(m, n-m+1)} p(m)^{m-1} [1 - p(m)]^{n-m}$$

$$\text{con } 0 \leq p(m) \leq 1$$

de donde la probabilidad incondicional de $P^*(n, m, N, k)$ vendrá dada por

$$P^*(n, m, N, k) = \int_0^1 P^*[n, m, N, k / p(m)]\, dF_{p(m)} [p(m)] =$$

$$= \int_0^1 \begin{bmatrix} N \\ k \end{bmatrix} p(m)^k [1 - p(m)]^{N-k} \frac{1}{\beta(m, n-m+1)} p(m)^{m-1}$$

$$[1 - p(m)]^{n-m} dp(m) = \begin{bmatrix} N \\ k \end{bmatrix} \frac{\beta(k+m, N-k+n-m+1)}{\beta(m, n-m+1)}$$

Con valor medio de $E P^*(n, m, N, k) = E_{p(m)} \{E[P^*(n, m, N, k / p(m))]\} = E_{p(m)} [Np(m)] = NE[p(m)] = NE(U_{m:n}) = Nm / (n+1)$.

Y una varianza de;

$$\operatorname{Var} P^*(n, m, N, k) = E_{p(m)} \{\operatorname{Var}[P^*(n, m, N, k / p(m))]\} +$$

$$+ \operatorname{Var}_{p(m)} \{E[P^*(n, m, N, k / p(m))]\} = E[p(m)] [np(m)] [1 - p(m)] +$$

$$+ \operatorname{Var}[Np(m)] = Nm(n-m+1)(N+n+1) / (n+1)^2 (n+2)$$

Ejemplo 2: La máxima tasa bruta de natalidad en un determinado país durante los últimos 40 años ha sido 28 nacimientos por 1.000 habitantes. Determinar la media y la varianza del número de excedencias de la mayor tasa anual, 28, durante los siguientes 30 años.

$$E P^*(n, m, N) = E P^*(40, 1, 30) = 0{,}75$$

$$\operatorname{Var} P^*(40, 1, 30) = 1{,}20$$

Períodos de retorno

Supóngase un suceso tal que la probabilidad de ocurrencia en un período unitario de tiempo es *p*, y que las ocurrencias de los sucesos en diferentes períodos

son independientes. En tal situación el tiempo, medido en períodos unitarios, hasta la ocurrencia del primer suceso sigue una distribución geométrica de media $1/p$.

En estas condiciones: dado un suceso *S* y sea *T* el tiempo aleatorio hasta su ocurrencia. El valor medio τ de *T* se denomina período de retorno del suceso *S*.

Nótese que si F (x) es la función de distribución de la variable aleatoria «Máxima anual», el período de retorno hasta que la variable aleatoria exceda el valor *x* es $1 / [1 - F(x)]$ períodos.

De esta manera la probabilidad de que ocurra el suceso *S* antes que el período de retorno es: $F(\tau) = 1 - (1-p)^\tau = 1 - (1-p)^{1/p}$, que en el límite vale 0,6321.

Ejemplo 3: La función de distribución de las máximas cuantías anuales en miles de pesetas que una compañía de seguros de automóviles debe pagar en concepto de primas a terceros viene dada por:

$$F(x) = \exp\{-\exp[-0{,}00284(x - 1.297{,}4648)]\}$$

los períodos de retorno de las máximas cuantías de 1,7 y 2 millones de pesetas son:

$$\tau\, 1.700 = 1 / [1 - F(1.700)] = 3{,}6 \text{ años}$$

$$\tau\, 2.000 = 1 / [1 - F(2.000)] = 7{,}9 \text{ años}$$

Esto significa que en media una cuantía de 1,7 millones y de 2 millones ocurren una vez cada 3,6 y 7,9 años.

3. DISTRIBUCIONES EXACTAS DE ESTADISTICAS DE ORDEN. EXTREMOS

Sea $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ la muestra ordenada correspondiente a la variable *X*.

Introducimos el indicador: $l_j(x) = 1$ si $X_j < x$; $l_j(x) = 0$ en otro caso, que verifica: $p[l_{j+}(x) = 1] = F(x)$.

