ESTUDIOS DE ECONOMÍA APLICADA N.º 18, 2001. PÁGS. 175-187

Buenos y malos riesgos en seguros: el punto de vista bayesiano basado en distribuciones bimodales

*GÓMEZ DÉNIZ, E. y **PÉREZ SÁNCHEZ, J.M.

Departamento de Métodos Cuantitativos en Economía y Gestión.

Universidad de Las Palmas de G.C.

*Telf.: +34 28 45 18 03-Fax: +34 28 45 18 29 • *E-mail: emilio@empresariales.ulpgc.es

**E-mail: josema@empresariales.ulpgc.es

RESUMEN

El modelo de buenos-malos riesgos considera que la población o colectivo de asegurados de una cartera de seguros tiene dos tipos de riesgos. Un porcentaje de ellos tiene una baja probabilidad de siniestralidad, son los buenos riesgos; mientras que el resto tiene una alta probabilidad de siniestralidad, son los malos riesgos. En esta situación, una única distribución de probabilidad no resulta lo más adecuado para modelar el proceso de siniestralidad, siendo más sugerente una combinación convexa de dos o más distribuciones. Hewitt (1.966) ha utilizado la combinación log-gamma+log-gamma; Hewitt y Lefkowitz (1.979), la combinación gamma+log-gamma y gamma+log-normal; Venter (1.991) la combinación Poisson+Poisson, etc. Sin embargo, no aparecen en la literatura estudios de esta naturaleza considerados desde una perspectiva bayesiana.

En este trabajo se analiza el modelo de los buenos-malos riesgos desde un punto de vista bayesiano y se realiza un análisis de robustez del mismo utilizando una clase de distribuciones a priori dada por contaminaciones de una distribución fija.

Palabras Clave: Teoría de la Credibilidad; Buenos-malos riesgos; Robustez bayesiana; Clase de contaminación.

ABSTRACT

Good and Bad models consider population of an insurances portfolio has two types of risks. There are good risks, which has low claims probabilities and bad risks with high claims probabilities. So we need a convex combination of two or more distributions to model this situation. Hewitt (1.966) has used combination log-gamma+log-gamma; Hewitt and Lefkowitz (1.979) used combination gamma+log-gamma and gamma+log-normal; Venter (1.991) used combination Poisson+Poisson, etc. However, bayesian methodology has not studied this model.

In this work we analyzed the model of the good-bad risks from a Bayesian point of view and we will measure the sensitivity of Bayesian Premium with respect to disturbances in the prior distribution from the risk parameter using a fixed contaminated class.

Key Words: Theory of Credibility; Good and Bad Risks; Bayesian Robustness; Contaminated Class.

Clasificación AMS: 62P05,62F15.

Artículo recibido el 15 de julio de 2000. Aceptado el 14 de diciembre de 2000.

1. Introducción

La teoría de la credibilidad, en palabras de Hickmann (1974), se ocupa del ajuste sistemático de las primas de seguros a medida que se obtiene la experiencia de reclamaciones. Sea una póliza (individuo) caracterizada por un parámetro de riesgo desconocido q perteneciente a una cartera de seguros (colectivo) cuyas reclamaciones vienen representadas por una variable aleatoria X que sigue una distribución de Poisson de parámetro q. En teoría de la credibilidad se asume que el parámetro de riesgo q tiene una distribución a priori $p_0(q)$, generalmente denominada función estructura. Esta distribución modela cómo se distribuye la media de reclamaciones en la cartera de seguros.

La idea central de la teoría de la credibilidad es estimar la prima de riesgo desconocida $\mathrm{P}(q) = E[X|q]$, dado t años de experiencia $x_1, x_2, \ldots, x_t$, la prima del colectivo $\mathrm{P}^*_{p_0} = \int_Q \mathrm{P}(q) p_0(q) dq$ y la prima bayesiana $\mathrm{P}^*_{p_0}(x_1, x_2, \ldots, x_t) = \int_Q \mathrm{P}(q) p_0(q|x_1, \ldots, x_t) dq$ donde $p_0(q|x_1, \ldots, x_t)$ es la distribución a posteriori.

