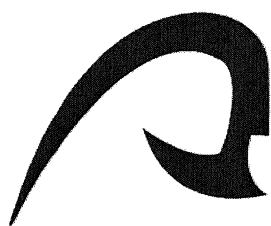


TESIS DOCTORAL

ESTADÍSTICA BAYESIANA EN LA CIENCIA ACTUARIAL CON ÉNFASIS EN ROBUSTEZ LOCAL E INFERENCIA

Enrique Calderín Ojeda

2007



UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
MÉTODOS CUANTITATIVOS EN ECONOMÍA Y GESTIÓN

54 1495 P.R.

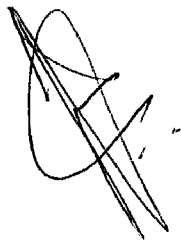
BIBLIOTECA UNIVERSITARIA
LAS PALMAS DE G. CANARIA
N.º Documento <u>502978</u>
N.º Copia <u>868379</u>

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
DEPARTAMENTO DE MÉTODOS CUANTITATIVOS
EN ECONOMÍA Y GESTIÓN

Los abajo firmantes certifican que han leído y recomendado a la Facultad de Ciencias Económicas y Empresariales para su aceptación la tesis titulada Estadística Bayesiana en la Ciencia Actuarial con Énfasis en Robustez Local e Inferencia, realizada por Enrique Calderín Ojeda en cumplimiento de los requerimientos para el grado de doctor por la Universidad de Las Palmas de Gran Canaria.

2007

Directores:



Emilio Gómez Déniz



Ignacio Cabrera Ortega

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Fecha: Febrero 2006

Autor: Enrique Calderín Ojeda.

Título: Estadística Bayesiana en la Ciencia Actuarial con Énfasis en Robustez Local e Inferencia

Dpto: Métodos Cuantitativos en Economía y Gestión

Grado: Doctor

Firma del Autor

Prólogo

El estudio realizado en esta Memoria se encuentra enmarcada dentro de la estadística bayesiana robusta aplicada a las Ciencias Actuariales. Esta disciplina, se ha desarrollado para proporcionar las herramientas necesarias en el estudio de ciertas actividades económicas que realizan las compañías de seguros. Desde el prisma actuarial, esta Memoria enfoca dos elementos importantes en el campo de los seguros generales. Por una parte, la búsqueda de nuevas distribuciones de probabilidad discreta con las que obtener buenos ajustes para modelizar datos de carteras de seguros. Por otra parte, nos centraremos en estudiar si bajo determinados modelos y métodos de cálculo, la primas de seguros obtenidas son o no robustas.

La metodología utilizada en esta Memoria para determinar la robustez de las primas de seguros puede considerarse novedosa desde la óptica actuarial. Tradicionalmente el estudio de robustez ha sido estudiado desde una perspectiva global. Esta Memoria se centra en el análisis local. Para llevar a cabo este estudio, procederemos a perturbar una de las entradas del modelo (la distribución a priori) y observar cómo se comporta la salida (magnitud a posteriori).

La Memoria está estructurada en cinco capítulos, de los que cabe destacar de cada uno de ellos los siguientes aspectos. En el capítulo 1 se enfatiza en los elementos básicos de la metodología actuarial y bayesiana. Se comentan, asimismo, los objetivos de la Memoria. En el capítulo 2 se analizan los elementos fundamentales del análisis bayesiano robusto desde una perspectiva local que se precisan para abordar el análisis local de las primas de seguros. En el capítulo 3 estudiaremos el problema de la inferencia a través de una nueva distribución de probabilidad discreta en su versión univariada y multivariada. En el capítulo 4 se desarrollarán diversas aplicaciones numéricas teniendo en cuenta la metodología desarrollada en los capítulos anteriores. Por último, en el capítulo 5 presentamos las conclusiones más relevantes y algunas líneas abiertas que consideramos merecerían ser objeto de investigación en el futuro. Además incluimos un Apéndice en el que realizamos una descripción detallada del *Método Scoring* que ha sido utilizado en las aplicaciones numéricas de esta Memoria.

X

A mi familia

Agradecimientos

Esta Memoria ha sido el resultado de cuatro años de trabajo. Huelga señalar que durante este tiempo, muchas personas me han apoyado y a todas ellas quiero expresar mi gratitud.

En primer lugar, quiero expresar mi más sincero agradecimiento a los directores de esta Tesis Doctoral, los doctores D. Emilio Gómez Déniz del Departamento de Métodos Cuantitativos en Economía y Gestión (Universidad de Las Palmas de Gran Canaria) y D. Ignacio Cabrera Ortega del Departamento de Matemáticas (Universidad de Las Palmas de Gran Canaria) por la confianza depositada en mí para la realización de esta Memoria. Sus comentarios, aportaciones, sugerencias y correcciones han sido enriquecedores para poder completar este trabajo.

Agradezco la colaboración del profesor Dr. D. José María Sarabia del Departamento de Economía (Universidad de Cantabria) por su ayuda en las demostraciones de algunos resultados en el capítulo de inferencia.

Finalmente debo hacer una especial mención a mi familia. Debo agradecer a mis padres todo el apoyo y la ayuda recibida durante todos estos años. A tí, Lidia, por tu comprensión, cariño y por la inestimable cantidad de tiempo invertido en las correcciones.

Índice general

Prólogo	ix
1. Introducción	1
1.1. Elementos de estadística actuarial	3
1.2. Introducción al análisis bayesiano robusto local	8
1.3. Objetivos de la Memoria	11
2. Elementos del análisis bayesiano robusto	15
2.1. Introducción	15
2.2. Preliminares	17
2.3. Robustez local para cociente de magnitudes a posteriori	17
2.3.1. Incorporación de la propiedad de unimodalidad	24
2.4. Robustez local bajo pérdida cuadrática ponderada	25
2.4.1. El caso particular de la norma del supremo esencial	28
2.5. Conclusiones	30
3. Inferencia	33
3.1. Modelo binomial negativa–inversa gaussiana	33
3.1.1. Resultados básicos	35
3.1.2. Relaciones recursivas para las probabilidades	40
3.1.3. Distribución a posteriori	43
3.1.4. Versión multivariada	44
3.1.5. Regresión	46
3.1.6. Estimación de parámetros	47
3.2. Modelo binomial–inversa gaussiana	54
3.2.1. Versión multivariada del modelo <i>BIG</i>	54
3.2.2. Relación recursiva para el cálculo de probabilidades	57
3.2.3. Estimación de parámetros	60
3.3. Conclusiones	62

4. Ilustraciones numéricas	63
4.1. Robustez local. Modelo Poisson-gamma	63
4.2. Robustez local de la PBM	66
4.3. Aplicaciones de la distribución <i>BNIG</i>	70
4.3.1. Ajustes para el caso univariado	70
4.3.2. Ajuste para el caso bivariado	73
4.3.3. Regresión	74
4.3.4. Cálculo de PBM bajo el modelo <i>BNIG</i>	75
4.3.5. Análisis de robustez	78
4.4. Conclusiones	78
5. Conclusiones y líneas abiertas	81
5.1. Conclusiones	81
5.2. Sobre la distribución perturbadora	84
5.3. Sobre el comportamiento asintótico de las normas	84
5.4. Sobre los principios de cálculo de primas	85
5.5. Sobre la incorporación de condiciones de momentos	85
5.6. Realizar mezclas con diferentes distribuciones	85
Apéndice	87
Bibliografía	89

Índice de figuras

3.1. Algunos ejemplos de la función de masa de probabilidad de la distribución \mathcal{BNTG} para diferentes valores de r , ψ y μ	41
4.1. Primas y normas para $p = \infty$, $q = 1$ y supremo esencial	66
4.2. Absentismo predicho en 1948 y diferencias entre nuestro modelo y el de Arbous y Sichel (1954)	76

Índice de cuadros

4.1. Frecuencias observadas (Lemaire, 1979)	65
4.2. Frecuencias observadas (Bühlmann, 1970)	67
4.3. PBM bajo los principios de prima neta (en negrita) y de varianza y el modelo \mathcal{PG}	67
4.4. Normas de la PBM bajo el principio de prima neta y el modelo \mathcal{PG}	68
4.5. Normas de la PBM bajo el principio de varianza y el modelo \mathcal{PG}	69
4.6. Siniestros observados y ajustados	71
4.7. Siniestros observados y ajustados	72
4.8. Frecuencias observadas y esperadas de los datos de absentismo (Método de los momentos)	74
4.9. Frecuencias observadas y esperadas de los datos de absentismo (Método de máxima verosimilitud)	75
4.10. PBM bajo los principios de prima neta (en negrita) y de varianza bajo el modelo \mathcal{BNIG}	77
4.11. PBM bajo los principios de prima neta (en negrita) y de varianza bajo el modelo \mathcal{PG}	77
4.12. PBM bajo los principios de prima neta (en negrita) y de varianza bajo el modelo \mathcal{PIG}	78
4.13. Porcentaje de incremento de la prima cuando el asegurado pasa de $x = 0$ a $x = 1$ bajo el principio de prima neta (en negrita) y el de varianza	78
4.14. Normas de la PBM bajo principio de prima neta del modelo \mathcal{BNIG}	79
4.15. Normas de la PBM bajo principio de varianza del modelo \mathcal{BNIG}	79

Capítulo 1

Introducción

La presente Memoria de Tesis Doctoral se desarrolla en el marco de la estadística bayesiana robusta aplicada a las Ciencias Actuariales. Esta disciplina, que ha tenido un auge extraordinario en las últimas décadas, está destinada a proporcionar las herramientas necesarias para el estudio de ciertas actividades económicas que llevan a cabo las compañías de seguros. Estas actividades incluyen, entre otras, los seguros no vida o generales y los seguros vida.

Los primeros cubren la práctica totalidad de los seguros que habitualmente se contratan: seguros de automóviles, contra incendios, hogar, etc. El término riesgo aparece aquí de manera natural, identificándose éste con una variable aleatoria que nos da generalmente el número de reclamaciones o siniestros en que incurre un asegurado en particular. En los seguros vida, por otro lado, el asegurado se compromete a hacer uno o más pagos de primas a la compañía aseguradora, comprometiéndose a su vez ésta a pagar a la muerte de aquél una suma fija al o a los beneficiarios designados por el asegurado.

