Estudio de la siniestralidad y aplicaciones económicas en las entidades aseguradoras (*)

ANA ISABEL CID CID

FAC. DE CIENCIAS JURÍDICAS Y SOCIALES, UNIVERSIDAD REY JUAN CARLOS. MADRID

La solvencia de las entidades aseguradoras es uno de los factores fundamentales para medir su estabilidad y garantía de operación. Los ingresos por primas y los gastos de indemnizaciones tienen una influencia crítica en su capacidad financiera.

Mientras las primas pueden presupuestarse con un ligero margen de variación, la siniestralidad está sujeta a una mayor aleatoriedad, que puede ser acotada mediante modelos analíticos.

Introducción

El seguro, tal y como establece el artículo 1 de la Ley 50/80 de 8 de octubre, de Contrato de Seguro, se define como «una operación por la que el asegurador se obliga, mediante el cobro de una prima y para el caso de que se produzca el evento cuyo riesgo es objeto de la cobertura, a indemnizar el daño producido al asegurado».

La inversión del proceso productivo, cobrando las primas a priori como precio del servicio de aseguramiento que prestan, para indemnizar al asegurado posteriomente si se produce el siniestro, obliga a la compañía a tener un conocimiento lo más exacto posible de la evolución de los siniestros a lo largo del tiempo. De esta forma se podrá contar con reservas suficientes que garanticen el cumplimiento de los contratos suscritos. El estudio del fenómeno de

(*) Resumen del texto de la beca concedida por la Fundación MAPFRE Estudios.

la siniestralidad en no vida, en una empresa aseguradora es por tanto de gran importancia dadas las características de operación de la misma.

La percepción anticipada de la prima por parte del asegurador obliga a éste a determinarla según estimaciones de carácter aleatorio sobre la siniestralidad futura. La estimación del coste total de los riesgos a que están expuestos los asegurados hace que se puedan determinar de forma justa y suficiente las primas que éstos deben aportar. Un conocimiento insuficiente de la siniestralidad puede hacer que las primas estén infravaloradas llevando a una situación insolvente del asegurador.

La indemnización es la obligación fundamental del asegurador y para poder indemnizar dentro de los límites pactados se debe disponer de reservas suficientes. De nuevo el conocimiento, lo más preciso posible, de la distribución de siniestralidad permite a la entidad hacer frente a las obligaciones que ésta tiene con el asegurado.

La Teoría del Riesgo (incluida en la Matemática Actuarial) es esencial en todo este proceso. Su objeto es el tratamiento del riesgo (característica básica de las empresas de seguros) y proporcionar modelos matemáticos para analizar las fluctuaciones aleatorias de variables expuestas a riesgos en general. Fundamentalmente es un conjunto de ideas relacionadas para diseñar, administrar y regular una empresa de riesgo, como es la entidad aseguradora. Tiene especial aplicación en el análisis de la entidad aseguradora por tratarse de una operación aleatoria y estar determinada por un conjunto de variables que llevan asociada incertidumbre.

En concreto, destaca la Teoría del Riesgo Colectivo que trata a la entidad aseguradora como un todo, estudiando el desarrollo del negocio del seguro desde el punto de vista probabilístico. Se examina el comportamiento estadístico de las dos variables aleatorias que intervienen en la determinación de la siniestralidad: el número de siniestros y la cuantía o costes de cada uno de ellos, introduciendo modelos generales de probabilidad que permiten dar respuesta a preguntas tales como cuál es la probabilidad de ruina de la compañía en un momento determinado.

Además, el análisis de la estabilidad de la empresa de seguros (requisito fundamental para la solvencia) desde el enfoque financiero-actuarial, identifica a la empresa con la cartera de seguros que posee y se aplica de nuevo la Teoría del Riesgo Colectivo para lograr el objetivo de la garantía. Esto es, evitar que para unos precios dados y equitativos (primas), la probabilidad de ruina (probabilidad de que los siniestros sean superiores a las primas) supere un valor prefijado.

La distribución de siniestralidad es por tanto una herramienta estadística que permite estimar los gastos por siniestros, ayuda a medir la solvencia financiera y posibilita poner precio o tarificar los distintos riesgos o pólizas de la cartera.

Variables que intervienen en la distribución de siniestralidad

El estudio realizado se centra en los seguros generales o no vida, con una estructura estocástica diferente a la de los seguros de vida. La probabilidad de los sucesos no depende de una sola característica o variable única y las desviaciones aleatorias debidas a siniestros extraordinarios, en comparación con las primas, son de mayor importancia que en los seguros de vida.