Los modelos frecuentemente utilizados en teoría de la credibilidad asumen que el número de reclamaciones en una póliza de seguros tiene la distribución de Poisson, con una distribución gamma a lo largo del colectivo de pólizas para el parámetro de riesgo. En el modelo de buenos-malos riesgos se supone que la cartera de seguros tiene dos tipos de riesgos, los buenos riesgos con bajas probabilidades de reclamaciones, y los malos riesgos con altas probabilidades de reclamaciones. Si suponemos, siguiendo los modelos habituales de teoría de la credibilidad, una distribución gamma para el parámetro de riesgo en el colectivo, un modelo como el anterior puede representarse como

$$p_0(q) = a_1 \, \mathrm{Ga}(a_1, b_1) + a_2 \, \mathrm{Ga}(a_2, b_2),$$

donde $\mathrm{Ga}(a_i, b_i)$ representa la función de densidad gamma de parámetros $a_i > 0, b_i > 0$ y a_1 y a_2 son números reales no negativos que suman la unidad.

Muchos trabajos se han publicado en teoría de la credibilidad acerca de la aproximación bayesiana (Gómez et al., (2.000); Klugman, (1.992); Young, (1.998); entre otros), ninguno de ellos en el modelo de los buenos-malos riesgos. El problema de interés en este artículo será el análisis bayesiano de los buenos-malos riesgos que se ampliará con un análisis basado en información a priori parcial. El análisis de robustez bayesiano ha recibido una atención considerable en las últimas décadas, ver por ejemplo Berger (1994) para una paronámica de diversas metodologías. En nuestro caso, para realizar un análisis de esta naturaleza supondremos que el actuario es incapaz de especificar una única distribución de probabilidad para el parámetro de riesgo y únicamente es capaz de especificar una clase de distribuciones a priori. Una clase particularmente intuitiva y flexible es la clase de contaminación dada por $G_e = \{p(q) = (1-e)p_0(q) + e\, q(q) : q \in \mathrm{Q}\}$, donde ε mide la incertidumbre sobre la distribución inicial y Q es la clase de contaminaciones posibles. Esta clase ha sido utilizada en un análisis bayesiano general en Berger (1994); Berger y Berliner (1986); Moreno y Cano (1991), Sivaganesan y Berger (1989), entre otros.

El artículo está organizado de la siguiente manera. En la sección 2 se presentan los distintos tipos de primas de seguros, obteniéndose sus expresiones para el principio de prima neta. La sección 3 consta de una descripción del modelo de los buenos-malos riesgos en su versión clásica y bayesiana. En la sección 4 se proporcionan los resultados fundamentales para llevar a cabo un análisis de sensibilidad de la prima bayesiana. Finalmente, las secciones 5 y 6 concluyen con diversas ilustraciones, comentarios finales y posibles extensiones.

2. Prima de riesgo, prima colectiva y prima bayesiana

Sean $X_1, X_2, \ldots, X_t$ variables aleatorias real-valuadas independientes e idénticamente distribuidas. Estas variables representan el número de reclamaciones de un asegurado en los últimos *t* años de vigencia de la póliza o asegurado. Sea la distribución de $X_i,\ i = 1,\ldots,t$ dada por $f(x|q)$, dependiente de un parámetro real $q \in Q \subseteq \Re$, y denominado en términos actuariales parámetro de riesgo.

Un principio de cálculo de prima P asigna a cada riesgo una prima que depende del parámetro *q* que se denomina prima de riesgo. En el supuesto de utilizar el principio de prima neta, la función P viene dada por

$$\mathrm{P}(q) = \int_{\mathrm{X}} x f(x|q)\,dx\,. \tag{1}$$

Obsérvese que esta prima no está expresada en términos monetarios. En la práctica, para calcular la prima monetaria se asume una indemnización fija por siniestro, *c*, de aquí que la prima que la compañía cobraría al asegurado sería $c\mathrm{P}(q)$.

Uno de los principales tópicos en teoría de la credibilidad es precisamente la incertidumbre del parámetro de riesgo (Freifelder, 1974; Gómez et al., 2000; Herzog, 1996; Klugman et al., 1998, entre otros). Si el actuario conoce la distribución del parámetro de riesgo *q* dentro del colectivo de la compañía de seguros a la que pertenezca el asegurado, la prima de riesgo se podrá estimar mediante la denominada prima colectiva, notada por $\mathsf{P}^{*}_{p_0}$, que se define como

$$\mathsf{P}^{*}_{p_0} = \int_{\grave{E}} \mathsf{P}(\boldsymbol{q})\boldsymbol{p}_0(\boldsymbol{q})d\boldsymbol{q}. \tag{2}$$