Aunque los seguros vida han experimentado una gran evolución en los últimos años, su tratamiento matemático no es comparable al de los seguros generales. Así, el complejo entramado matemático en el que se desenvuelven estos últimos los ha convertido en una disciplina científica que cuenta actualmente con sus propios medios de difusión, destacando las revistas *Insurance: Mathematics and Economics*, *Scandinavian Actuarial Journal*, *Astin Bulletin*, *Journal of Risk and Insurance*, amén de numerosos Congresos anuales también específicos en la materia.

Desde el punto de vista actuarial la Memoria tratará de tocar dos elementos importantes que aparecen en el terreno de los seguros generales. Por un lado, el problema de buscar distribuciones de probabilidad discretas (univariadas y multivariadas) con las que se obtengan buenos ajustes de distribuciones empíricas de frecuencias relacionadas con el número de siniestros en que incurre un asegurado. Por otro lado, nos centraremos en el precio para el seguro vendido por la aseguradora, esto es, la prima de seguro, para tratar de averiguar si, bajo determinados

modelos y métodos de cálculo de prima asumidos, aquéllas son o no robustas. Y es aquí donde entra en juego la estadística bayesiana robusta.

A ninguna persona que haya trabajado en las Ciencias o Estadísticas Actuariales se le escapa que la metodología bayesiana es una herramienta necesaria e imprescindible en determinados problemas actuariales, sobre todo los relacionados con el cálculo de primas de seguros. Pensemos, para poner un ejemplo, en el seguro de automóviles. Las compañías de seguros agrupan las pólizas referentes a un mismo riesgo con una serie de características comunes en un colectivo o cartera, al cual le corresponde como tal una determinada prima colectiva. Pero, a su vez, cada póliza tiene un conjunto de características específicas que la diferencia de las demás pólizas o asegurados. Estas características, en la mayoría de las ocasiones, son inobservables o difíciles de cuantificar, aunque se deben tener en cuenta a la hora de calcular las primas de cada asegurado. La estadística actuarial aborda este problema asignando una distribución de probabilidad a esas características inobservables, la distribución estructura o distribución a priori. Ahora, utilizando la información muestral que se disponga sobre el asegurado y la distribución a priori, el teorema de Bayes permite calcular la distribución a posteriori a partir de la cual se obtiene la prima a cobrar al asegurado. Bajo determinados modelos y métodos de cálculo de primas se obtienen sugerentes fórmulas que vienen dadas como una suma convexa de la prima colectiva y la media de reclamaciones del asegurado. Son las denominadas fórmulas de credibilidad que han generado un extenso campo de investigación en la estadística actuarial denominado teoría de la credibilidad. Obsérvese, como señalaremos más adelante, que las primas obtenidas mediante esta manera podrán actualizarse a medida que se incorpora la experiencia de siniestralidad del asegurado. Esta es la forma de proceder en los denominados sistemas de tarificación *bonus-malus*. El asegurado paga una prima inicial al entrar en la compañía de seguros, prima que se verá penalizada o bonificada dependiendo de si el asegurado ha experimentado o no reclamación.

Una de las críticas a las que se ha visto sometida la metodología bayesiana recae precisamente en la incorporación de la distribución a priori. Aunque no es esta la única entrada del análisis bayesiano (las otras, como veremos, son el modelo y la función de pérdida) nos ocuparemos también del problema de sensibilidad o robustez de la salida del proceso, esto es de la prima de seguro.

La estadística bayesiana robusta, que tuvo un auge espectacular en los años de las décadas de los 80 y 90 del siglo pasado, se ocupa del estudio de variación de la magnitud a posteriori de interés cuando variamos alguna o algunas de las entradas del sistema. Este estudio puede llevarse a cabo desde dos puntos de vista, global y local. En ambos, el problema consiste en estudiar la variación de la magnitud

a posteriori de interés, utilizando clases de distribuciones de probabilidad en el primer caso, mientras que variando infinitesimalmente la distribución a priori en el segundo.

La Memoria está dividida en cinco capítulos, de los que cabe destacar en cada uno de ellos los siguientes aspectos. En el presente capítulo 1 se hace énfasis en los elementos básicos de la metodología actuarial y bayesiana, así como en los objetivos que se pretenden alcanzar con esta Memoria. En el capítulo 2 se exponen los elementos fundamentales del análisis bayesiano robusto desde un punto de vista local que se precisan para abordar el estudio de sensibilidad local de las primas de seguros. En el capítulo 3 nos ocupamos del problema de la inferencia a través de una nueva distribución de probabilidad discreta en su versión univariada y multivariada. Diversas aplicaciones numéricas desarrolladas teniendo en cuenta la metodología desarrollada en los capítulos precedentes se mostrarán en el capítulo 4. Finalmente, en el capítulo 5 presentamos las conclusiones más relevantes y algunas líneas abiertas que consideramos merecerían ser objeto de investigación en el futuro.

1.1. Elementos de estadística actuarial

En esta sección introduciremos los elementos fundamentales de estadística actuarial que se utilizarán en esta Memoria. Empezamos por establecer el término riesgo, habitual en los escenarios actuariales. Por riesgo entenderemos sencillamente una variable aleatoria, que podrá estar ligada al número de reclamaciones o siniestros, a la cuantía de las mismas o a la cantidad total reclamada, $X = \sum_{i=1}^N X_i$, donde N es la variable aleatoria asociada al número de reclamaciones y X_i la variable aleatoria asociada a la cuantía del i -ésimo siniestro o reclamación. De ahora en adelante, y hasta que no se especifique lo contrario, nos referiremos a un riesgo X indistintamente a la variable aleatoria asociada al número, la cuantía o la cantidad total reclamada.

La prima es el precio para el seguro vendido por la compañía aseguradora. En la práctica, incluye el coste que puede suponer para la compañía aseguradora los siniestros declarados más el margen de beneficio, aunque aquí nos centraremos siempre en el primer elemento y nunca haremos mención al segundo, que tiene un carácter menos estadístico, y por así decirlo, más contable.

Si denotamos mediante X la variable aleatoria representativa del riesgo, un principio de cálculo de prima se define como sigue (Gerber, 1979).

Definición 1.1 *Un principio de cálculo de prima es un funcional H que asigna*

a un riesgo X un número real $P = H(X)$, que se denomina prima asignada al riesgo X .

Consideremos ahora una función de pérdida $L : \mathbb{R}^2 \rightarrow \mathbb{R}$ que atribuya a $(x, P) \in \mathbb{R}^2$ la pérdida $L(x, P)$ sostenida por un decisor que toma la acción P y se encuentra con el resultado x de algún experimento aleatorio. La prima de riesgo se define (Heilmann, 1989) de la siguiente manera.

Definición 1.2 Dado un riesgo X con función de distribución $F(x)$ y una función de pérdida $L : \mathbb{R}^2 \rightarrow \mathbb{R}$, la prima de riesgo es el valor de P que minimiza la pérdida esperada

$$\int_{\mathbb{X}} L(x, P) dF(x) = \mathbb{E}[L(X, P)] \quad (1.1)$$

donde x es el resultado del experimento aleatorio y P la prima cobrada por tomar x .

En la práctica, la distribución $F(x)$ depende de un parámetro o vector de parámetros desconocido θ , y entonces $F(x) \equiv F(x|\theta)$. Ahora la prima P obtenida de (1.1) será obviamente desconocida y dependiente del vector de parámetros θ por lo que la notaremos mediante $P(\theta)$.

Si notamos por $\pi_0(\theta)$ la distribución a priori de θ , se define la prima colectiva o a priori como sigue (Heilmann, 1989).

Definición 1.3 Dado un riesgo X con distribución $F(x|\theta)$ siendo θ un vector de parámetros desconocido con distribución a priori $\pi_0(\theta)$ y una función de pérdida $L : \mathbb{R}^2 \rightarrow \mathbb{R}$, la prima colectiva o a priori es el valor de PC que minimiza la pérdida esperada

$$\int_{\Theta} L(P(\theta), PC) \pi_0(\theta) d\theta, \quad (1.2)$$

con $P(\theta)$ la verdadera prima individual obtenida de (1.1).

Esta prima colectiva representa hasta ahora la mejor y única decisión que estima la prima de riesgo, desconocida. En la práctica se corresponde con la prima que teóricamente debería pagar un asegurado que pertenezca a un determinado colectivo o cartera y del que no se posea información individual. Si se dispone de esta información, información muestral acerca de su propia experiencia de reclamaciones, podemos calcular, vía teorema de Bayes la distribución a posteriori del parámetro.

Así, si $\pi_0(\theta)$ es la densidad a priori del parámetro θ y \mathbf{x} es la observación muestral de una población cuya distribución depende de \mathbf{x} , denotaremos la verosimilitud por $f(\mathbf{x}|\theta)$ y por medio del teorema de Bayes podremos obtener la distribución a posteriori de la siguiente manera:

$$\pi_0(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi_0(\theta)}{\int_{\Theta} f(\mathbf{x}|\theta)\pi_0(\theta)} \propto f(\mathbf{x}|\theta)\pi_0(\theta),$$

siendo $p(\mathbf{x}|\pi_0) = \int_{\Theta} f(\mathbf{x}|\theta)\pi_0(\theta)$ la distribución predictiva.

Ahora la distribución a posteriori puede utilizarse para calcular lo que Heilmann (1989) denomina prima Bayes, que se define como sigue.

Definición 1.4 Dado un riesgo X con distribución $F(x|\theta)$ siendo θ un vector de parámetros desconocido con distribución a priori $\pi_0(\theta)$, una función de pérdida $L: \mathbb{R}^2 \rightarrow \mathbb{R}$, y una muestra \mathbf{x} , la prima a posteriori es el valor PB que minimiza

$$\int_{\Theta} L(P(\theta), PB)\pi_0(\theta|\mathbf{x})d\theta, \quad (1.3)$$

siendo $\pi_0(\theta|\mathbf{x})$ la distribución a posteriori de θ dada la muestra y $P(\theta)$ la prima de riesgo dada en (1.1).