Si se analiza el comportamiento estadístico de la siniestralidad desde la perspectiva de la Teoría del Riesgo Individual, la ganancia o pérdida total de la empresa sigue una distribución de probabilidad aproximadamente normal si el número de pólizas es suficientemente grande. En este caso nos limitamos a una única función de distribución.

Hay que resaltar en este punto la importancia de la distribución normal puesto que muchos fenómenos que se presentan en la naturaleza tienen una distribución de probabilidad de este tipo. Si bien es cierto que el cálculo de probabilidades de forma analítica es complicado, mediante un sencillo cálculo se puede transformar cualquier distribución normal en otra que se caracteriza por tener una media igual a 0 y desviación típica 1. De esta forma las probabilidades se pueden obtener mediante el uso de cualquiera de las múltiples tablas de probabilidad que sobre dicha distribución existen.

Por el contrario, si se examina el comportamiento estadístico de las dos variables aleatorias que intervienen en la determinación de la siniestralidad (número de siniestros y cuantía) existen múltiples modelos que se ajustan a ambas distribuciones y a la cuantía total.

Respecto a la variable aleatoria discreta «v», que representa el número de siniestros, decir que se trata de una variable aleatoria discreta de tal forma que se puede calcular la probabilidad de que tome valores concretos, es decir, la probabilidad de que ocurran un número determinado de siniestros entre todos los entes expuestos al riesgo. Se puede partir de los esquemas de las urnas de Bernouilli y de Polya-Eggenberger, que se basan en calcular la probabilidad de extraer bolas de distinto color de una urna, para determinar su función de distribución. Si «N» representa el número de entes expuestos al riesgo (finito pero desconocido), diferenciando entre que se produzca o no contagio después de cada extracción de bola (si un siniestro afecta a otros expuestos al riesgo o no), se tiene una distribución de Polya-Eggenberger en el caso de contagio positivo o, en el caso más general, una distribución de Polya-Eggenberger generalizada. Si no hay contagio, la distribución de la variable aleatoria es binomial.

En general no es posible conocer el número «N» de objetos expuestos al riesgo. Surge de esta forma el proceso de paso al límite, haciendo crecer «N» hasta infinito, que define la distribución de Poisson (ausencia de contagio) a partir de la distribución binomial, o bien la distribución binomial negativa (contagio positivo débil) como límite de la distribución Polya-Eggenberger.

Destacar la distribución de Poisson que resulta adecuada en muchas ocasiones para representar el número de siniestros que tienen lugar en una cartera de pólizas. En un periodo actuarial de observación «t» (t puede ser un día, una semana, un mes...), la probabilidad de que ocurran n siniestros es:

$$P_n(t) = P\,(v = n) = \frac{\lambda^n}{n!} \cdot e^{-\lambda}$$

El parámetro «λ» es la media de la distribución de Poisson, pudiéndose estimar mediante la media muestral, a partir de los datos disponibles sobre el número de siniestros en la entidad aseguradora. «λ» será por tanto el número medio de siniestros ocurridos en una muestra de expuestos al riesgo durante el periodo de observación actuarial considerado.

El coste de cada uno de los siniestros se representa por medio de la variable aleatoria continua «ζ». Por ser una variable continua, la probabilidad de que un siniestro tenga un coste exacto de «x» unidades monetarias (x arbitrario) es cero. No se puede calcular por tanto la probabilidad de cuantías unitarias aunque si es posible calcular la probabilidad de que los siniestros sean de coste inferior a un valor dado o que su coste pertenezca a un intervalo determinado.

Existen también distintas distribuciones de probabilidad que representan el coste de cada uno de los siniestros como la distribución gamma, la logarítmico-normal, distribución de Pareto o la distribución de Weibull.

La distribución gamma es una de las más importantes desde el punto de vista actuarial. Además de ser muy manejable matemáticamente es una de las que mejor se ajustan, en múltiples casos, a la distribución empírica de la cuantía de los siniestros. Como caso especial de la distribución gamma, destacar la distribución exponencial de parámetro «h», con función de densidad:

$$f(x) = h \cdot e^{-hx}; \quad \text{para } x > 0$$

La probabilidad de que la cuantía de un siniestro no sea superior a x unidades monetarias se determina mediante la integral:

$$P(\zeta \leq x) = \int_0^x h \cdot e^{-hx}$$

El parámetro «h» se puede estimar de diferentes formas, llegando al mismo valor si se hace por el método de los momentos o por máxima verosimilitud:

$$h = \frac{1}{\bar{x}}$$

Es decir, se estima como el inverso del coste medio de los siniestros de una muestra aleatoria.