Si se incorpora la experiencia de siniestralidad dada por el número de reclamaciones del asegurado en los últimos t años, notadas $x_1, x_2, \ldots, x_t$, se puede determinar la prima bayesiana que se define como

$$\mathsf{P}^{*}_{p_0}(\mathrm{m}) \equiv \mathsf{P}^{*}_{p_0}(x_1, x_2, \ldots, x_t) = \int_{\grave{E}} \mathsf{P}(\boldsymbol{q})\boldsymbol{p}_0(\boldsymbol{q}|x_1, x_2, \ldots, x_t)d\boldsymbol{q}, \tag{3}$$

en la que $\mathrm{m} = \sum_{\mathrm{i}=1}^{\mathrm{t}} x_{\mathrm{i}} / t$ es el estadístico minimal suficiente dado por la media muestral y $\boldsymbol{p}_0(\boldsymbol{q}|x_1, x_2, \ldots, x_t)$ la distribución a posteriori, que se calcula

$$\boldsymbol{p}_0(\boldsymbol{q}|x_1, x_2, \ldots, x_t) = \frac{f(x_1, x_2, \ldots, x_t|\boldsymbol{q})\boldsymbol{p}_0(\boldsymbol{q})}{p(x_1, x_2, \ldots, x_t|\boldsymbol{p}_0)},$$

donde $p(x_1, x_2, \ldots, x_n|\boldsymbol{p}_0)$ es la distribución marginal o predictiva.

3. Modelo de los buenos-malos riesgos

El modelo de los buenos-malos riesgos considera que la población o colectivo de asegurados de una cartera de seguros tiene dos tipos de riesgos, los buenos riesgos, con una baja probabilidad de reclamaciones y los malos riesgos, con altas probabilidades de reclamaciones. Un modelo de esta naturaleza puede representarse mediante una mixtura de dos distribuciones. Esto es, el modelo que asumiremos será el siguiente:

$$X_i \sim f(x|\boldsymbol{q}), \ \ i = 1, \ldots, t, \ \ \boldsymbol{q} \in \boldsymbol{Q} \subseteq \Re,$$

$$\boldsymbol{p}_0(\boldsymbol{q}) = \sum_{i=1}^{2} \boldsymbol{a}_i \boldsymbol{p}_i(\boldsymbol{q}), \ \ \sum_{i=1}^{2} \boldsymbol{a}_i = 1,$$

donde X_i es la variable aleatoria que representa el número de reclamaciones del asegurado en el período i, i.i.d. con función de densidad $f(x|\theta)$ dependiente de un parámetro θ. Este parámetro, a su vez, se distribuye en el colectivo como una mixtura de dos distribuciones (distribuciones a priori) $\pi_1(\theta)$ y $\pi_2(\theta)$.

La prima de riesgo es ahora igual que en (1) y la prima colectiva es, de (2)

$$\begin{aligned} \Pi^*_{\pi_0} &= \int_\Theta \Pi(\theta)\pi_0(\theta)d\theta = \int_\Theta \Pi(\theta)\{\alpha_1\pi_1(\theta)+\alpha_2\pi_2(\theta)\}d\theta \\ &= \alpha_1\int_\Theta \Pi(\theta)\pi_1(\theta)d\theta + \alpha_2\int_\Theta \Pi(\theta)\pi_2(\theta)d\theta, \end{aligned}$$

es decir, una suma convexa de la prima colectiva obtenida para cada uno de los riesgos en que se considera dividido el colectivo, los buenos y malos riesgos.

El cálculo de la prima bayesiana requiere obtener primero la distribución a posteriori de θ. Esta viene dada por,

$$\begin{aligned} \pi_0(\theta|\mathrm{m}) &= \frac{f(\mathrm{m}|\theta)\pi_0(\theta)}{\int_\Theta f(\mathrm{m}|\theta)\pi_0(\theta)d\theta} \\ &= \frac{\alpha_1 f(\mathrm{m}|\theta)\pi_1(\theta)+\alpha_2 f(\mathrm{m}|\theta)\pi_2(\theta)}{\alpha_1\int_\Theta f(\mathrm{m}|\theta)\pi_1(\theta)d\theta+\alpha_2\int_\Theta f(\mathrm{m}|\theta)\pi_2(\theta)d\theta} \\ &= \frac{\alpha_1\pi_1(\theta|\mathrm{m})\int_\Theta f(\mathrm{m}|\theta)\pi_1(\theta)d\theta+\alpha_2\pi_2(\theta|\mathrm{m})\int_\Theta f(\mathrm{m}|\theta)\pi_2(\theta)d\theta}{\alpha_1\int_\Theta f(\mathrm{m}|\theta)\pi_1(\theta)d\theta+\alpha_2\int_\Theta f(\mathrm{m}|\theta)\pi_2(\theta)d\theta} \\ &= \alpha_1^*\pi_1(\theta|\mathrm{m})+\alpha_2^*\pi_2(\theta|\mathrm{m}), \end{aligned} \tag{4}$$

donde $\alpha_i^* = \dfrac{\alpha_i\, p(\mathrm{m}|\pi_i)}{\sum_{i=1}^2 \alpha_i\, p(\mathrm{m}|\pi_i)}$, con $\sum_{i=1}^{2}\alpha_i^* = 1.$