De ahora en adelante consideraremos que la distribución del riesgo X depende de un solo parámetro desconocido, y por tanto escribiremos $f(x|\theta)$ y $P(\theta)$ para denotar a la función de densidad de X y la prima de riesgo, respectivamente.

En cuanto a la función de pérdida se refiere, es usual (véase Heilmann (1989); Gómez et al. (2000, 2002, 2003); Makov (1994); entre otros) considerar la función de pérdida cuadrática ponderada, dada por

$$L(x, P) = m(x)(n(x) - P)^2, \quad (1.4)$$

donde $m(x)$ y $n(x)$ son funciones apropiadas cuyas esperanzas bajo f existen.

Derivando con respecto a P la expresión (1.1) cuando utilizamos la función de pérdida (1.4) resulta:

$$P(\theta) = \frac{\int m(x)n(x)f(x|\theta)dx}{\int m(x)f(x|\theta)dx} = \frac{\mathbb{E}_{f(x|\theta)}[m(X)n(X)]}{\mathbb{E}_{f(x|\theta)}[m(X)]}. \quad (1.5)$$

Las siguientes expresiones son consecuencia directa de aplicar el resultado (1.5) a funciones particulares $m(x)$ y $n(x)$.

1. Si consideramos la función de pérdida cuadrática, dada por

$$L(x, P) = (x - P)^2,$$

resulta,

$$P(\theta) = E_{f(x|\theta)}(X), \quad (1.6)$$

denominado principio de prima neta o de *equivalencia*.

2. Si consideramos la función de pérdida cuadrática ponderada, con peso $m(x) = e^{\alpha x}$, dada por

$$L(x, P) = e^{\alpha x} (x - P)^2,$$

con $\alpha > 0$, entonces

$$P(\theta) = \frac{E_{f(x|\theta)} (X e^{\alpha X})}{E_{f(x|\theta)} (e^{\alpha X})}, \quad (1.7)$$

denominado principio Esscher.

El parámetro $\alpha > 0$ de (1.7), se denomina constante de aversión al riesgo o medida de Arrow-Pratt.

3. Si consideramos la función de pérdida cuadrática ponderada, con peso $m(x) = x$, dada por

$$L(x, P) = x(x - P)^2,$$

entonces:

$$P(\theta) = \frac{E_{f(x|\theta)} (X^2)}{E_{f(x|\theta)} (X)}, \quad (1.8)$$

denominado principio de varianza.

Utilizando ahora (1.6), (1.7) y (1.8) y (1.2), la prima colectiva para los principios de prima neta, Esscher y de varianza vienen dadas por

$$PC = \int_{\Theta} \left[\int_{\mathcal{X}} x f(x|\theta) dx \right] \pi(\theta) d\theta = E_{\pi(\theta)} [E_{f(x|\theta)} (X|\theta)], \quad (1.9)$$

$$PC = \frac{\int_{\Theta} P(\theta) e^{\alpha P(\theta)} \pi(\theta) d\theta}{\int_{\Theta} e^{\alpha P(\theta)} \pi(\theta) d\theta} = \frac{E_{\pi(\theta)} [P(\theta) e^{\alpha P(\theta)}]}{E_{\pi(\theta)} [e^{\alpha P(\theta)}]}, \quad (1.10)$$

$$PC = \frac{\int_{\Theta} P(\theta)^2 \pi(\theta) d\theta}{\int_{\Theta} P(\theta) \pi(\theta) d\theta} = \frac{E_{\pi(\theta)} [P(\theta)^2]}{E_{\pi(\theta)} [P(\theta)]}, \quad (1.11)$$

respectivamente.

Finalmente, la prima Bayes para los principios de cálculo de prima considerados se obtienen sin más que intercambiar en (1.9), (1.10) y (1.11) la distribución a priori $\pi_0(\theta)$ por la distribución a posteriori $\pi_0(\theta|\mathbf{x})$. De ahí que resulte conveniente trabajar con distribuciones a priori conjugadas, esto es, distribuciones tales que para una verosimilitud dada, la distribución a posteriori es de la misma familia que la distribución a priori inicial. En este caso, pues, la prima Bayes se obtendrá sin

más que sustituir en la prima colectiva los parámetros de la distribución a priori inicial por los parámetros de la distribución a posteriori.

Obsérvese que, salvo para la expresión (1.9) la prima Bayes vendrá dada en general como un cociente de dos magnitudes a posteriori, que puede escribirse de la forma siguiente llamando $g(\theta) = m(\theta)n(\theta)$ y $h(\theta) = n(\theta)$

$$PB = \frac{\int_{\Theta} g(\theta)f(\mathbf{x}|\theta)\pi_0(\theta)d\theta}{\int_{\Theta} h(\theta)f(\mathbf{x}|\theta)\pi_0(\theta)d\theta}. \quad (1.12)$$

Para el principio de prima neta la prima es también una expresión como en (1.12) con $h(x) = 1$.

Las expresiones de la forma (1.12) bajo determinadas formas funcionales para $g(x)$ y $h(x)$ constituyen las primas que las compañías aseguradoras cobran a sus clientes.

Por otro lado, en los sistemas de tarificación bonus-malus (SBM, en adelante), utilizados mayoritariamente en el sector de seguros de automóviles, la prima se actualiza a medida que se incorpora la experiencia de siniestralidad además de cumplir determinadas reglas de transición. Así, si denominamos mediante PBM a la prima bonus-malus, esta ha de verificar:

$$\frac{\partial PBM}{\partial \mathbf{x}} > 0, \quad (1.13)$$

$$\frac{\partial PBM}{\partial t} < 0, \quad (1.14)$$

en la que \mathbf{x} representa la información muestral, usualmente el número de reclamaciones, y t es el período de tiempo transcurrido, usualmente el año. La desigualdad (1.13) fuerza a que la prima aumente con el número de reclamaciones, mientras que la desigualdad (1.14) exige que la prima disminuya con el tiempo si el número de reclamaciones permanece constante.

Una forma natural y habitual (véase Gómez et al., 2002, 2003, 2004a, 2004b, 2005a, 2005b, 2006b; Lemaire, 1979, 1985, 1995; Meng et al., 1999; Sarabia et al., 2004; Tremblay, 1992; entre otros) de construir primas que satisfagan estas sencillas reglas de transición, utilizando la metodología bayesiana, consiste en dividir la prima Bayes entre la prima colectiva para los principios de cálculo de prima estudiados:

$$PBM = \frac{PB}{PC}. \quad (1.15)$$

Así, teniendo en cuenta (1.12) las primas bonus-malus vendrán expresadas como un cociente entre una magnitud a posteriori y una magnitud a priori, en la

forma:

$$PBM = \frac{\int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \pi_0(\theta) d\theta}{\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_0(\theta) d\theta} \frac{\int_{\Theta} h(\theta) \pi_0(\theta) d\theta}{\int_{\Theta} g(\theta) \pi_0(\theta) d\theta}. \quad (1.16)$$

De esta manera, bajo los principios de cálculo de primas neta, Esscher y varianza, la prima bonus-malus a aplicar vendrá dada por:

$$PBM = \frac{\mathbb{E}_{\pi(\theta|\mathbf{x})} [\mathbb{E}_{f(x|\theta)} (X|\theta)]}{\mathbb{E}_{\pi(\theta)} [\mathbb{E}_{f(x|\theta)} (X|\theta)]}, \quad (1.17)$$

$$PBM = \frac{E_{\pi(\theta|\mathbf{x})} [P(\theta)e^{\alpha P(\theta)}]}{E_{\pi(\theta|\mathbf{x})} [e^{\alpha P(\theta)}]} \frac{E_{\pi(\theta)} [e^{\alpha P(\theta)}]}{E_{\pi(\theta)} [P(\theta)e^{\alpha P(\theta)}]}, \quad (1.18)$$

$$PBM = \frac{E_{\pi(\theta|\mathbf{x})} [P(\theta)^2]}{E_{\pi(\theta|\mathbf{x})} [P(\theta)]} \frac{E_{\pi(\theta)} [P(\theta)]}{E_{\pi(\theta)} [P(\theta)^2]}, \quad (1.19)$$

respectivamente.

1.2. Introducción al análisis bayesiano robusto local

Se introducen en esta sección, y de manera somera, los elementos básicos del análisis bayesiano robusto, haciendo hincapié en el análisis local o infinitesimal, que será la herramienta a utilizar en esta Memoria. La metodología necesaria para llevar a cabo un estudio de esta naturaleza se introducirá con detalle en el capítulo 2 de esta Memoria.

Es bien sabido que la metodología bayesiana requiere calcular la distribución a posteriori, que constituye la herramienta fundamental de trabajo y sobre la que se realizan las estimaciones convenientes. La introducción de la distribución a priori, que se apoya en creencias subjetivas del investigador, ha sido criticada por los estadísticos no bayesianos desde hace décadas (Klugman, 1987, p. 318; Ríos et al., 1999; entre otros). Como respuesta a ello se introdujo el análisis de robustez o de sensibilidad bayesiano que, *grosso modo*, consiste en estudiar el comportamiento de una magnitud a posteriori de interés cuando se incorpora incertidumbre sobre la distribución a priori.

Este estudio puede ser abordado desde dos perspectivas, robustez global y local, conducentes en muchas ocasiones a resultados análogos aunque con diferentes interpretaciones, y que aparecen interconectados (véase Gustafson, 1996a, 1996b).

La ventaja que tiene el análisis local frente al análisis global radica fundamentalmente en que puede ser aplicado en cualquier situación, mientras que el análisis global adolece en muchas ocasiones a estar limitado a ciertas magnitudes a posteriori que aparecen expresadas en términos sencillos.

En líneas generales, el planteamiento de un análisis de robustez bayesiano global consiste en considerar que existe imprecisión sobre la distribución a priori. Esta imprecisión puede estar originada porque no se está seguro de ella o porque la decisión sobre la misma ha de llevarse a cabo no por un investigador sino por dos o más, que no tienen porqué estar de acuerdo en sus impresiones acerca de la distribución a priori. En líneas generales, la metodología de robustez global consiste en sustituir la distribución a priori inicial por una clase suficientemente grande de distribuciones a priori y calcular el rango de variación de la magnitud a posteriori de interés. Si bien, podríamos considerar incertidumbre en cualquiera de las entradas del modelo, ya sea la función de verosimilitud y/o la función de pérdida y analizar el comportamiento de la magnitud a posteriori. Si el rango de variación es suficientemente pequeño podemos considerar que la magnitud a posteriori resultante es robusta respecto a la distribución a priori inicial. En caso contrario, si el rango es grande, se concluye que no es robusta.