Analicemos ahora cuál es la distribución de probabilidad de la cuantía del daño total, variable aleatoria «ξ», sufrido por la cartera del ente asegurador cuando el número de siniestros es una variable aleatoria que sigue una ley de Poisson. La distribución del daño total recibe el nombre de *Distribución de Poisson Generalizada*, siendo su función de distribución:

$$F(x,t) = \sum_{n=0}^{\infty} \frac{e^{-\lambda} \cdot (\lambda)^n}{n!} \cdot V^{n^*}(x)$$

«$F(x,t)$» representa la probabilidad de que en el periodo de observación actuarial «t» el coste de todos los siniestros ocurridos no sea superior a «x» unidades monetarias. «$V^{n^*}(x)$» es la probabilidad de que habiendo sucedido «n» siniestros su cuantía total no exceda de «x» unidades monetarias (suponiendo que las cuantías de cada siniestro son variables aleatorias independientes e idénticamente distribuidas con función de distribución «$V(x)$ $V^{n^*}(x)$» es la convolución n-ésima de «$V(x)$» que verifica la relación de recurrencia:

$$V^{n^*}(x) = \int_0^x V^{(n-1)^*}(x-z) \cdot dV(z)$$

con:

$$V^{0^*}(x) = \begin{cases} 0 & si\ x < 0 \\ 1 & si\ x \geq 0 \end{cases} \text{ y } V^{1^*}(x) = V(x)$$

En general la expresión de $F(x,t)$ no se puede determinar de forma explícita. Para estudiar dicha función existen diferentes métodos de aproximación (útiles para estimar probabilidades máximas de pérdidas y probabilidades de ruina) como las series de Edgeworth, aproximación normal, aproximación NP, método de Montecarlo o aproximación de Esscher y métodos exactos como la inversión de la función característica de Fourier o la fórmula de recurrencia de Panjer.

Es importante la aproximación normal (construida mediante la aplicación del Teorema Central del Límite) por ser la más clásica y una de las más frecuentemente aplicadas, además de ser bastante intuitiva. Si las cuantías de los siniestros son variables aleatorias estocásticamente independientes (el coste de un siniestro no influye en el coste de cualquier otro), el Teorema Central del Límite permite afirmar que «$F(x,t)$» tiende a una distribución normal cuando «t» tiende a infinito. Esto es, la función de distribución, que proporciona la probabilidad de que el coste de todos los siniestros ocurridos en la entidad aseguradora durante el período de observación considerado sea inferior a «x» unidades monetarias, es:

$$F(x,t) \cong \phi\left(\frac{x-\mu}{\sigma}\right)$$

siendo «ϕ» la función de distribución de una variable aleatoria con distribución de probabilidad Normal (0,1) y «μ» y «σ», respectivamente, el valor probable y la desviación típica de «$F(x,t)$». La probabilidad buscada se obtiene mediante las tablas de la distribución normal buscando en ellas el valor:

$$\frac{x-\mu}{\sigma}$$

En general «μ» y «σ» son desconocidos pudiéndose estimar su valor a partir de los correspondientes valores muestrales.

Aplicación de las mixturas de distribuciones al estudio de la siniestralidad

Como alternativa a las aproximaciones clásicas que consideran, en el caso del número de siniestros y coste de cada uno de ellos, una única función de distribución, se pueden utilizar mixturas de distribuciones que consisten básicamente en construir una distribución de probabilidad que es combinación lineal de distintas componentes.

Este tipo de distribuciones no ha sido muy estudiado en el ámbito asegurador, quizá debido a la dificultad que aparece en la estimación de los parámetros y, sin embargo, pueden mejorar en gran medida las aproximaciones clásicas. Su utilidad se manifiesta particularmente cuando los datos proceden de distintas poblaciones (en este caso distintas carteras de pólizas) o se dispone de observaciones anómalas.