ÍNDICE

Ahora la prima bayesiana es, de (3) y (4)

$$\begin{aligned}\Pi^{*}_{\pi_0}(\mathrm{m}) &= \int_{\Theta} \Pi(\theta)\left\{\alpha_1^{*}\pi_1(\theta|\mathrm{m}) + \alpha_2^{*}\pi_2(\theta|\mathrm{m})\right\}d\theta \\ &= \alpha_1^{*}\int_{\Theta}\Pi(\theta)\pi_1(\theta|\mathrm{m})d\theta + \alpha_2^{*}\int_{\Theta}\Pi(\theta)\pi_2(\theta|\mathrm{m})d\theta \\ &= \alpha_1^{*}\Pi^{*}_{\pi_1}(\mathrm{m}) + \alpha_2^{*}\Pi^{*}_{\pi_2}(\mathrm{m}),\end{aligned}$$

es decir, de nuevo, una suma convexa ahora de la prima bayesiana obtenida para cada uno de los riesgos en que se considera dividido el colectivo, los buenos y malos riesgos.

4. Análisis bayesiano con entrada menos rígida

La metodología bayesiana habitual, como hemos comprobado en la sección anterior, se compone de los siguientes pasos:

- La elección de una distribución a priori inicial. En nuestro caso $\pi_0(\theta)$, que viene dada por la suma convexa de dos distribuciones a priori, $\pi_1(\theta)$ y $\pi_2(\theta)$.

- La combinación de la distribución a priori con la información muestral, utilizando el teorema de Bayes, para obtener la distribución a posteriori, que nos permite calcular la prima bayesiana.

Sin embargo, el análisis bayesiano ha recibido numerosas críticas por parte, sobre todo, de los no bayesianos. Éstos frecuentemente argumentan que la especificación en el modelo de la distribución a priori inicial es difícil. Para los bayesianos, por el contrario, supone la conducta adecuada, pero sólo si el investigador es capaz de discriminar de manera muy fina su juicio acerca de diferentes distribuciones de probabilidad. La conclusión es realizar un análisis de robustez o de sensibilidad.

El análisis de robustez bayesiano se ocupa de estudiar la sensibilidad de las respuestas bayesianas (en nuestro caso la prima bayesiana) ante la incertidumbre de las entradas (en nuestro caso la distribución a priori). Nuestra aproximación al problema, pues, consistirá en:

- Especificar una clase de distribuciones a priori compatible con las creencias iniciales del actuario, en vez de una distribución a priori singular; esta clase recoge la incertidumbre en la especificación de la distribución a priori.

- Obtener el rango de variación de la prima bayesiana cuando la distribución a priori varía en la clase de distribuciones a priori especificada.

Es decir, si Γ es la clase de distribuciones a priori especificada, el objetivo es determinar los extremos $inf\left\{\mathsf{P}_p^*(\mathrm{m}); \boldsymbol{p} \in G\right\}$ y el $sup\left\{\mathsf{P}_p^*(\mathrm{m}); \boldsymbol{p} \in G\right\}$, donde m es la observación muestral. Si la diferencia entre estos dos valores es pequeña se dice que el modelo es robusto , en otro caso se habla de falta de robustez.

Una familia que permite realizar un análisis de sensibilidad como el señalado es la clase de contaminación (Berger (1994); Berger y Berliner (1986); Moreno y Cano (1991); Sivaganesan y Berger (1989); entre otros). En ella la distribución a priori del parámetro está dentro de una clase de distribuciones a priori de la forma

$$\Gamma_{\varepsilon,\mathrm{Q}} = \{\pi = (1-\varepsilon)\pi_0 + \varepsilon q : q \in \mathrm{Q}\}, \tag{5}$$

en la que $\boldsymbol{e} \in [0,1]$ determina la cantidad de incertidumbre en la distribución a priori inicial y Q es una clase se posibles contaminaciones. El siguiente teorema muestra la manera de obtener el rango de variación de la prima bayesiana para la clase contaminante Q* = {todas las distribuciones de probabilidad}.