Un estudio de esta naturaleza puede llevarse a cabo utilizando la clase de contaminación:

$$\Gamma_\epsilon = \{\pi(\theta) = (1 - \epsilon)\pi_0(\theta) + \epsilon q(\theta) : q \in \mathcal{Q}\}, \quad (1.20)$$

donde $\epsilon \in [0, 1]$ mide la incertidumbre sobre la distribución a priori inicial π_0 en la que inicialmente confía el investigador y \mathcal{Q} es una clase de distribuciones a priori llamada clase contaminante.

Usualmente el estudio suele iniciarse tomando como clase contaminante \mathcal{Q} la clase de todas las distribuciones de probabilidad. El cálculo requerido para esta clase suele ser bastante sencillo, ya que es una clase flexible que juega con la idea de proximidad entre distribuciones de probabilidad, concepto bastante intuitivo. En ocasiones, los rangos de variación de la magnitud a posteriori son amplios puesto que se están incorporando distribuciones de probabilidad que con seguridad nada tienen que ver con la modelización del problema bajo estudio. Por ejemplo, si la distribución a priori bajo estudio es discreta, entonces tomando la clase de todas las distribuciones de probabilidad, no podemos discernir entre distribuciones discretas y continuas. O, si la distribución a priori es unimodal, la clase podría incorporar por supuesto a distribuciones que no poseen esta característica.

Esta clase ha sido usada en un análisis bayesiano general en Berger (1985, 1994), Berger y Berliner (1986), Moreno y Cano (1991), entre otros. En escenarios actuariales se ha usado por Gómez et al. (1999, 2000).

El estudio, obviamente, puede refinarse tomando como clase contaminante la

clase de distribuciones de probabilidad con ciertas características conocidas, como la media, la moda, etc, teniendo en cuenta que algunas de estas características pueden ser intuitivas para un investigador entrenado. Por ejemplo, en los escenarios actuariales la multimodalidad es una propiedad de la distribución de probabilidad del número de reclamaciones y puede implementarse, sin ningún género de dudas, en un análisis como el detallado.

Por otra parte, el estudio de robustez se puede llevar a cabo también, como ya se comentó, desde un punto de vista local o infinitesimal. Esta técnica presenta la ventaja de permitir el estudio de sensibilidad o robustez cuando las características del problema hace muy difícil el estudio bajo robustez global. El análisis de sensibilidad o de robustez local de (1.12) y (1.16) será el objetivo fundamental de esta Memoria. Para ello, y teniendo en cuenta el marcado carácter dependiente de estas expresiones respecto de la distribución a priori, se estudiará la variación infinitesimal de (1.12) y (1.16) cuando la distribución a priori inicial se perturba infinitesimal y linealmente mediante una medida signada.

Contemplando (1.12) y (1.16) como un funcional dependiente de la distribución a priori inicial, se calculará la derivada funcional de aquella con respecto a una medida signada (perturbadora), cuantificando la magnitud de la variación a través de la norma de la derivada. Obviamente, las herramientas del cálculo diferencial tradicionales no pueden ser utilizadas aquí, así que el problema se abordará en el marco de los espacios funcionales, más concretamente los espacios L^p . Para calcular la derivada se utilizará el concepto de derivada de Fréchet, que ya ha sido utilizado anteriormente en problemas similares en el marco de la estadística bayesiana (Gustafson, 1996a, 1996b) y nunca, hasta que comenzó a realizarse esta Memoria en el escenario de la estadística actuarial (Calderín et al., 2007 y Gómez et al., 2004c, 2004d y 2006b).

Es usual, en el análisis funcional, utilizar las derivadas funcionales de Gateaux y de Fréchet cuando se requiere de la herramienta del cálculo diferencial. Sin embargo, y puesto que la derivada de Fréchet es más fuerte que la de Gateaux, en el sentido de que si un funcional es derivable Fréchet es derivable Gateaux, en esta Memoria haremos uso de la derivada de Fréchet solamente.

Introduciremos a continuación el concepto de derivada de Fréchet (Diaconis y Freedman, 1986; Gustafson, 1996a, Martín et al., 2003, Milne, 1980; entre otros).

En la siguiente definición denotaremos por \mathcal{U} y \mathcal{V} a dos espacios vectoriales topológicos y por $\mathcal{D}(\mathcal{P}) \subset \mathcal{U}$ un subconjunto abierto de \mathcal{U} .

Definición 1.5 *Un operador $\mathcal{P} : \mathcal{U} \rightarrow \mathcal{V}$ es Fréchet diferenciable en $x \in \mathcal{D}(\mathcal{P}) \subset \mathcal{U}$ si existe un operador lineal continuo $\dot{\mathcal{P}}(x) \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ de forma que, para todo $s \in \mathcal{U}$,*

$$\mathcal{P}(x + s) - \mathcal{P}(x) = \dot{\mathcal{P}}(x)s + \epsilon(x; s),$$

con

$$\lim_{\|s\|_{\mathcal{U}} \rightarrow 0} \frac{\|\varepsilon(x; s)\|_{\mathcal{V}}}{\|s\|_{\mathcal{U}}} = 0,$$

donde $\|\cdot\|_{\mathcal{V}}$ y $\|\cdot\|_{\mathcal{U}}$ denotan la norma en los espacios \mathcal{V} y \mathcal{U} respectivamente.

El operador $\hat{P}(x)$ se llama derivada de Fréchet o derivada fuerte de \mathcal{P} en x .

En concreto el estudio que procederemos a llevar a cabo puede resumirse de la siguiente manera. La prima Bayes, PB , así como la prima bonus-malus, PBM , pueden contemplarse como un funcional dependiente de la distribución a priori inicial, $P(\pi_0)$. Entonces, perturbaremos la distribución a priori inicial a través de una medida signada, esto es, intercambiaremos $\pi_0(\theta)$ por $\pi(\theta) = \pi_0(\theta) + \delta(\theta)$, siendo $\int_{\Theta} \delta(\theta) d\theta = 0$. Obsérvese que la distribución $\pi(\theta)$ obtenida de (1.20) puede escribirse como

$$\pi(\theta) = \pi_0(\theta) + \varepsilon(q(\theta) - \pi_0(\theta)) = \pi_0(\theta) + \delta(\theta),$$

donde $\delta(\theta) = \varepsilon(q(\theta) - \pi_0(\theta))$ es la medida signada. Es aquí donde radica el nexo existente entre el análisis global y local que comentamos anteriormente.

Finalmente, una vez perturbada la distribución a priori a través de $\delta(\theta)$ procederemos a calcular la derivada funcional (derivada de Fréchet) de la magnitud a posteriori de interés, en nuestro caso de (1.12) y (1.16), con respecto a la distribución perturbadora en 0. Es decir, calcularemos la variación infinitesimal de la magnitud a posteriori ante cambios infinitesimales en la distribución a priori. Evidentemente, lo que se obtiene es un vector, así que para medir la variación en términos cuantitativos calcularemos la norma de este vector.

Para ello consideremos el conjunto de todas las funciones medibles $f : \Theta \rightarrow \mathbb{R}$ cuyo valor absoluto elevado a la p -ésima potencia es integrable en el sentido Lebesgue, y tomaremos, dentro del amplio abanico de normas disponibles, la siguiente:

$$\|\delta/\pi_0; \Pi_0\|_p = \begin{cases} \left(\int_{\Theta} (\delta/\pi_0)^p d\Pi_0 \right)^{1/p}, & p < \infty, \\ \text{ess sup}_{\Theta} (\delta/\pi_0), & p = \infty, \end{cases}$$

siendo $p, q \in [0, \infty]$ los clásicos exponentes conjugados del análisis funcional tal que $1/p + 1/q = 1$ y ess sup representa el supremo esencial.

La norma se elegirá de esta manera porque satisface (Gustafson, 1996a) propiedades deseables en este contexto.

1.3. Objetivos de la Memoria

Las compañías aseguradoras, como cualquier otra empresa que opere en el mercado, son fundamentalmente empresas de servicios. Esto es ofrecen un servicio al

cliente, en este caso la seguridad, a cambio de un precio que es la prima. Como tal han de mostrarse competitivas y más si tenemos en cuenta que nos movemos en un mundo cada vez más globalizado, en el que está a la orden del día las fusiones y absorciones de compañías, unas a manos de otras.

La prima de riesgo representa el precio teórico que la compañía aseguradora cobraría a un individuo al suscribir éste un contrato de seguro determinado. Para conocer este precio exacto, la aseguradora debe conocer con precisión la distribución de siniestralidad individual con sus parámetros respectivos. Si esta información le es asequible a la aseguradora, entonces no hay problema que tratar y la prima que el cliente ha de pagar está fijado con precisión. En la práctica esto no ocurre ya que, o bien la distribución de probabilidad no es la adecuada o, conociendo la misma, se desconoce el valor del parámetro o parámetros de los que depende. Si es este el caso, el actuario o investigador asignará una distribución a priori al mismo y podrá calcular la prima colectiva, que es la mejor estimación, como ya se dijo, de la prima de riesgo, obviamente desconocida. La información acerca de la distribución a priori usualmente se obtiene de datos de contratos similares o de la experiencia anterior sobre los mismos contratos. Una vez especificada la distribución a priori y, si se dispone de información muestral, el teorema de Bayes permite calcular la distribución a posteriori a partir de la cual podemos, a su vez, obtener la prima Bayes. Esta prima combina dos cosas: la información a priori o del colectivo y la información muestral. Por tanto puede ser utilizada para calcular primas que puedan corregirse o actualizarse mediante la incorporación de la experiencia de siniestralidad. Esta es la base de los SBM.

Sin embargo, todo este trabajo puede venirse abajo si la modelización inicial asumida no resulta adecuada. Por ejemplo, si modelizamos la variable aleatoria número de reclamaciones mediante una distribución de Poisson, por ejemplo, y los datos empíricos nos dicen que la varianza es mayor que la media, obviamente la distribución de Poisson no es la apropiada ya que para ella la media y la varianza tienen el mismo valor. Se requiere en este caso considerar una distribución sobredispersa, $\text{Var}(X) > \mathbb{E}(X)$.