Una *mixtura finita de distribuciones* se define mediante la siguiente función de cuantía, cuando la variable aleatoria es discreta (o de densidad si la variable aleatoria es continua):

$$f(x)=\sum_{i=1}^{c} P(\theta_i)g(x;\theta_i), \quad \sum_{i=1}^{c} P(\theta_i)=1$$

En la definición aparecen tres tipos de parámetros diferentes: «c» (número de componentes de la mixtura finita), «$P(\theta_i)$» (proporción con que la componente «$g(x;\theta_i)$» aparece en la mixtura, denotada generalmente por «p_i»)» y «θ_i» (vectores paramétricos de los que dependen las densidades componentes) que se pueden estimar mediante el método de máxima verosimilitud, estimación bayesiana, método de inversión y minimización de error, etc...

Las mixturas de un número finito de distribuciones de Poisson o exponenciales se definen mediante la función (de densidad o cuantía):

$$f(x;p,\mu)=\sum_{i=1}^{c} p_i f_i(x;\mu_i), \; x\geq 0, \; m>0$$

donde «$p=(p_1, p_2, ..., p_c)$» es el vector de las «c» proporciones de mixtura que verifica $0 < pi < 1$ y $\sum_{i=1}^{c} p_i = 1$ y μ $(\mu_1, \mu_2, ..., \mu_c)$ es el vector de los parámetros de las distribuciones componentes. Estas distribuciones componentes se caracterizan por la función de cuantía, en el caso de considerar distribuciones de Poisson:

$$P_i\,(\xi_i = x)=\frac{\mu_i^x \cdot e^{-\mu}}{x!}$$

o por la función de densidad, si se trata de exponenciales:

$$f_i(x)=\frac{1}{\mu_i}\cdot e^{-\frac{x}{\mu_i}}, \; x\geq 0, \; \mu_i>0$$

En el ejemplo que se desarrolla a continuación se explica de forma más sencilla el concepto de mixtura de distribuciones y su aplicación en la entidad aseguradora.

Ejemplo

Como ejemplo se estudian dos de las distribuciones que modelan la siniestralidad en una cartera de seguros multirriesgo del hogar en una compañía de seguros privada: número de siniestros y cuantía de los mismos. Como alternativa a las aproximaciones clásicas se buscan mixturas de distribuciones que mejoren las mismas.

La base de datos correspondiente, como se ha comentado, a una cartera de seguros multirriesgo hogar en el territorio español, contiene información referente a la causa del siniestro, fecha de ocurrencia, coste del mismo y fecha de la póliza.

Se considerarán las pólizas durante un periodo de observación actuarial de un año, esto es, desde que se firma el contrato hasta transcurrido un año del mismo. Para el estudio se rechazan las pólizas con vida inferior.

En un primer análisis de los datos disponibles se observa que el 42% de los 17.500 siniestros ocurridos han sido causados por el agua, muy por encima de cualquier otro. El mayor porcentaje de los mismos (12,3%) se registran en el segundo mes de vida de la póliza, a partir del cual el número de siniestros decrece hasta llegar al 5% en el último mes. Comentar también que durante el primer mes se registran el 10,1% de todos los siniestros.

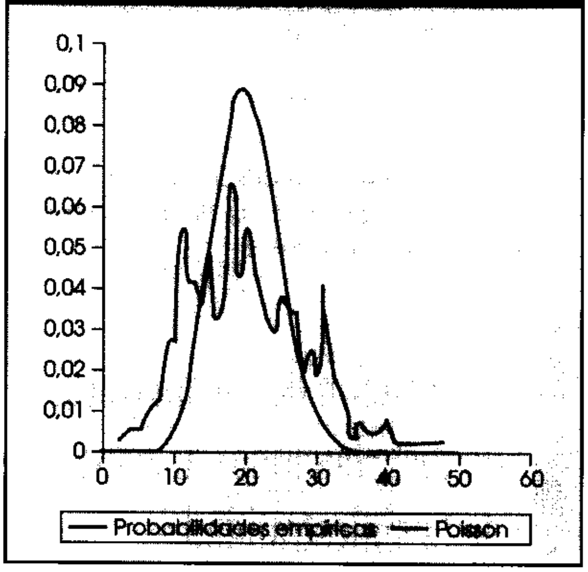
El estudio de la variable aleatoria número de siniestros se realiza para los siniestros comentados (aquellos cuya causa es el agua) en un periodo de observación actuarial de un día. Es decir, se calculará la probabilidad de que ocurran un número determinado de siniestros causados por agua en un día. Respecto a estos siniestros se dispone de 7.350 observaciones.