Teorema

$$\inf_{\text{ðÎ Ã}_{\text{åQ}^*}} \mathsf{P}_\pi^*(\mathrm{m}) = \inf_{\theta\in\Theta} \mathrm{R}(\theta),$$

$$\sup_{\pi\in\Gamma_{\varepsilon,\mathrm{Q}^*}} \mathsf{P}_\pi^*(\mathrm{m}) = \sup_{\theta\in\Theta} \mathrm{R}(\theta),$$

donde

$$\mathrm{R}(\boldsymbol{q}) = \frac{(1-\boldsymbol{e})\left[\sum_{i=1}^{2} \boldsymbol{a}_i\, p(\mathrm{m}|\boldsymbol{p}_i)\right]\mathsf{P}_{p_0}^*(\mathrm{m}) + \boldsymbol{e}\,\mathsf{P}(\boldsymbol{q})\,f(\mathrm{m}\boldsymbol{q})}{(1-\boldsymbol{e})\left[\sum_{i=1}^{2} \boldsymbol{a}_i\, p(\mathrm{m}|\boldsymbol{p}_i)\right] + \boldsymbol{e}\, f(\mathrm{m}|\boldsymbol{q})}.$$

Demostración. Para la clase en (5), la distribución predictiva de *m* dada *p* es

$$p(\mathrm{m}|\boldsymbol{p}) = (1-\boldsymbol{e})[\boldsymbol{a}_1 p(\mathrm{m}|\boldsymbol{p}_1) + \boldsymbol{a}_2 p(\mathrm{m}|\boldsymbol{p}_2)] + \boldsymbol{e}\; p(\mathrm{m}|q).$$

Luego, del cálculo de la distribución a posteriori $\boldsymbol{p}(\boldsymbol{q}|\mathrm{m})$ resulta sin muchas complicaciones,

$$\mathsf{P}_p^*(\mathrm{m}) = \frac{(1-\boldsymbol{e})\,p(\mathrm{m}|\boldsymbol{p}_0)\mathsf{P}_{p_0}^*(\mathrm{m}) + \boldsymbol{e}\int \mathsf{P}(\boldsymbol{q})\,f(\mathrm{m}|\boldsymbol{q})q(\boldsymbol{q})d\boldsymbol{q}}{(1-\boldsymbol{e})\,p(\mathrm{m}|\boldsymbol{p}_0) + \boldsymbol{e}\int f(\mathrm{m}|\boldsymbol{q})q(\boldsymbol{q})d\boldsymbol{q}}$$

$$= \frac{(1-\boldsymbol{e})\left[\sum_{i=1}^{2}\boldsymbol{a}_i\,p(\mathrm{m}|\boldsymbol{p}_i)\right]\mathsf{P}_{p_0}^*(\mathrm{m}) + \boldsymbol{e}\int \mathsf{P}(\boldsymbol{q})\,f(\mathrm{m}|\boldsymbol{q})q(\boldsymbol{q})d\boldsymbol{q}}{(1-\boldsymbol{e})\left[\sum_{i=1}^{2}\boldsymbol{a}_i\,p(\mathrm{m}|\boldsymbol{p}_i)\right] + \boldsymbol{e}\int f(\mathrm{m}|\boldsymbol{q})q(\boldsymbol{q})d\boldsymbol{q}}.$$

Ahora, $inf\left\{\mathsf{P}_p^*(\mathrm{m}); \boldsymbol{p}\in G_{e,\mathrm{Q}^*}\right\}$ y $sup\left\{\mathsf{P}_p^*(\mathrm{m}); \boldsymbol{p}\in G_{e,\mathrm{Q}^*}\right\}$ se obtienen de aplicar el lema A.1. en Sivaganesan y Berger (1989).

Para el modelo que nos ocupa, en que X tiene la distribución de Poisson de parámetro q y q sigue en el colectivo una distribución $\boldsymbol{p}_0(\boldsymbol{q}) = \boldsymbol{a}_1\,\mathrm{Ga}(a_1, b_1) + \boldsymbol{a}_2\,\mathrm{Ga}(a_2, b_2)$, el rango de variación de la prima bayesiana se obtiene aplicando el siguiente corolario, que es consecuencia inmediata del teorema anterior.

Corolario.