En la literatura actuarial es extensa la bibliografía existente sobre nuevas distribuciones de probabilidad que tratan de ajustar datos actuariales y que sean sobredispersas. Algunos de estos trabajos son Denuit (1997), Kokonendji y Khoudar (2004), Meng et al. (1999), Pérez et al. (2006), Sarabia et al. (2004), Simon (1961), Tremblay (1992), entre otros. En la mayoría de estos trabajos se generan nuevas distribuciones de probabilidad mezclando otras distribuciones conocidas. El ejemplo típico al que generalmente se hace alusión cuando se quiere explicar esa metodología es el Poisson-gamma. Así, si la variable aleatoria X sigue una distribución de Poisson, $f(x|\theta)$, con parámetro $\theta > 0$, y considerando el parámetro θ aleatorio con distribución $\pi(\theta)$ gamma con parámetros $a > 0$, $b > 0$, es sencillo

comprobar que la distribución incondicional de X dada por la mezcla

$$f(x) = \int_0^{\infty} f(x|\theta)\pi(\theta)d\theta,$$

es una binomial negativa con parámetros a y $b/(b+1)$.

En esta Memoria se abordará un problema de esta naturaleza, que llamaremos de inferencia, y que consistirá en obtener una distribución de probabilidad discreta mezclando la distribución binomial negativa con parámetros $r > 0$ y $p = e^{-\theta}$, $\theta > 0$, con la distribución inversa gaussiana. La distribución resultante, como veremos, es sobredispersa y, aunque no admite una expresión sencilla, los parámetros de la misma pueden estimarse por el método de los momentos y el de máxima verosimilitud de una manera relativamente fácil.

Para esta distribución, además, será posible obtener una fórmula recursiva para el cálculo de probabilidades que permitirá expresar la distribución de la cantidad total reclamada $X = \sum_{i=1}^N X_i$ en términos de una ecuación integro diferencial similar a la obtenida en Panjer (1981).

La distribución binomial negativa-inversa gaussiana, como así la hemos denominado, admite también una extensión multivariada para la que de nuevo no resulta complicado estimar los parámetros del modelo.

Por otro lado, en esta Memoria se aborda también, y como ya se comentó, el estudio de robustez o sensibilidad de las primas en general y de las primas bonus-malus bajo determinados modelos y principios de cálculo de primas asumidos. En particular estudiaremos cómo se comporta la prima Bayes y la prima bonus-malus cuando el conocimiento sobre la distribución a priori no es completo, admitiéndose como distribución a priori una lo suficientemente cercana a la inicial. El análisis bayesiano robusto local permite hacer un estudio de estas características en el que el término parecido se asume en el sentido infinitesimal. Es más, análisis de este tipo ha sido llevado a cabo por Gustafson (1996a, 1996b) para magnitudes a posteriori de la forma

$$\rho = \int_{\Theta} g(\theta)\pi_0(\theta|\mathbf{x})d\theta. \quad (1.21)$$

Sin embargo, no existe en la literatura al respecto un procedimiento teórico que nos permita estudiar desde un punto de vista local o infinitesimal la sensibilidad de magnitudes que vengan expresadas como cocientes de magnitudes a posteriori o cociente entre una magnitud a posteriori y una magnitud a priori, como así son las expresiones (1.12) y (1.16), respectivamente.

En el capítulo 2 abordaremos dicho problemas y se dará solución al mismo. Además, obviamente, se llegará a un resultado, para el caso del cociente (1.12)

que será una formulación general al problema resuelto en Gustafson (1996a, 1996b) correspondiente a la expresión (1.21).

Además, desarrollaremos la metodología apropiada que permita introducir características sobre la distribución a priori que puedan ser adecuadas para el problema que ocupe al investigador. Esta iniciativa, que es común en el análisis de robustez global, es novedosa bajo la perspectiva local. Las características a las que nos referimos son la unimodalidad y la simetría. Con ello se pretende refinar la clase Γ sobre la que haremos el análisis de sensibilidad local y plantear el problema en un contexto más realista.

Capítulo 2

Elementos del análisis bayesiano robusto

En este capítulo se introducirán las herramientas estadísticas necesarias para llevar a cabo un análisis bayesiano de sensibilidad desde el punto de vista local o infinitesimal. La literatura al respecto es escasa, siguiendo básicamente dos líneas de actuación. Una de ellas, basada en el uso de los espacios L^p en el contexto del análisis funcional será la que se seguirá en esta memoria.

2.1. Introducción

Es sabido que la metodología bayesiana requiere de:

1. Una distribución muestral o función de verosimilitud $f(x|\theta)$ dependiente de un parámetro θ o vector de parámetros desconocido θ
2. Una función de pérdida $L(\theta, \hat{\theta})$ que atribuye a un decisor la pérdida soportada por tomar como valor del parámetro $\hat{\theta}$ en lugar del verdadero valor θ
3. Una distribución a priori $\pi(\theta)$ asociada al parámetro.

La función de pérdida generalmente utilizada es

$$L(\theta, \hat{\theta}) = m(\theta)[n(\theta) - \hat{\theta}]^2$$

denominada en la literatura bayesiana función de pérdida cuadrática ponderada y que ya hemos tenido ocasión de comentar en el capítulo 1. También se comentó en dicho capítulo cómo resulta la magnitud a posteriori dependiendo de las funciones $m(\theta)$ y $n(\theta)$, sabiendo que dicha magnitud se obtiene minimizando el riesgo Bayes.

De manera simplificada podemos pues argumentar que el análisis bayesiano consta de tres entradas y una salida. Las entradas son la verosimilitud de los datos, la función de pérdida y la distribución a priori. La salida es la magnitud a posteriori de interés.

Visto de esta manera, podemos considerar la magnitud a posteriori como un funcional

$$\text{Magnitud a posteriori} \equiv \rho(f, L, \pi)$$

dependiente de tres funciones f , L y π

El análisis de sensibilidad o de robustez bayesiana consiste en estudiar la variación experimentada por la magnitud a posteriori cuando se modifica alguna o algunas de las entradas del modelo. Este estudio, que tuvo un auge extraordinario en la década de los 80 y principios de los 90 del siglo XX, y que aún hoy día sigue siendo fuente de numerosos trabajos de investigación, puede llevarse a cabo desde dos puntos de vista. Desde un punto de vista global, utilizando clase de distribuciones de probabilidad o clase de funciones de pérdidas, y desde un punto de vista local, variando infinitesimalmente alguna o algunas de las entradas del modelo. En el primer caso el problema se engloba dentro del cálculo variacional y en el segundo en el cálculo diferencial, conectados obviamente.

En esta Memoria, y como ya se comentó en el capítulo 1, el punto de vista adoptado es el infinitesimal, denominado también local. Ya se comentó que la literatura al respecto es escasa (Calderín et al. (2007), Gómez et al. (2005e, 2006b); Gustafson (1996a, 1996b); Gustafson y Wasserman (1995); Peña y Zamar (1997); Ruggeri y Wasserman (1993); Sivaganesan (2000)), basada además únicamente en la variación de la magnitud a posteriori ante cambios infinitesimales en la distribución a priori y utilizando la función de pérdida cuadrática. Como se sabe, en este caso la magnitud a posteriori es la media a posteriori del parámetro.

En este capítulo se introducirán las herramientas estadísticas necesarias para llevar a cabo un estudio de sensibilidad local ante cambios infinitesimales en la distribución a priori cuando se utiliza la función de pérdida cuadrática ponderada para valores de $m(\cdot)$ y $n(\cdot)$ arbitrarios. Este estudio se extenderá al caso en que la magnitud de interés venga dada por un cociente entre una magnitud a posteriori y una magnitud a priori, como ocurre con los sistemas de tarificación bonus-malus.

El capítulo se organiza de la siguiente manera. En la sección 2 se introducen las herramientas necesarias para llevar a cabo un estudio de robustez local como el que se ha comentado. En las secciones 3 y 4 se utilizan las herramientas anteriores para analizar la sensibilidad bayesiana local de magnitudes a posteriori y de expresiones dadas por cocientes entre una magnitud a posteriori y una magnitud a priori, respectivamente. El capítulo concluye con un resultado que conecta el análisis de sensibilidad local y global y otras consideraciones de interés.

2.2. Preliminares

Consideremos un espacio de medida $(\Theta, \mathcal{B}, \mu)$ y una función de densidad de probabilidad π_0 con respecto a $(\Theta, \mathcal{B}, \mu)$, que induce una medida de probabilidad Π_0 , entonces $\frac{d\Pi_0(\theta)}{d\theta} = \pi_0(\theta)$. Una manera sencilla de incorporar el análisis bayesiano local en un modelo paramétrico para el espacio muestral X y el espacio paramétrico $\Theta \subset \mathbb{R}$, $\{f(\mathbf{x}|\theta), \theta \in \Theta\}$, es utilizando la clase de distribuciones $\Gamma = \{\pi : \pi = \pi_0 + \delta\}$, donde δ es una medida signada, y entonces $\delta(\Theta) = 0$. Para el caso general en que $\delta(\theta)$ sea una medida no signada bastaría considerar

$$\pi(\theta) = \frac{\pi_0(\theta)}{1 + \int_{\Theta} \delta(\theta) d\theta}.$$

Obsérvese que si la medida signada δ viene dada en la forma $\delta = \varepsilon(q - \pi_0)$ con $q \in \mathcal{Q}$, $\varepsilon \in [0, 1]$ tenemos la conocida clase de contaminación (Martín et al., 2003; Gustafson, 1996a y Gómez et al., 2002; entre otros)

$$\pi(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon q(\theta),$$

usualmente utilizada en el estudio de robustez global.

Tanto si queremos llevar a cabo un estudio de robustez local del funcional (1.12) como del funcional (1.16), el camino que seguiremos es tomar $\pi \in \Gamma$ y hacer que δ varíe infinitesimalmente. Esta variación infinitesimal se hará en el sentido diferencial. No vale, por supuesto, la derivada usual del cálculo diferencial, y se hablará de derivadas funcionales.