Entonces la distribución del estadístico ordenao r-ésimo viene dada por:

$$F_{Xr:n}(x) = p(X_{r:n} \leq x) = p(\text{al menos r de los } X_i \text{ son } \leq x) =$$

$$= p\left[\sum_{i=1}^{n} 1_i(x) \geq r\right] = \sum_{i=r}^{n} \begin{bmatrix} n \\ i \end{bmatrix} F^i(x)\,[1 - F(x)]^{n-i}$$

En particular, las distribuciones del máximo y del mínimo serán:

$$F_{X_{n:n}}(t) = p\,X_{n:n} \leq x) = [F(t)]^n$$
$$F_{X_{1:n}}(t) = p(X_{1:n} \leq x) = 1 - [1 - F(t)]^n$$

La distribución de probabilidad de estos estadísticos de orden también admite la representación habitual:

$$F_{X_{r:n}}(x) = \frac{1}{\beta\,(r, n - r + 1)} \int_0^{F(x)} t^{r-1}\,(1 - t)^{n-r}\,dt$$

y la densidad es, por tanto:

$$f_{X_{r:n}}(x) = \frac{1}{\beta\,(r, n - r + 1)}\,F(x)^{r-1}\,[1 - F(x)]^{n-r}\,f(x)$$

Veamos algunas aplicaciones de estas distribuciones:

Ejemplo 1: Distribución exponencial.

Supongamos que la variable aleatora X representa el tiempo de vida de algún tipo de componente con la propiedad de pérdida de memoria.

Consideremos n realizaciones independientes de la variable X. La distribución de la duración máxima y mínima de dicha componente vendrá dada por:

$$F_{X_{n:n}}(t) = [1 - \exp(-\lambda t)]^n \text{ si } t \geq 0$$
$$F_{X_{1:n}}(t) = 1 - \exp(-\lambda n t) \text{ si } t \geq 0$$

con una distribución del rango de variación $R_n = X_{n:n} - X_{1:n}$ dada por:

$$F_{R_n}(t) = [1 - \exp(-\lambda t)]^{n-1} \text{ si } t \geq 0$$

El estadístico de orden r-ésimo de esta distribución admite una interesante representación como suma de variables independientes también exponenciales:

La densidad conjunta de la muestra ordenada (4) viene dada por:

$$f(x_{1:n}, \ldots, x_{n:n}) = n!\,\exp\{-\sum_{i=1}^{n} x_{i:n}\} =$$

$$= n!\,\exp\{-\sum_{i=1}^{n} (n - i + 1)\,(x_{i:n} - x_{i-1:n})\} \text{ con } x_{1:n} \leq \ldots \leq x_{n:n}$$

considerando ahora la transformación $y_i = (n - i - 1)\,(x_{i:n} - x_{i-1:n})$ de jacobiano $1/n!$ llegamos a que:

$$f(y_1, \ldots, y_n) = \exp\{-\sum_{i=1}^{n} y_i\} \quad \text{si} \quad y_1, \ldots, y_n \geq 0$$

de donde se sigue que las variables y_i $1 \leq i \leq n$ son independientes, dada la factorización de la función de densidad conjunta. A partir de esta relación podemos escribir el estadístico ordenado r-ésimo como:

$$X_{r:n} = \sum_{i=1}^{r} \frac{1}{n - r + 1} Y_i \quad \text{si} \quad 1 \leq r \leq n$$

A partir de esta representación los tiempos medios de recurrencia de cualquier estadístico de orden vienen dados por:

$$EX_{r:n} = \frac{1}{\lambda} \sum_{k=n-r+1}^{n} \frac{1}{k} \quad \text{y en particular:}$$

$$EX_{n:n} = \frac{1}{\lambda} \sum_{k=1}^{n} \frac{1}{k}; \; EX_{1:n} = \frac{1}{\lambda n}$$

4. DISTRIBUCIONES ASINTOTICAS DE EXTREMOS. DETERMINACION DEL DOMINIO DE ATRACCION. EJEMPLOS

En este apartado vamos a tratar el tema de determinar la distribución asintótica de los valores extremos. Vamos a verlo únicamente para el caso de máximos. Si quisiéramos estudiarlo para el caso de mínimos no habría más que tener en cuenta que mín $\{X\} = -$ máx $\{-X\}$.

Denotando por $F_{X_{n:N}}(x)$ la distribución de máximos, una primera consideración nos lleva a que tal distribución es degenerada, ya que al ser $F_{X_{n:n}}(x)$ la potencia n-ésima de F, y estar F acotada en [0, 1] al ser una probabilidad, en el límite tiende a 1 si $F(x) = 1$ y a 0 si $F(x)$ es estrictamente menor que 1.