Si X_i sigue una distribución de Poisson de parámetro q y $\boldsymbol{q} \sim \boldsymbol{p}_0(\boldsymbol{q}) = \sum_{i=1}^{2}\boldsymbol{a}_i\,\mathrm{Ga}(a_i, b_i)$ se tiene:

$$\inf_{p\in G_{e,Q^*}} \mathsf{P}_p^*(\mathrm{m}) = \inf_{q\in Q}\frac{\mathrm{R}_1\mathsf{P}_{p_0}^*(\mathrm{m}) + \mathrm{R}_2(\boldsymbol{q})}{\mathrm{R}_1 + \mathrm{R}_3(\boldsymbol{q})},$$

$$\sup_{p\in G_{e,Q^*}} \mathsf{P}_p^*(\mathrm{m}) = \sup_{q\in Q}\frac{\mathrm{R}_1\mathsf{P}_{p_0}^*(\mathrm{m}) + \mathrm{R}_2(\boldsymbol{q})}{\mathrm{R}_1 + \mathrm{R}_3(\boldsymbol{q})},$$

donde $\mathrm{R}_1 = (1-\boldsymbol{e})\sum_{i=1}^{2}\boldsymbol{a}_i\frac{a_i^{b_i}}{(b_i-1)!}\frac{(b_i+t\mathrm{m}-1)!}{(a_i+t)^{b_i+t\mathrm{m}}}$, $\mathrm{R}_2(\boldsymbol{q}) = \boldsymbol{e}\,e^{-tq}\boldsymbol{q}^{t\mathrm{m}+1}$ y $\mathrm{R}_3(\boldsymbol{q}) = \frac{\mathrm{R}_2(\boldsymbol{q})}{\boldsymbol{q}}$.

Como medida de la robustez, Sivaganesan (1991) propone el factor de sensibilidad relativa RS, que adaptado a nuestro modelo viene dado por

$$\mathrm{RS} = \frac{sup\left\{\mathrm{P}_{p}^{*}(\mathrm{m}); \boldsymbol{p} \in G_{e,Q^{*}}\right\} - inf\left\{\mathrm{P}_{p}^{*}(\mathrm{m}); \boldsymbol{p} \in G_{e,Q^{*}}\right\}}{2\mathrm{P}_{p_0}^{*}(\mathrm{m})} \times 100\%, ,$$

y que indica el porcentaje de variación de la prima bayesiana calculada para $\boldsymbol{p} \in G_{e,Q^{*}}$ alrededor de la calculada para $\boldsymbol{p}_0$.

5. Ilustración numérica

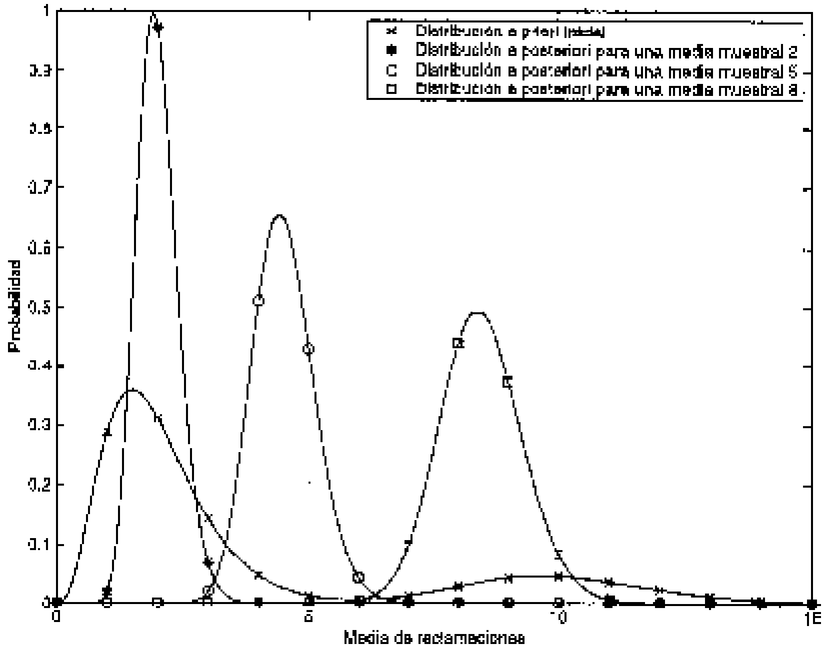
En esta sección ilustraremos las ideas expuestas anteriormente con un ejemplo numérico. Supongamos que el número de reclamaciones de una póliza de seguros está representado por una variable aleatoria X que se distribuye según una Poisson de parámetro q.

El actuario está convencido de que el número medio de reclamaciones en el colectivo, la cartera de seguros que incluye a la póliza anterior, es tal que las reclamaciones superiores a 4 son muchísimo menos probables que reclamaciones por debajo de dicha cantidad. Una distribución de la media de reclamaciones en el colectivo que verifique dicha característica puede venir dada por

$$\boldsymbol{p}_0(\boldsymbol{q}) = 0.8\,\mathrm{Ga}(2,4) + 0.2\,\mathrm{Ga}(3,30),$$

cuya gráfica aparece en la figura 1, así como la distribución a posteriori de la misma para las medias muestrales observadas 2, 5 y 8.