Podemos elegir Γ para $c > 0$ de la forma $\Gamma_{L^p} = \{\pi : \|\delta/\pi_0; \Pi_0\|_p \leq c\}$, la clase de todas las perturbaciones sobre π_0 y acotadas, i.e. $\Gamma_{L^p} \subset \Gamma$.

A partir de la siguiente definición (Milne, 1980; Griffel, 1985; entre otros) obtenemos la norma de la derivada de Fréchet

Definición 2.1 Si $\dot{P}(\delta_0)$ es la derivada de $P(\delta)$ en δ_0 en la dirección δ entonces:

$$\|\dot{P}(\delta_0)\| = \sup_{\|\delta\|=1} \|\dot{P}(\delta_0)\delta\|_{\mathcal{V}}.$$

Para concluir esta sección recordaremos un aspecto que nos será de utilidad más adelante, el conocido resultado en el análisis funcional que afirma que para cualquier función I de Θ en \mathbb{R} podemos escribir $I = I^+ - I^-$ y $|I| = I^+ + I^-$, donde $I^+ = \max\{I, 0\}$ y $I^- = \max\{-I, 0\}$.

2.3. Robustez local para cociente de magnitudes a posteriori

En esta sección obtendremos la norma de la derivada de Fréchet del funcional (1.12), esto es del cociente de magnitudes a posteriori, y que no es más que la

magnitud a posteriori que se obtiene bajo pérdida cuadrática ponderada, en el que $g(\cdot)$ y $h(\cdot)$ son funciones arbitrarias.

Recordemos que habíamos denotado por \mathcal{U} y \mathcal{V} a dos espacios vectoriales topológicos y por $\mathcal{D}(\mathcal{P}) \subset \mathcal{U}$ un subconjunto abierto de \mathcal{U} .

A partir de ahora denotaremos al conjunto \mathcal{U} como

$$\mathcal{U} = \left\{ \delta \geq 0 : \|\delta/\pi_0; \Pi_0\|_p < \infty \right\},$$

y al conjunto \mathcal{V} por $\mathcal{V} = \mathbb{R}$. Adoptaremos $\|\delta/\pi_0; \Pi_0\|_p$ como la norma sobre \mathcal{U} y el valor absoluto como la norma sobre \mathbb{R} .

Se necesita, previo a obtener la derivada de Fréchet de (1.12), el siguiente lema,

Lema 2.1 *La derivada de*

$$T^h(\delta) = \int_{\Theta} h(\theta)[\pi(\theta) + \delta(\theta)]d\theta$$

en la dirección δ viene dada por

$$\dot{T}^h(\delta_0)\delta = \int_{\Theta} h(\theta)\delta(\theta)d\theta.$$

Demostración: De la definición de derivada de Fréchet tenemos que

$$\dot{T}^h(\delta_0)\delta = T^h(\pi + \delta) - T^h(\pi) + o(\|\delta\|).$$

Debemos pues probar que se cumple:

$$\left\| T^h(\delta + \delta_0) - T^h(\delta_0) - \dot{T}^h(\delta_0)\delta \right\|_{\mathcal{V}} = o(\|\delta\|_{\mathcal{U}}),$$

lo que resulta evidente teniendo en cuenta que:

$$\begin{aligned} & \left\| \int_{\Theta} h(\theta) [\pi_0(\theta) + (\delta + \delta_0)(\theta)] d\theta - \int_{\Theta} h(\theta) [\pi_0(\theta) + \delta_0(\theta)] d\theta \right. \\ & \quad \left. - \int_{\Theta} h(\theta)\delta(\theta)d\theta \right\|_{\mathcal{V}} = o(\|\delta_0\|_{\mathcal{V}}). \end{aligned}$$

Para simplificar la notación, a partir de ahora denotaremos por π_{δ_0} y $\pi_{\delta_0}^x$ las distribuciones a priori y a posteriori respectivamente, cuando $\pi = \pi_0 + \delta_0$. ■

Ahora ya estamos en disposición de obtener la derivada de Fréchet de (1.12) en la dirección de δ_0 , y que se muestra en el siguiente resultado.

Teorema 2.1 *Supongamos que gf y hf son integrables, entonces la derivada de (1.12) en el sentido Fréchet viene dada por*

$$\dot{P}B(\delta_0)\delta = \frac{\mathbb{E}_{\pi_{\delta_0}^x} \left[g(\theta) \frac{\delta(\theta)}{\pi_{\delta_0}(\theta)} \right]}{\mathbb{E}_{\pi_{\delta_0}^x} [h(\theta)]} - \frac{\mathbb{E}_{\pi_{\delta_0}^x} [g(\theta)] \mathbb{E}_{\pi_{\delta_0}^x} \left[h(\theta) \frac{\delta(\theta)}{\pi_{\delta_0}(\theta)} \right]}{\mathbb{E}_{\pi_{\delta_0}^x} [h(\theta)] \mathbb{E}_{\pi_{\delta_0}^x} [h(\theta)]}. \quad (2.1)$$

Demostración: Derivando (1.12) respecto de δ en la dirección δ_0 tenemos

$$\begin{aligned} & \dot{P}B(\delta_0)\delta \\ &= \frac{\int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \delta(\theta) d\theta \int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta}{\left[\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta \right]^2} \\ &- \frac{\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \delta(\theta) d\theta \int_{\Theta} g(\theta) f(\mathbf{x}|\theta) [\pi_{\delta_0}(\theta)] d\theta}{\left[\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta \right]^2} \\ &= \frac{\int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \delta(\theta) d\theta}{\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta} - \frac{\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \delta(\theta) d\theta \int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta}{\left[\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta \right]^2} \\ &= \frac{\int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \frac{\delta(\theta)}{\pi_{\delta_0}(\theta)} \pi_{\delta_0}(\theta) d\theta}{\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta} \\ &- \frac{\int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta \int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \frac{\delta(\theta)}{\pi_{\delta_0}(\theta)} (\theta) d\theta}{\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta \int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta} \\ &= \frac{E_{\pi_{\delta_0}^x} \left[g(\theta) \frac{\delta(\theta)}{\pi_{\delta_0}(\theta)} \right]}{E_{\pi_{\delta_0}^x} [h(\theta)]} - \frac{E_{\pi_{\delta_0}^x} [g(\theta)] E_{\pi_{\delta_0}^x} \left[h(\theta) \frac{\delta(\theta)}{\pi_{\delta_0}(\theta)} \right]}{E_{\pi_{\delta_0}^x} [h(\theta)] E_{\pi_{\delta_0}^x} [h(\theta)]}. \end{aligned}$$

Parece lógico que si lo que se pretende es estudiar la variación del funcional $PB(\theta)$ ante cambios infinitesimales en la distribución a priori, el cambio ha de ser lo suficientemente pequeño. Así que supondremos en $\pi = \pi_0 + \delta_0$ que $\delta_0 = 0$. Ahora (2.1) puede escribirse de una forma más práctica como mostramos en el siguiente resultado. ■

Proposición 2.1 *La derivada de (1.12) cuando $\delta_0 = 0$ viene dada por*

$$\dot{P}B(0)\delta = \frac{1}{N_{\pi_0}^h} \left(N_{\delta}^g - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h \right). \quad (2.2)$$

siendo

$$N_{\pi_0}^g = \int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \pi_0(\theta) d\theta,$$

$$\begin{aligned}
 N_{\pi_0}^h &= \int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_0(\theta) d\theta, \\
 N_{\delta}^g &= \int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \delta(\theta) d\theta, \\
 N_{\pi_0}^h &= \int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \delta(\theta) d\theta.
 \end{aligned}$$

Demostración: La expresión (2.1) es

$$\dot{P}B(0)\delta = \frac{E_{\pi_0^*} [g(\theta) \frac{\delta(\theta)}{\pi_0(\theta)}]}{E_{\pi_0^*} [h(\theta)]} - \frac{E_{\pi_0^*} [g(\theta)]}{E_{\pi_0^*} [h(\theta)]} \frac{E_{\pi_0^*} [h(\theta) \frac{\delta(\theta)}{\pi_0(\theta)}]}{E_{\pi_0^*} [h(\theta)]}.$$

Sacando factor común y cancelando las distribuciones marginales tenemos

$$\begin{aligned}
 \dot{P}B(0)\delta &= \frac{1}{N_{\pi_0}^h} \left(\int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \frac{\delta(\theta)}{\pi_0(\theta)} \pi_0(\theta) d\theta \right. \\
 &\quad \left. - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} \int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \frac{\delta(\theta)}{\pi_0(\theta)} \pi_0(\theta) d\theta \right) \\
 &= \frac{1}{N_{\pi_0}^h} \left(\int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \delta(\theta) d\theta - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} \int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \delta(\theta) d\theta \right) \\
 &= \frac{1}{N_{\pi_0}^h} \left(N_{\delta}^g - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h \right).
 \end{aligned}$$

La conexión entre el trabajo de Gustafson (1996a, 1996b) y el realizado en esta Memoria viene dada en el presente comentario: ■

Comentario 2.1 Si $h(\theta) = 1$, la expresión (2.2) se reduce a $a_L(\theta)$ en el resultado 1 en Gustafson (1996b).

Obsérvese que a la expresión (2.2) también podía haberse llegado de una forma alternativa como mostramos en el siguiente resultado.