Sin embargo, definiendo una transformación lineal de los datos $a_n + b_n x$, donde a_n y $b_n > 0$ son dos sucesiones de números reales que varían continuamente con el tamaño de la muestra (idea debida a Fisher & Tippet), nos encontramos únicamente con tres distribuciones límites, conocidas como distribuciones de extremos de Frechet Weibull y Gumbel con función de distribución:

$$H_{1\gamma}(x) = \exp(-x^{-\gamma}) \quad \text{si} \quad x \geq 0$$
$$H_{2\gamma}(x) = \exp[-(-x)^{\gamma}] \quad \text{si} \quad x \leq 0$$
$$H_{30}(x) = \exp[-\exp(-x)] \quad \text{si} \quad -\infty < x < \infty$$

que podemos englobar dentro de una misma familia de distribuciones conocida como distribución de Von-Mises, con distribución:

$$H_c(x) = \exp[-\{1 + c(x - \lambda / \delta)\}^{-1/c}] \quad \text{si} \quad 1 + c(x - \lambda / \delta) \geq 0$$

siendo λ un parámetro de localización, δ un parámetro de escala y c un parámetro de forma que, dependiendo del signo de c, positivo, negativo o cero, nos da las tres distribuciones antes citadas.

Es decir, según lo visto anteriormente, se tiene el siguiente límite:

$$\lim_{n \to \infty} F_{X_{n:n}}(a_n + b_n x) = H_c(x)$$

de donde podemos tomar la aproximación:

$$F_{X_{n:n}}(x) \approx H_c[(x - a_n) / b_n] \quad \text{si} \quad n \text{ «grande»}$$

La relación anterior se lee " X pertenece al dominio de atracción para máximos de H_c, Frechet, Gumbel o Weibull. Obsérvese que las constantes de normalización a_n, b_n lo que hacen es, de alguna manera, recoger la distribución de la degeneración obvia y situarla en su verdadero dominio de atracción.

Hay que notar que precisamente estas tres distribuciones son las únicas que presentan un comportamiento estable frente a máximos (existen otras equivalentes para el caso de mínimos

$$L_{1\gamma}(x) = 1 - \exp[(-(-x)^{-\gamma}] \quad \text{si} \quad x \leq 0$$

$$L_{2\gamma}(x) = 1 - \exp(-x)^{\gamma} \quad \text{si} \quad x \geq 0$$

$$L_{30}(x) = 1 - \exp[-\exp(x)] \quad \text{si} \quad -\infty < x < \infty$$

llamadas distribuciones de Frechet, Weibull y Gumbel para mínimos, que poseen unas propiedades totalmente paralelas que las de máximos), lo que querrá decir que la distribución de valores máximos de una Frechet, Gumbel y Weibull sigue siendo una Frechet, Gumbel y Weibull.

Daremos a continuación el teorema que nos da las condiciones para la determinación de la distribución asintótica de una dada, así como las constantes de normalización: denotaremos por, $\alpha(F) = \inf\{x / F(x) > 0\}$ y por $w(F) = \sup\{x / F x) < 1\}$ los límites inferior y superior de la función de distribución.

Teorema

i) Supongamos que $w(F) = \infty$ y que existe constante $\gamma > 0$ tal que:

$$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\gamma}$$

entonces x pertenece al dominio de atracción de una Frechet.

ii) Si $w(F)$ es finita y la función $F^*(x) = F[w(F) - 1 / x]$ verifica el límite anterior, entonces la variable aleatoria X pertenece al dominio de atracción de una Weibull.

iii) Supongamos que la vida residual media $R(t) = E(X - t / X \geq t)$ de la variable X es finita y, además, verifica:

$$\lim_{t \to w} \frac{1 - F[t + x R(t)]}{1 - F(t)} = e^{-x}$$

entonces X pertenece al dominio de atracción de una distribución de Gumbel.

Las constantes de normalización en los tres casos vienen dadas por:

i) $a_n = 0$, $b_n = \inf\{x / 1 - F(x) \leq 1 / n\}$

i) $a_n = w(F)$, $b_n = w(F) - \inf\{x / 1 - F(x) \leq 1 / n\}$

i) $a_n = \inf\{x / 1 - F(x) \leq 1 / n\}$, $b_n = R(a_n)$

Una consecuencia del resultado *ii)* es que toda variable aleatoria acotada superiormente tiene como distribución límite de extremos una Weibull, suponiendo verifique el límite antes citado.