Figura 1. Distribuciones a priori y a posteriori para las distintas medias muestrales.



Tomaremos como valor de *c*, la cantidad media de indemnización por siniestro, el valor de 100 unidades monetarias (u.m.).

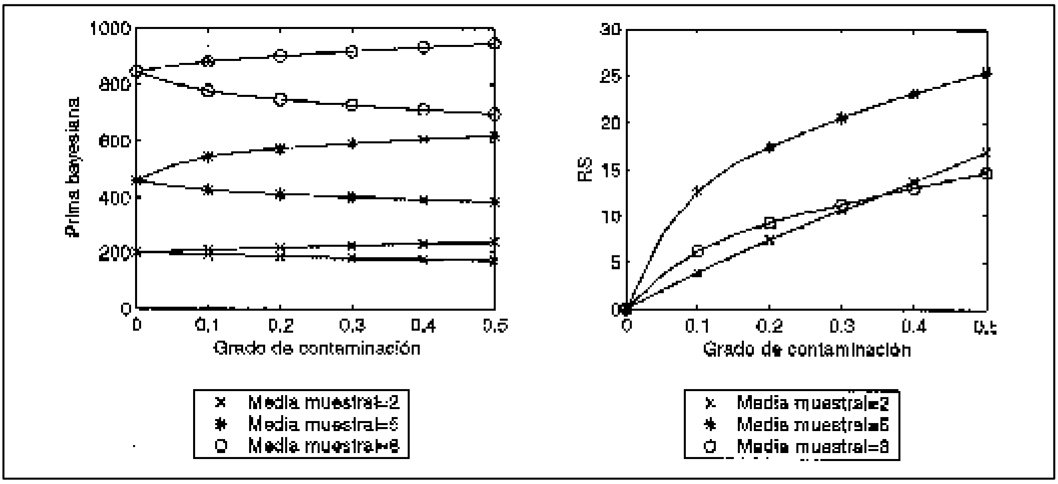
Antes de obtener la información muestral, la prima que cobraría la compañía aseguradora a los clientes de esta póliza sería la prima neta colectiva, cuyo valor es de 360 u.m. Después de observar el número de reclamaciones en los últimos *t*=10 años se obtienen las primas bayesianas que aparecen en la tabla 1 para las medias muestrales 2,5 y 8. Obsérvese que $\mathrm{P}^*_{p_0}(\mathrm{m})$ está más próxima a $\mathrm{P}^*_{p_i}(\mathrm{m})$ para el mayor a^*_i , como era previsible.

Tabla 1. Prima colectiva y bayesiana para diferentes medias muestrales

m	$\mathrm{P}^*_{p_1}$	$\mathrm{P}^*_{p_2}$	$\mathrm{P}^*_{p_0}=\sum_{i=1}^{2}a_i\mathrm{P}^*_{p_i}$	$\mathrm{P}^*_{p_1}(\mathrm{m})$	$\mathrm{P}^*_{p_2}(\mathrm{m})$	a^*_1	a^*_2	$\mathrm{P}^*_{p_0}(\mathrm{m})=\sum_{i=1}^{2}a^*_i\mathrm{P}^*_{p_i}(\mathrm{m})$
2	200	1000	360	200	384.61	0.999	0.111	200.000
5	200	1000	360	450	615.38	0.938	0.062	460.250
8	200	1000	360	700	846.153	0.01	0.99	844.771

En la figura 2 se muestra el rango de variación de la prima bayesiana, así como los valores del factor RS para las distintas medias muestrales. Hemos tomado como grado de contaminación valores desde 0.1 hasta 0.5 con pasos de 0.1 unidades. Dicho factor toma los mayores valores para m=5, seguido de m =8 y de m =2. Cierta falta de robustez conduce a pensar en la posibilidad de incorporar al modelo propiedades de las distribuciones a priori que permitiesen reducir el rango de variación de la prima bayesiana, como pudiera ser la bimodalidad.

Figura 2. Rangos de variación de las primas bayesianas monetarias en función del grado de contaminación



Comparando las figuras 1 y 2 se observa como la mayor falta de robustez (caso m=5) coincide con la mayor distancia entre la distribución a priori inicial y la distribución a posteriori para este caso. La menor distancia ocurre para el caso m=2, que es el que presenta mayor robustez. Esto es así para valores del grado de contaminación de hasta el 40%. A partir de aquí el caso m=8 es más robusto. La falta de robustez del caso m=5 puede interpretarse considerando que con la distribución a priori inicial elegida, las reclamaciones más probables son 2 y 10, luego una media de reclamación de valor 5 es muy poco frecuente, y de ahí la carencia de robustez manifiesta en este caso.