Teorema 2.2 Supongamos que gf y hf son integrables, entonces

$$\dot{P}B(0)\delta = \frac{1}{N_{\pi_0}^h} \left(N_{\delta}^g - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h \right).$$

Demostración:

$$PB(\pi_0 + \delta) - PB(\pi_0) = \frac{N_{\pi_0 + \delta}^g}{N_{\pi_0 + \delta}^h} - \frac{N^g \pi_0}{N^h \pi_0} = \frac{N_{\pi_0}^g + N_{\delta}^g}{N_{\pi_0}^h + N_{\delta}^h} - \frac{N_{\pi_0}^g}{N_{\pi_0}^h}$$

$$\begin{aligned}
&= \frac{1}{N_{\pi_0}^h} \left(\frac{N_{\pi_0}^g N_{\pi_0}^h + N_{\delta}^g N^h \pi_0}{N_{\pi_0}^h + N_{\delta}^h} - N_{\pi_0}^g \right) \\
&= \frac{1}{N_{\pi_0}^h} \left(\frac{N_{\pi_0}^g N_{\pi_0}^h + N_{\delta}^g N^h \pi_0 - N_{\pi_0}^h N_{\pi_0}^g - N_{\pi_0}^g N_{\delta}^h}{N_{\pi_0}^h + N_{\delta}^h} \right) \\
&= \frac{1}{N_{\pi_0}^h} \left(\frac{N_{\delta}^g N_{\pi_0}^h - N_{\pi_0}^g N_{\delta}^h \frac{N_{\pi_0}^h}{N_{\pi_0}^h} + N_{\delta}^g N_{\delta}^h - N_{\delta}^g N_{\delta}^h + \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h N_{\delta}^h - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h N_{\delta}^h}{N_{\pi_0}^h + N_{\delta}^h} \right) \\
&= \frac{1}{N_{\pi_0}^h} \left(\frac{N_{\delta}^g (N_{\pi_0}^h + N_{\delta}^h) - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} (N_{\delta}^h N_{\pi_0}^h + N_{\delta}^h N_{\delta}^h) - N_{\delta}^g N_{\delta}^h + \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h N_{\delta}^h}{N_{\pi_0}^h + N_{\delta}^h} \right) \\
&= \frac{1}{N_{\pi_0}^h} \left(\frac{N_{\delta}^g (N_{\pi_0}^h + N_{\delta}^h) - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h (N_{\pi_0}^h + N_{\delta}^h) + \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h N_{\delta}^h - N_{\delta}^g N_{\delta}^h}{N_{\pi_0}^h + N_{\delta}^h} \right) \\
&= \frac{1}{N_{\pi_0}^h} \left(N_{\delta}^g - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h \right) - \frac{1}{N_{\pi_0}^h} \frac{N_{\pi_0}^g N_{\delta}^h - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h N_{\delta}^h}{N_{\pi_0}^h + N_{\delta}^h} \\
&= \frac{1}{N_{\pi_0}^h} \left(N_{\delta}^g - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h \right) - \frac{N_{\pi_0}^h N_{\delta}^g N_{\delta}^h - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h N_{\delta}^h N_{\pi_0}^h}{(N_{\delta}^h)^2 (N_{\pi_0}^h + N_{\delta}^h)} \\
&= \frac{1}{N_{\pi_0}^h} \left(N_{\delta}^g - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h \right) - \frac{N_{\delta}^h (N_{\delta}^g N_{\pi_0}^h - N_{\delta}^h N_{\pi_0}^g)}{(N_{\pi_0}^h)^2 (N_{\pi_0}^h + N_{\delta}^h)}.
\end{aligned}$$

El segundo miembro de la última expresión puede reescribirse de la forma siguiente

$$R = \frac{N_{\delta}^h (N_{\delta}^g N_{\pi_0}^h - N_{\delta}^h N_{\pi_0}^g)}{(N_{\pi_0}^h)^2 (N_{\pi_0}^h + N_{\delta}^h)}.$$

Sólo queda por probar que $\|R\| \rightarrow 0$, pero

$$PB(\pi_0 + \delta) - PB(\pi_0) \approx \frac{1}{N_{\pi_0}^h} \left(N_{\delta}^g - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h \right)$$

$$\begin{aligned}
\|R\| &\leq \frac{|N_{\delta}^h|}{|N_{\pi_0}^h + N_{\delta}^h| |N_{\pi_0}^h|} \frac{|N_{\delta}^g N_{\pi_0}^h - N_{\pi_0}^g N_{\delta}^h|}{N_{\pi_0}^h} \\
&= \frac{|N_{\delta}^h|}{|N_{\pi_0}^h + N_{\delta}^h| |N_{\pi_0}^h|} \left| N_{\delta}^g - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h \right|.
\end{aligned}$$

Podemos acotar la última desigualdad puesto que hf y gf están acotadas.

$$\begin{aligned} |N_\delta^h| &\leq \int_{\Theta} |h(\theta)| |\delta(\theta|\mathbf{x})| d\theta \leq \int_{\Theta} |h(\theta)| f(\mathbf{x}|\theta) |\delta(\theta)| d\theta \\ &\leq \int_{\Theta} a |\delta(\theta)| d\theta = a \int_{\Theta} |\delta(\theta)| d\theta = a \|\delta\|. \end{aligned}$$

$$\begin{aligned} |N_\delta^g| &\leq \int_{\Theta} |g(\theta)| |\delta(\theta|\mathbf{x})| d\theta \leq \int_{\Theta} |g(\theta)| f(\mathbf{x}|\theta) |\delta(\theta)| d\theta \\ &\leq \int_{\Theta} b |\delta(\theta)| d\theta = b \int_{\Theta} |\delta(\theta)| d\theta = b \|\delta\|. \end{aligned}$$

$$\begin{aligned} \|R\| &\leq \frac{a\|\delta\|}{|N_{\pi_0}^h + N_\delta^h| |N_{\pi_0}^h|} \left| b\|\delta\| - \frac{N_{\pi_0}^g}{N^h \pi_0} a\|\delta\| \right| \\ &= \frac{a\|\delta\|^2}{|N_{\pi_0}^h + N_\delta^h| |N_{\pi_0}^h|} \left| b - \frac{N_{\pi_0}^g}{N^h \pi_0} a \right| \leq C_{\pi_0} \|\delta\|^2 = o(\|\delta\|). \end{aligned}$$

Puesto que C_{π_0} es una constante que depende de π_0

$$\|R\| \leq o(\pi_0).$$

si $\delta \rightarrow 0$ entonces $\|R\| \rightarrow 0$. ■

De esta forma hemos obtenido la misma expresión para la derivada de (1.12) que la obtenida mediante las propiedades de la derivada de Fréchet.

La siguiente Proposición nos permite expresar la derivada del funcional en una notación más conveniente, modificando la dirección de la derivada.

Proposición 2.2 *Se tiene que*

$$\dot{P}B(0)\delta = \dot{P}B(0) \frac{\delta}{\pi_0}. \quad (2.3)$$

Demostración: Se obtiene de forma inmediata intercambiando en (2.2) δ y $d\theta$ por $\frac{\delta}{\pi_0}$ y $d\Pi_0(\theta)$. Así:

$$\dot{P}B(0)\delta = \frac{N_\delta^g}{N_{\pi_0}^h} - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} \frac{N_\delta^h}{N_{\pi_0}^h} = \frac{N_{\frac{\delta}{\pi_0}}^g}{N_{\pi_0}^h} - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} \frac{N_{\frac{\delta}{\pi_0}}^h}{N_{\pi_0}^h} = \dot{P}B(0) \frac{\delta}{\pi_0}.$$

El siguiente Lema nos permite estudiar la derivada de (1.12) en una notación más cómoda ■

Lema 2.2 Sea $E_{\pi_0^x}[g(\theta)] = \rho_g$ y $E_{\pi_0^x}[h(\theta)] = \rho_h$ entonces

$$\dot{P}B(0)\delta = \frac{1}{\rho_h} \int_{\Theta} I(\theta)\delta(\theta)d\theta, \quad (2.4)$$

siendo

$$I(\theta) = \left(g(\theta) - h(\theta) \frac{\rho_g}{\rho_h} \right) \frac{\pi_0^x(\theta)}{\pi_0(\theta)}. \quad (2.5)$$

Demostración: Si expresamos la distribución marginal por

$$M_{\pi_0} = \int_{\Theta} f(\mathbf{x}|\theta)\pi_0(\theta)d\theta.$$

La expresión (2.2) puede reescribirse de la forma

$$\begin{aligned} \dot{P}B(0)\delta &= \frac{1}{N_{\pi_0}^h} \left(N_{\delta}^g - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h \right) \\ &= \frac{1}{N_{\pi_0}^h \rho_h} \left(N_{\delta}^g - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h \right) \\ &= \frac{1}{N_{\pi_0}^h \rho_h} \int_{\Theta} \left(g(\theta) - h(\theta) \frac{\rho_g}{\rho_h} \right) f(\mathbf{x}|\theta)\delta(\theta)d\theta \\ &= \frac{1}{\rho_h} \int_{\Theta} \left(g(\theta) - h(\theta) \frac{\rho_g}{\rho_h} \right) \frac{f(\mathbf{x}|\theta)}{M_{\pi_0}} \delta(\theta)d\theta \\ &= \frac{1}{\rho_h} \int_{\Theta} \left(g(\theta) - h(\theta) \frac{\rho_g}{\rho_h} \right) \frac{\pi_0^x(\theta)}{\pi_0(\theta)} \delta(\theta)d\theta \\ &= \frac{1}{\rho_h} \int_{\Theta} I(\theta)\delta(\theta)d\theta. \end{aligned}$$

Para obtener la norma de la derivada (2.4), se requiere previamente del siguiente resultado que aparece en Gustafson (1996b) ■

Teorema 2.3 (Gustafson (1996b)) Sea $T\delta = \int_{\Theta} g(\theta)\pi^x(\theta)d\theta$, entonces

$$\|T(\dot{\delta})\|_p = \max\{\|a^+; \Pi\|, \|a^-; \Pi\|\}$$

donde

$$a(\theta) = \frac{(g(\theta) - \rho_g)\pi^x(\theta)}{\pi(\theta)}, \quad (2.6)$$

siendo

$$\rho_g = \int_{\Theta} g(\theta) \pi_0^x(\theta) d\theta$$

y

$$a^+(\theta) = \max\{a(\theta), 0\}, \quad a^-(\theta) = -\min\{a(\theta), 0\}$$

Ahora podemos obtener la norma de la derivada de Fréchet de (2.4) y, por tanto, de (1.12) como aparece a continuación.