Se supone que la distribución del número de siniestros es una distribución de Poisson y se estima la media de la misma como la media muestral, *20,1*. Es decir, ocurren diariamente una media de *20,1* siniestros causados por agua. La probabilidad de que un día determinado tengan lugar «*n*» siniestros es por tanto:

$$P_n(t = 1) = P(v = n) = \frac{\lambda^n}{n!} \cdot e^{-\lambda} \frac{20,1^n}{n!} \cdot e^{-20,1}$$

En el gráfico 1 se representan las verdaderas frecuencias y las probabilidades teóricas obtenidas a partir de la distribución de Poisson.

Gráfico 1. Ajuste del número de siniestros a una distribución de Poisson.



Si se supone que la función de cuantía del número de siniestros que se producen diariamente es una mixtura de dos distribuciones de Poisson:

$$P_n(1) = P(v = n) = p \cdot \frac{\lambda_1^n}{n!} \cdot e^{-\lambda_1} + $$
$$+ (1 - p) \cdot \frac{\lambda_2^n}{n!} \cdot e^{-\lambda_2}; \quad \text{cuando } n = 1, 2, 3, \ldots$$

para estimar los parámetros «λ_1», «λ_2» y «p» por el método de los momentos, se igualan los tres primeros momentos factoriales muestrales

$$V_1 = \frac{1}{n}\sum x_i; \quad V_2 = \frac{1}{n}\sum x_i (x_i - 1),$$
$$V_3 = \frac{1}{n}\sum x_i (x_i - 1)(x_i - 2)$$

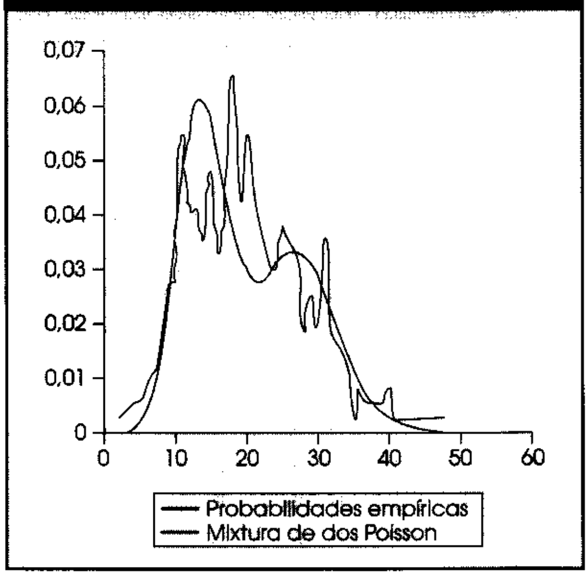
a los momentos poblacionales, de los cuales los momentos muestrales son estimadores insesgados, ($v_1 = \alpha\lambda_1 + (1 - \alpha)\lambda_2$, $v_2 = \alpha\lambda_1^2 + (1 - \alpha)\lambda_2^2$ y $v_3 = \alpha\lambda_1^3 + (1 - \alpha)\lambda_2^3$) resultando un sistema de ecuaciones no lineales cuya solución son las estimaciones de los parámetros desconocidos.

La función de cuantía de la mixtura resultante es:

$$P_n(1) = P(v = n) = 0{,}57 \cdot \frac{14{,}23^n}{n!} \cdot e^{14{,}23} +$$
$$+ 0{,}43 \cdot \frac{27{,}66^n}{n!} \cdot e^{-27{,}66}$$

Esto quiere decir que el 57% de los días hay una media de *14,23* siniestros y en el resto de días una media de *27,66* siniestros causados por agua. En el Gráfico 2 se puede observar, al igual que en el caso anterior, cómo se ajusta esta mixtura de dos distribuciones de Poisson a los datos disponibles. Se observa que la distancia entre los valores de la mixtura y los verdaderos valores correspondientes al número de siniestros es menor que si se considera una única distribución de Poisson.

Gráfico 2. Ajuste del número de siniestros a una mixtura de dos distribución de Poisson.