6. Conclusiones finales y extensiones

En este artículo se ha ilustrado el uso de la metodología bayesiana clásica y robusta para analizar la prima bayesiana en un modelo de los buenos-malos riesgos en seguros.

Para calcular la prima bayesiana se requiere de la especificación de una distribución a priori inicial para el parámetro de riesgo por parte del actuario. La incorporación de esta distribución a priori en el modelo ha sido objeto de crítica por parte de los no bayesianos, de ahí que se trate de salvar esta situación utilizando una clase de posibles distribuciones a priori. Este es el fundamento del análisis bayesiano robusto, cuya metodología proporciona un rango de variación de la prima bayesiana, asequible para que la compañía de seguros pueda tarificar un riesgo.

La clase de todas las distribuciones de probabilidad utilizada en este artículo puede contener, sin duda, numerosas distribuciones en nada compatibles con las creencias iniciales del actuario, por lo que el factor RS puede estar tomando valores mayores de los que debiera. Una modificación sustancial en el teorema 1 puede llevarse a cabo si el actuario considera conveniente utilizar únicamente aquellas distribuciones a priori que, como la inicial, sean bimodales. Esto llevaría consigo una reducción en los valores del factor RS. Esta modificación es posible dada la flexibilidad de la clase de ε-contaminación, que permite modelar las creencias iniciales del actuario sin que surjan demasiadas complicaciones de índole matemática.

Por último, digamos que la metodología seguida en este artículo es susceptible de ser aplicada a otros principios de cálculo de primas distintos del de prima neta presentado aquí (una excelente revisión de los mismo puede consultarse en Heilmann, 1989).

* Los autores agradecen muy sinceramente los comentarios de un evaluador anónimo que han ayudado a mejorar este trabajo.

Bibliografía

BERGER, J. (1994): *An Overview of Robust Bayesian Analysis*. Test, 3, 5 -120.

BERGER, J. y BERLINER, L. (1986): *Robust Bayes and empirical bayes analysis with ε-contaminated priors*. The Annal of Statistics, 14, 2, 461-486.

FREIFELDER, L. (1974): *Statistical Decision Theory and Credibility Theory and Procedures*. Proceedings of the Berkley Actuarial Research Conference on Credibility. Credibility Theory and Applications. Edited by P.M. Kahn. Academic Press,71-88. University of California, Berkley.

GÓMEZ, E.; HERNÁNDEZ, A. y VÁZQUEZ, F. (2000): *Robust Bayesian Premium Principles in Actuarial Science*. Journal of the Royal Statistical Society. Series D. The Statistician, 49, 2, 241-252.

HEILMANN, W. (1989): *Decision Theoretic Foundations of Credibility Theory*. Insurance: Mathematics and Economics, 8, 77-95.

HERZOG, T. (1996): *Introduction to Credibility Theory*. ACTEX Publications (Second Edition).

HEWITT, C. (1966): *Distribution by Size of Risk*. Proceedings of the Casualty Actuarial Society, LIII,106-117 (with discussion).

HEWITT, C. y LEFKOWITZ, B. (1979): *Methods for Fitting Distributions to Insurance Loss Data*. Proceedings of the Casualty Actuarial Society, LXVI, 139-160.

HICKMANN, J. (1974): *Introduction and Historical Overview of Credibility*. Credibility Theory and Applications. Edited by P.M. Kahn. Academic Press,181-193.

KLUGMAN, S. (1992): *Bayesian Statistics in Actuarial Science*. Kluwer Academic Publisher.

KLUGMAN, S., PANJER, H. y WILLMOT, G. (1998): *Loss Model from Data to Decisions*. Willey, New York.

MORENO, E. y CANO, J. (1991): *Robust Bayesian Analysis for ε-contaminations Partially Known*. Journal of the Royal Statistical. Series B, 53,143-155.

SIVAGANESAN, S. (1991): *Sensitivity of some posterior summaries when the prior is unimodal with specified quantiles*. The Canadian Journal of Statistics,19,1,57-65.

SIVAGANESAN, S. y BERGER, J. (1989): *Ranges of Posterior Measures for Priors with Unimodal Contaminations*. The Annals of Statistics, 17, 2, 868-889.

VENTER, G. (1991): *Effects of Variations from Gamma-Poisson Assumptions*. Proceedings of the Casualty Actuarial Society, LXXVIII, 41-56.

YOUNG, V. (1998): *Robust Bayesian Credibility Using Semiparametric Models*. Astin Bulletin, 28,1,187-203.