Es interesante señalar dos resultados que podríamos llamar «rápidos», deducidos del teorema anterior, que también nos permiten determinar el dominio de atracción.

Resultado rápido 1

Sea $F(x)$ una distribución de probabilidad con $W(F) = \infty$, para la cual existe x_1 tal que si $x \geq x_1$ $f(x) = F'(x)$, existe y es continua. Entonces si:

$$\lim_{x \to \infty} \frac{x f(x)}{1 - F(x)} = \gamma$$

$F(x)$ pertenece al dominio de atracción de una $H_{1,\gamma}(x)$ si γ es un valor finito y positivo.

Resultado rápido 2

Sea $F(x)$ una distribución de probabilidad para la cual existe un número x_1 tal que $x_1 \leq x < W(F)$, $f(x) = F'(x)$ y $F''(x)$ existe. Entonces si;

$$\lim_{x \to w} \frac{d}{dx}\left[\frac{1 - F(x)}{f(x)}\right] = 0$$

$F(x)$ pertenece al dominio de atracción de una Gumbel, $H_{3,0}(x)$.

Veamos algunos ejemplos de la determinación de los dominios de atracción:

Ejemplo 1: (Distribución beta de primera especie). La distribución beta de primera especie viene dada por la p.d.f.:

$$f_{p,q}(x) = [1 / \beta(p, q)]\, x^{p-1} (1 - x)^{q-1} \quad \text{con } 0 \leq x \leq 1$$

de múltiples aplicaciones, dada su flexibilidad funcional. Como w(F) = 1:

$$\lim_{t \to \infty} \frac{1 - F^*(tx)}{1 - F^*(t)} = \lim_{t \to \infty} \frac{f(1 - 1/tx)}{xf(1 - 1/t)} =$$

$$= \lim_{t \to \infty} \frac{(1 - 1/tx)^{p-1} (1/tx)^{q-1}}{x(1 - 1/t)^{p-1} (1/t)^{q-1}} = x^{-q}$$

de donde se deduce que la distribución asintótica es una Weibull.

Ejemplo 2: (Distribución de Pareto). $F(x) = 1 - (\beta / x)^k$ con $x \geq \beta$, usada para ajustar distribuciones de la renta.

$$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = \lim_{t \to \infty} \frac{\beta^k t^k}{t^k x^k \beta^k} = x^{-k}$$

de donde la distribución de Pareto pertenece al dominio de atracción de una Frechet, $H_{1,k}$, con constantes de normalización: $a_n = 0$ y $b_n = \beta n^{1/k}$.

Ejemplo 3: (Distribución exponencial).F(x) = 1 — exp(— X / μ) si $x \geq 0$. La vida residual media de la variable viene dada por;

$$R(t) = E(X - t / X \geq t) = \frac{1}{1 - F(t)} \int_t^{w(F)} [1 - F(x)]\, dx = \mu$$

$$\lim_{t \to \infty} \frac{1 - F[t + x R(t)]}{1 - F(t)} = \lim_{t \to \infty} \frac{\exp[-(t + x\mu)/\mu]}{\exp(-t/\mu)} = e^{-x}$$

de donde la exponencial pertenece al dominio de atracción de una Gumbel, con constantes de normalización $a_n = \mu \mathrm{Ln}(n)$, $b_n = R(a_n) = \mu$. En otras palabras:

$$\lim_{n \to \infty} F^n[\mu x + \mu \mathrm{Ln}(n)] = e^{-e^{-x}}$$

Para acabar esta sección dedicada a la determinación de la distribución asintótica del máximo vamos a dar un concepto que puede ser de gran utilidad a la hora de determinar tal distribución asintótica: se trata del concepto de equivalencia en la cola de funciones de distribución.

Hasta ahora para calcular la distribución asintótica del máximo de una variable *X* únicamente hemos hecho uso de su función de distribución F(x) y de las sucesiones a_n, b_n que determinan el dominio de atracción. Sin embargo, si pretendemos comparar las distribuciones asintóticas de dos variables, una comparación entre sus distribuciones nos traería problemas de identificación, ya que

cualquier función de distribución tiende a la unidad cuando *x* tiende a w (F). Por esta razón es más lógico comparar sus funciones de azar $1 - F(x) = p(X > x)$, que se convertirán en infinitésimos cuando *x* tiende a w (F). Esto es equivalente a comparar los inversos de las funciones de azar, es decir, los períodos de retorno.