Si se consideran tres distribuciones de Poisson la función de cuantía:

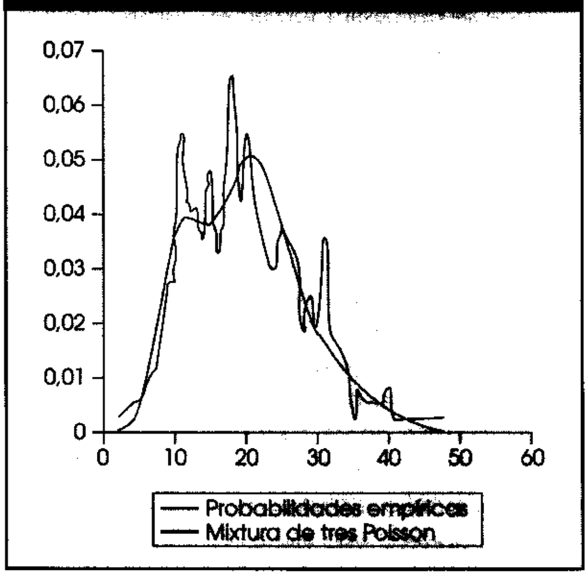
$$P(v = n) = p_1 \cdot \frac{\lambda_1^n \cdot e^{-\lambda_1}}{n!} + p_2 \cdot \frac{\lambda_2^n \cdot e^{-\lambda_2}}{n!} + p_3 \cdot \frac{\lambda_3^n \cdot e^{-\lambda_3}}{n!}$$

los parámetros desconocidos son «λ_1», «λ_2», «λ_3», «p_1», «p_2» y «p_3». Para estimar dichos parámetros se plantea un sistema de seis ecuaciones añadiendo a las tres anteriores las que resultan de igualar los momentos muestrales y poblacionales de orden cuatro, cinco y seis, cuya solución permite determinar los parámetros de la mixtura que es:

$$P(v = n) = 0{,}3 \cdot \frac{11{,}54^n \cdot e^{-11{,}54}}{n!} +$$
$$+ 0{,}56 \cdot \frac{21{,}56^n \cdot e^{-21{,}56}}{n!} + 0{,}14 \cdot \frac{32{,}42^n \cdot e^{32{,}42}}{n!} \quad (V.3)$$

Esto es, el *30%* de los días ocurren como media *11,54* siniestros, el *56%* de los días *21,56* siniestros y tan solo en el *14%* de los días ocurren como media *32,42* siniestros. Gráficamente la representación de la distribución empírica de probabilidades asociadas a los números de siniestros y el correspondiente ajuste a una mixtura de tres componentes de Poisson es la siguiente:

Gráfico 3. Ajuste del número de siniestros a una mixtura de tres distribución de Poisson.



Para determinar cuál es la bondad de ajuste de estas aproximaciones (con qué error se aproxima la distribución teórica a la empírica) se utiliza el estadístico «χ^2» que analiza la diferencia entre los verdaderos valores de la distribución y los teóricos proporcionados por el modelo, válido cuando cada valor (o cada intervalo si se consideran datos agrupados) aparece con una frecuencia absoluta mayor o igual que cinco. El estadístico:

$$\chi^2 = \sum \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$$

donde «n_i» son las frecuencias empíricas de cada valor y «np_i» las teóricas proporcionadas por el modelo («p_i» es la probabilidad asignada por el modelo teórico a un valor concreto). Se considera la diferencia elevada al cuadrado para evitar que se compensen valores positivos y negativos, no proporcionando de esta forma un valor global del error cometido, y se divide entre las frecuencias teóricas. Se dispone así para cada dato el error cometido por valor teórico de la probabilidad. El estadístico definido sigue una distribución de probabilidad «χ^2» con «n-k-l» grados de libertad, siendo «n» el número total de observaciones y «k» el número de parámetros que se han estimado en el modelo.

Si el valor obtenido es menor que «$\chi^2_{\alpha,\ n-k-l}$» (aquel que verifica, para una variable aleatoria «X» con distribución Ji-cuadrado con n-k-l grados de libertad que $P(X > \chi^2_{\alpha,\ n-k-l}) = \alpha$, valor que se obtiene de forma sencilla a partir de las tablas de probabilidad de la distribución Ji-cuadrado) con un nivel de significación «α», se puede admitir que la distribución teórica escogida representa bien a la distribución empírica. El nivel de significación representa la probabilidad de no aceptar que la distribución teórica se ajusta a los datos cuando en realidad si lo hace. En este caso tomamos como probabilidad de cometer este error el valor $\alpha = 0{,}05$.