Esto nos lleva a la siguiente definición: dos funciones de distribución $F_X(x)$, $F_Y(x)$ se dice que son equivalentes en la cola si $w(F_X) = w(F_Y)$ y:

$$\lim_{x \to w(F)} \frac{p(X > x)}{p(Y > x)} = 1$$

El siguiente teorema dado por Resnick (9) nos demuestra la relación entre dos dominios de atracción y equivalencia en la cola:

Teorema

Sean $F_X(x)$, $F_Y(x)$ dos funciones de distribución equivalentes en la cola. Entonces si:

$$\lim_{n \to \infty} F_{X_{n:n}}(a_n + b_n x) = H_c(x)$$

se verifica que:

$$\lim_{n \to \infty} F_{Y_{n:n}}(a_n + b_n x) = H_c(x)$$

es decir, las dos distribuciones tienen el mismo dominio de atracción.

En la tabla adjunta se presentan los dominios de atracción de las principales distribuciones continuas.

Dominios de atracción de algunas distribuciones continuas

Distribución	Aplicaciones de la distribución	Máximos	Mínimos
Beta de 1.ª especie ..	Tiempos de actividades en el PERT.	Weibull	Weibull
Beta de 2.ª especie ..	Hidrología y meteorología.	Frechet	Weibull
Cauchy	Distribuciones de ingresos.	Frechet	Frechet
Exponencial	Procesos con pérdida de memoria. Tiempos de espera en proceso de Poisson.	Gumbel	Weibull
Gamma	Tiempos de espera en proceso de Poisson.	Gumbel	Weibull
Logística	Modelos de crecimiento.	Gumbel	Gumbel
Log-Normal	Modelos de crecimiento.	Gumbel	Gumbel
Normal	Procesos suma de múltiples causas.	Gumbel	Gumbel
Pareto	Distribución de la renta.	Frechet	Weibull
Rayleigh	Teoría de la fiabilidad.	Gumbel	Weibull
Uniforme	Equiprobabilidad en intervalo.	Weibull	Weibull

5. EJEMPLO DE APLICACION

El número de siniestros de los suscriptores a una compañía de seguros durante un determinado período de tiempo se supone sigue una distribución:

$$p(N = n) = p^n(1 - p) \text{ si } n = 0, 1, 2, \ldots$$

Se ha estimado que la cuantía de cada uno de los accidentes sigue una distribución exponencial de media μ.

5.1. Distribución del coste anual que debe abonar la compañía al cabo de ese período de tiempo

Si X_i representa las cuantías de cada uno de los accidentes, se tiene que:

$$C_N = X_1 + X_2 + \ldots + X_N$$

$$p(C_N = 0) = p(\text{no haya siniestros}) = p(N = 0) = 1 - p$$

Es claro que $C_{N/N=n}$ sigue una distribución gamma, $G(n, 1/\mu)$, de donde para hallar la distribución de C_N aplicaremos el teorema de las probabilidades totales:

$$f_{C_N}(x) = \frac{d}{dx} F_{C_N}(x) = \sum_{n=1}^{\infty} \frac{d}{dx} F_{C_N}(x)\, p(N = n) =$$

$$= \sum_{n=1}^{\infty} \frac{1}{\mu^n \Gamma(n)} x^{n-1} \exp\left(-\frac{x}{\mu}\right) p^n(1 - p) =$$

$$= p(1 - p) \frac{1}{\mu} \exp\left(-\frac{x}{\mu}\right) \sum_{n=1}^{\infty} \left(\frac{xp}{\mu}\right)^{n-1} \frac{1}{(n-1)!} =$$

$$= \frac{1}{\mu} p(1 - p) \exp\left(-\frac{1}{\mu}\right) [(1 - p)x]$$

de donde podemos calcular la función de distribución:

$$F_{C_N}(x) = p(C_N \leq x) = p(C_N = 0) + p(0 < C_N \leq x) =$$

$$= 1 - p \exp\left\{-\frac{1}{\mu}(1 - p)x\right\}$$

El valor medio y la varianza de C_N vienen dados por:

$$EC_N = EXEN = \mu p(1 - p)^{-1}$$

$$\text{Var } C_N = E[\text{Var}(C_{N/N=n})] + \text{Var}[E(C_{N/N=n})] = E(N\mu^2) +$$

$$+ \text{Var}(\mu N) = \mu^2 p(2 - p)(1 - p)^{-2}$$

5.2. Distribución asintótica de los máximos costes anuales que debe abonar la compañía en un largo período de tiempo

Para determinar tal distribución límite vamos a aplicar el segundo resultado rápido:

Sea $x_1 > 0$, entonces si $x > x_1$, se tiene que:

$$\frac{d}{dx}\left(\frac{1 - F(x)}{f(x)}\right) = -1 - \frac{[1 - F(x)]\, f'(x)}{f(x)^2} =$$

$$= -1 - \frac{-p^2(1 - p)^2 \mu^{-2} \exp[-2(1 - p)\mu^{-1}x]}{p^2(1 - p)^2 \mu^{-2} \exp[-2(1 - p)\mu^{-1}x]} = -1 - (-1) = 0$$

de donde la distribución asintótica es una distribución de Gumbel $H_{3,0}(x)$.

5.3. Distribución exacta de la cuantía del mayor siniestro

Si denotamos esa mayor cuantía por $X_{N:N}$, $p(X_{N:N} = 0) = p(N = 0) = 1 - p$ para los demás valores de la distribución:

$$F_{X_{N:N}}(x) = p(X_{N:N} = 0) + p(0 < X_{N:N} \leq x) =$$

$$= 1 - p + \sum_{n=1}^{\infty} F_{X_{N:N/N=n}}(x)\, p(N = n) = 1 - p +$$

$$+ \sum_{n=1}^{\infty} (1 - e^{-x/\mu})^n\, p^n (1 - p) =$$

$$= 1 - p + (1 - p) \sum_{n=1}^{\infty} [p(1 - e^{-x/\mu})]^n = \frac{1 - p}{1 - p(1 - e^{-x/\mu})}$$

que es precisamente una distribución logística discretizada en un punto.

5.4. Distribución asintótica de las máximas cuantías de los mayores siniestros en dichos períodos de tiempo

Para determinar el dominio de atracción de la distribución anterior, basta tener en cuenta que la distribución logística, $F_\mu(x) = (1 + e^{-x/\mu})^{-1}$ es equivalente en la cola derecha a una exponencial, que pertenece al dominio de atracción de Gumbel.

Por otra parte, esta última distribución es equivalente a una logística, ya que:

$$\frac{1 - F(x)}{1 - F_\mu(x)} = \frac{pe^{-x/\mu}}{1 - p + pe^{-x/\mu}} \cdot \frac{1 + e^{-X/\mu}}{e^{-x/\mu}}$$

que en el límite tiende a 1.

Por tanto, su distribución asintótica es otra distribución de Gumbel con constantes de normalización: $a_n = \mu$ y $b_n = \mu \mathrm{Ln}(n)$.

BIBLIOGRAFIA

(1) ANSOFF, H. I.; BOSMAN, A., y STORN, P. M. (1982): *Understanding and Managing Strategic Chance: Contributions to the Theory and Practice of General Management.* North-Holland, Amsterdam.

(2) CASTILLO, E. (1987): *Extreme Value Theory in Engineering.* Academic Press.

(3) CASTILLO, E., y GALAMBOS, J. (1986): «Determining the Domain of Attraction of a Parent Distribution». *Technical Report.* Temple University.

(4) DAVID, H. A. (1981): *Order Statistics.* John Wiley and Sons. New York.

(5) FISHER, R. A., y TIPPET, L. H. C. (1928): *Limiting forms or the frequency distribution of the largest or smallest member of a sample.* Proc. Camb. Phil. Soc. 24, 180-90 (258, 264).

(6) GALAMBOS, J. (1978): *The asymptotic Theory of Extreme Order Statistics.* John Wiley and Sons (segunda edición, por Krieger, 1987).

(7) PATIL, G. P.; BOSWELL, M. T., y RATNAPARKHI, M. V. (1984): *Dictionary and Classified Bibliography of Statistical distributions in Scientific Word.* Vol. 2: «Continuous Univariate Models. International Cooperative Publishing House».

(8) LOPEZ MORENO, M. J. (1987): «La gestión de la empresa en un contexto de cambio». II Encuentro Luso-Español de Economía Empresarial. Madrid.

(9) RESNICK, S. I. (1971): *Tail equivalence and applications.* JAP 8, 136-156.

(10) TIAGO DE OLIVEIRA, J.: *Statistical Extremes and Applications.* Nato Asi Series. D. Reidel Company.