En el ajuste a una distribución de Poisson se ha estimado únicamente un parámetro «$(k = 1)$». El número de observaciones es de *366* (adminitiendo la posibilidad de años bisiestos), siendo el valor teórico obtenido a partir de las tablas de la distribución Ji-cuadrado con *364* grados de libertad «$\chi^2_{0.05;\ 364} = 409{,}488$» y el valor empírico obtenido a partir de los datos disponibles,

$$\chi^2 = \sum \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i} = 3.403{,}36.$$

Si se ajusta a una mixtura de dos distribuciones de Poisson se estiman cuatro parámetros siendo el valor teórico «$\chi^2_{0.05;\ 361} = 406{,}31$» y el valor empírico

$$\chi^2 = \sum \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i} = 1.388.$$

En consecuencia, se puede afirmar que la mixtura de dos distribuciones de Poisson representa bien a la distribución empírica, no mejorando el ajuste si se consideran tres componentes. En este caso concreto, si se representa el número de siniestros mediante una distribución de Poisson el ajuste no es bueno.

La base de datos contiene *7.350* valores referidos a la cuantía de los siniestros. En un primer análisis se estudia si alguna de las distribuciones teóricas clásicas (distribuciones causal, gamma, exponencial, logarítmico-normal, de Pareto y Weibull) es buena para aproximar dicho coste. Se observa que el mejor ajuste es a una distribución exponencial.

Para estimar los parámetros de la distribución exponencial que modela el coste de los siniestros, así como los parámetros de las mixturas que se consideren, se toma una muestra aleatoria simple de *1.000* observaciones (en lugar de todos los valores) y se determinará para ella el error cometido mediante el uso del estadístico «χ^2».

Mediante el método de los momentos la estimación de la media de la distribución exponencial es «$\mu = 30.657$», lo que significa que el coste medio de los siniestros ocurridos toma dicho valor, siendo por tanto la función de densidad:

$$f(x) = \frac{1}{30.657} \cdot e^{-\frac{x}{30.657}}; \quad \text{donde } x \geq 0$$

lo que permite calcular la probabilidad de que un siniestro tenga un coste inferior a *x* unidades monetarias como:

$$P(\zeta \leq x) = \int_0^x \frac{1}{30.657} \cdot e^{-\frac{x}{30.637}} \cdot dx$$

La expresión de las mixturas es igual al caso anterior (sustituyendo la función de cuantía de la distribución de Poisson por la función de densidad de la distribución exponencial). Para estimar como

$$V_r = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i^r}{\Gamma(r+1)} = \frac{a_r}{\Gamma(r+1)}$$

a los momentos poblacionales (de los que los primeros son estimadores insesgados) obteniéndose un sistema de ecuaciones no lineales cuya solución proporciona estimaciones de los parámetros desconocidos de la misma forma que en el caso anterior.

La función de densidad de la mixtura de dos exponenciales resulta:

$$f(x) = 0{,}958 \cdot \frac{1}{24.215} \cdot e^{-\frac{x}{24.215}} +$$

$$+ 0{,}042 \cdot \frac{1}{178.805} \cdot e^{-\frac{x}{178.805}}; \quad \text{donde } x \geq 0$$

por lo que el *95,8%* de los siniestros tienen una distribución exponencial siendo su media de *24.215* pts. (145 euros, 155 dólares) y el restante *4,2%* son de coste medio *178.805* pts. (1.077 euros, 1.146 dólares).

Si se considera una mixtura de tres distribuciones exponenciales se debe resolver un sistema de seis ecuaciones no lineales. La solución para la media de la primera distribución exponencial que compone la mixtura es negativa (el resultado es *–32.407*), por lo que no se puede realizar el ajuste considerando tres componentes porque no es posible que un siniestro tenga coste negativo.

El análisis de la bondad de ajuste de estas aproximaciones, al igual que en la estimación del número de siniestros, utiliza el estadístico «χ^2» con distribución de probabilidad «χ^2» con «*n-k-l*» grados de libertad, siendo en este caso «n = 1.000».

Fijando un nivel de significación de α = 0,05, en el ajuste a una distribución exponencial se estima únicamente un parámetro, por lo tanto «$\chi^2_{0.05;998} = 1.072{,}605711$», siendo el valor

$$\chi^2 = \sum \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i} = 904{,}8824777.$$

Si se ajusta a una mixtura de dos distribuciones exponenciales se estiman cuatro parámetros siendo «$\chi^2_{0.05;995} = 1.069{,}495144$» y

$$\chi^2 = \sum \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i} = 155{,}6258663.$$

En consecuencia, ambas aproximaciones son válidas para modelar el coste de los siniestros dado que el valor «χ^2» empírico es menor que $\chi^2 = \sum \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$. Se observa que la mixtura de dos distribuciones exponenciales mejora el ajuste a una sola exponencial porque el valor «χ^2» decrece considerablemente.

Conclusiones

La importancia del estudio de la siniestralidad en una entidad aseguradora mediante el uso de distribuciones de probabilidad se pone de manifiesto desde la consideración inicial del objeto de la empresa, es decir, el cumplimiento para con los asegurados de las obligaciones contraidas por la misma.

El conocimiento de la distribución de siniestralidad es de vital importancia tanto para determinar la prima que percibirá el asegurador para la cobertura de los riesgos a que están expuestos los asegurados como para poder prever la solvencia de la entidad, garantizando de esta forma la indemnización al asegurado dentro de los límites pactados.

La *Teoría del Riesgo* es fundamental dado que su objeto es el tratamiento del riesgo, característica básica de la empresa de seguros. Proporciona modelos matemáticos para analizar las fluctuaciones aleatorias de variables expuestas a riesgos en general. Destacar la importancia de la Teoría del Riesgo Colectivo que trata a la entidad aseguradora como un todo, estudiando el desarrollo del negocio de seguro desde el punto de vista probabilístico.

Si se examina el comportamiento estadístico de las dos variables aleatorias que intervienen en la determinación de la siniestralidad (número de siniestros y cuantía) existen múltiples modelos que se ajustan a ambas distribuciones y a la cuantía total.

Las mixturas de distribuciones mejoran en gran medida las aproximaciones clásicas a las distribuciones que caracterizan la siniestralidad ya que supone una generalización de las distribuciones que las componen. Permiten que los valores muestrales procedan de distintas poblaciones lo que es habitual ante grupos de riesgo heterogéneos.

Como se demuestra en el ejemplo considerado, una mixtura de dos distribuciones de Poisson mejora en gran medida el ajuste que si se considera una única distribución de Poisson. De la misma forma, en el estudio de la cuantía de los siniestros la mixtura de dos distribuciones exponenciales mejora el ajuste realizado a una sola distribución, aunque en este caso son válidas ambas aproximaciones.

Como consecuencia, se debe comentar que el uso novedoso de este tipo de distribuciones no se puede obviar en una compañía de seguros. En general se mejoran las aproximaciones clásicas, lo que permite el mejor conocimiento de la siniestralidad con todas las consecuencias que ello implica (garantizar la solvencia, un precio más justo de las primas, mejorar la toma de decisiones, ...).

Bibliografía

Ajne, B. (1974). «On the statistical estimation of cost of claims». *ASTIN Bulletin*, Vol. VII, n.° 3, pp. 181-191.

Bartlett, M. S. and MacDonald, P. D. M. (1968). «Least squares estimation of distribution mixtures». *Nature*, n.° 217, pp. 195-196.

Beard, R. E., Pentikäinen, T. y Pesonen, E. (1969). *Risk Theory*. Chapman y Hall. London. 3.ª edición, (1984).

Blischke, W. R. (1964). «Estimating the parameters of mixtures of binomial distributions». *Journal of American Statistical Association*, n.° 59, pp. 510-528.

Bouyssou, J. (1997). *Théorie génerale du risque*. Económica. París.

Daykin, C. D., Pentikäinen, T. y Pesonen, M. (1994). *Practical risk theory for actuaries*. Chapman & Hall. London.

Grandell, J. (1991). *Aspects of Risk Theory*. Springer Verlag. New York.

Hasselblad, V. (1969). «Estimation of finite mixtures of distributions from the exponential family». *Journal of American Statistical Association*, n.°64, pp. 1.459-1.471.

Legislación: Real Decreto 2.486/1998, de 20 de noviembre, por el que se aprueba el Reglamento de Ordenación y Supervisión de los Seguros Privados.

McLachlan, G. J., Basford, K. E. (1988). *Mixture Models: Inference and Applications to clustering (Statistics, textbooks and monographs; v. 84)*. Marcel Dekker. New York.

Latorre Llorens, L. (1992). *Teoría del riesgo y sus aplicaciones a la empresa aseguradora*. MAPFRE. Madrid.

Nieto de Alba, U. y Vegas, J. (1993). *Matemática actuarial*. MAPFRE. Madrid.

Panjer, H. y Wilmot, G. (1992). *Insurance Risk Models*. Society of Actuaries. Schaumburg. ■