Teorema 2.4 *Bajo las condiciones del Lema 2.2 se tiene:*

$$\|\dot{P}B(0)\|_p = \max \left\{ \left\| \frac{1}{\rho_h} I^+(\theta) \right\|_q, \left\| \frac{1}{\rho_h} I^-(\theta) \right\|_q \right\} \quad (2.7)$$

Demostración: Obsérvese que (2.6) puede escribirse como

$$a(\theta) = G(\theta) \frac{\pi^x(\theta)}{\pi(\theta)},$$

siendo $G(\theta) = g(\theta) - \rho_g$, de la misma forma que (2.5) puede reescribirse como

$$I(\theta) = G_1(\theta) \frac{\pi^x(\theta)}{\pi(\theta)},$$

con $G_1(\theta) = g(\theta) - h(\theta)\rho_g/\rho_h$. Ahora, el resultado se sigue de utilizar el Teorema 2.3. ■

2.3.1. Incorporación de la propiedad de unimodalidad

En muchas ocasiones el investigador puede estar seguro que el valor más probable del parámetro de riesgo está en el entorno de un valor concreto. Por ejemplo, al considerar el número de reclamaciones en una cartera de seguros de automóviles el valor más probable del parámetro está en la proximidad de 0. Por esta razón, se hace necesario usar una clase más refinada que la de todas las distribuciones de probabilidad, puesto que esta clase reduce el efecto de la distribución a priori usada, ya que ésta presenta un carácter subjetivo y depende de un parámetro desconocido. Todo esto nos conduce a usar la clase de las distribuciones unimodales.

El siguiente teorema nos expresa en una forma más adecuada la expresión $\dot{P}B(0)\delta$, en términos de una función de influencia, que puede usarse cuando consideramos las clases

$$\begin{aligned} \Gamma^A &= \{\text{Todas las distribuciones de probabilidad } \delta\}, \\ \Gamma^U &= \{\text{Todas las distribuciones de probabilidad unimodales } \delta \\ &\quad \text{con la misma moda } \theta_0 \text{ que } \pi_0\}, \end{aligned}$$

Teorema 2.5 Sea $\rho_h = E_{\pi_0^x}[h(z)]$, entonces $PB(0)\delta = \int I(z)\delta(z)dz$, donde la función de influencia, $I(z)$, viene dada por

$$\begin{aligned} I(z) &= \frac{1}{\rho_h} \left[g(z) - h(z) \frac{N_{\pi_0}^g}{N_{\pi_0}^h} \right] \frac{\pi_0^x(z)}{\pi_0(z)}, \quad \delta \in \Gamma^A, \\ I(z) &= \frac{1}{z\rho_h} \int_{\theta_0}^{\theta_0+z} \left[g(\theta) - h(\theta) \frac{N_{\pi_0}^g}{N_{\pi_0}^h} \right] \frac{\pi_0^x(\theta)}{\pi_0(\theta)} d\theta, \quad z \neq 0, \delta \in \Gamma^U, \\ I(z) &= \frac{1}{\rho_h} \left[g(\theta_0) - h(\theta_0) \frac{N_{\pi_0}^g}{N_{\pi_0}^h} \right] \frac{\pi_0^x(\theta_0)}{\pi_0(\theta_0)}, \quad z = 0, \delta \in \Gamma^U. \end{aligned} \quad (2.8)$$

Demostración: Si denotamos $m(\mathbf{x}|\pi_0) = \int f(\mathbf{x}|z)\pi_0(z)dz$, la expresión (2.2) puede reescribirse como

$$\begin{aligned} PB(0)\delta &= \frac{1}{m(\mathbf{x}|\pi_0)\rho_h} \left(N_{\delta}^g - \frac{N_{\pi_0}^g}{N_{\pi_0}^h} N_{\delta}^h \right) \\ &= \frac{1}{m(\mathbf{x}|\pi_0)\rho_h} \int \left[g(z) - h(z) \frac{N_{\pi_0}^g}{N_{\pi_0}^h} \right] f(\mathbf{x}|z)\delta(z)dz \\ &= \frac{1}{\rho_h} \int \left[g(z) - h(z) \frac{N_{\pi_0}^g}{N_{\pi_0}^h} \right] \frac{f(\mathbf{x}|z)}{m(\mathbf{x}|\pi_0)} \delta(z)dz \\ &= \frac{1}{\rho_h} \int \left[g(z) - h(z) \frac{N_{\pi_0}^g}{N_{\pi_0}^h} \right] \frac{\pi_0^x(z)}{\pi_0(z)} \delta(z)dz \\ &= \int I(z)\delta(z)dz, \end{aligned}$$

para $\delta \in \Gamma^A$. Para $\delta \in \Gamma^U$, es sólo necesario usar el Lemma 3.2.1 en Sivaganesan y Berger (1989) que afirma que la función $I(z)$ se puede representar como una mezcla de distribuciones uniformes. ■

2.4. Robustez local bajo pérdida cuadrática ponderada

Un estudio paralelo al desarrollado en la sección anterior se llevará a cabo para el caso en que la magnitud de interés aparezca expresada como en (1.16), obtenida a partir de utilizar la pérdida cuadrática ponderada.

Nuevamente, volveremos a escribir π_{δ_0} y $\pi_{\delta_0}^x$ para denotar las distribuciones a priori y a posteriori, respectivamente, cuando $\pi = \pi_0 + \delta_0$. Ahora, la derivada de Fréchet de (1.16) viene dada en el siguiente teorema.

Teorema 2.6 Sea $PBM(\pi)$ en la forma (1.16) y $\pi = \pi_0 + \delta \in \Gamma_{L^p}$, entonces:

$$\begin{aligned}
 P\dot{B}M(\delta_0)\delta &= E_{\pi_{\delta_0}^*} \left[g(\theta) \frac{\delta(\theta)}{\pi_{\delta_0}(\theta)} \right] \left\{ \frac{E_{\pi_{\delta_0}} [h(\theta)]}{E_{\pi_{\delta_0}} [g(\theta)] E_{\pi_{\delta_0}^*} [h(\theta)]} \right. \\
 &\quad \left. - \frac{E_{\pi_{\delta_0}^*} [g(\theta)] E_{\pi_{\delta_0}} [h(\theta)]}{E_{\pi_{\delta_0}^*} [h(\theta)] \{E_{\pi_{\delta_0}} [g(\theta)]\}^2} \right\} \\
 &+ E_{\pi_{\delta_0}^*} \left[h(\theta) \frac{\delta(\theta)}{\pi_{\delta_0}(\theta)} \right] \left\{ \frac{E_{\pi_{\delta_0}^*} [g(\theta)]}{E_{\pi_{\delta_0}^*} [h(\theta)] E_{\pi_{\delta_0}} [g(\theta)]} \right. \\
 &\quad \left. - \frac{E_{\pi_{\delta_0}} [h(\theta)] E_{\pi_{\delta_0}^*} [g(\theta)]}{E_{\pi_{\delta_0}} [g(\theta)] \{E_{\pi_{\delta_0}^*} [h(\theta)]\}^2} \right\}. \tag{2.9}
 \end{aligned}$$

Demostración: Reescribiendo (1.16) como:

$$PBM(\pi) = \mathcal{P}_1(\pi)\mathcal{P}_2(\pi), \tag{2.10}$$

donde

$$\mathcal{P}_1(\pi) = \frac{\int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta}{\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta}, \quad \mathcal{P}_2(\pi) = \frac{\int_{\Theta} h(\theta) \pi_{\delta_0}(\theta) d\theta}{\int_{\Theta} g(\theta) \pi_{\delta_0}(\theta) d\theta},$$

y usando propiedades usuales del operador derivada tenemos que:

$$P\dot{B}M(\delta_0)\delta = \mathcal{P}_2(\pi)\dot{\mathcal{P}}_1(\delta_0)\delta + \mathcal{P}_1(\pi)\dot{\mathcal{P}}_2(\delta_0)u. \tag{2.11}$$

Ahora, teniendo en cuenta el Lema 2.1 nos queda:

$$\begin{aligned}
 \dot{\mathcal{P}}_1(\delta_0)\delta &= \left[\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta \right]^{-2} \left\{ \int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \delta(\theta) d\theta \int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta \right. \\
 &\quad \left. - \int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta \int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \delta(\theta) d\theta \right\} \\
 &= \frac{\int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \delta(\theta) d\theta}{\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta} - \frac{\int_{\Theta} g(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta}{\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta} \frac{\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \delta(\theta) d\theta}{\int_{\Theta} h(\theta) f(\mathbf{x}|\theta) \pi_{\delta_0}(\theta) d\theta} \\
 &= \frac{E_{\pi_{\delta_0}^*} \left[g(\theta) \frac{\delta(\theta)}{\pi_{\delta_0}(\theta)} \right]}{E_{\pi_{\delta_0}^*} [h(\theta)]} - \frac{E_{\pi_{\delta_0}^*} [g(\theta)]}{E_{\pi_{\delta_0}^*} [h(\theta)]} \frac{E_{\pi_{\delta_0}^*} \left[h(\theta) \frac{\delta(\theta)}{\pi_{\delta_0}(\theta)} \right]}{E_{\pi_{\delta_0}^*} [h(\theta)]}.
 \end{aligned}$$

De la misma forma $\dot{\mathcal{P}}_2(\delta_0)\delta$ viene dado por:

$$\dot{\mathcal{P}}_2(\delta_0)\delta = \frac{E_{\pi_{\delta_0}} \left[h(\theta) \frac{\delta(\theta)}{\pi_{\delta_0}(\theta)} \right]}{E_{\pi_{\delta_0}} [g(\theta)]} - \frac{E_{\pi_{\delta_0}} [h(\theta)]}{E_{\pi_{\delta_0}} [g(\theta)]} \frac{E_{\pi_{\delta_0}} \left[g(\theta) \frac{\delta(\theta)}{\pi_{\delta_0}(\theta)} \right]}{E_{\pi_{\delta_0}} [g(\theta)]}.$$

Finalmente, sustituyendo estas expresiones en (2.11) y después de algunos cálculos se obtiene el resultado deseado. ■

Para expresar de una manera mas sencilla la derivada de (1.16) estudiaremos la derivada para $\delta_0 = 0$. La siguiente proposición nos da la derivada de Fréchet de (1.16) en $\delta_0 = 0$ escrita en una forma más sencilla.

Proposición 2.3 *La expresión $P\dot{B}M(0)\delta$ puede reescribirse como*

$$P\dot{B}M(0)\delta = \int_{\Theta} I(\theta)\delta(\theta)d\theta,$$

donde la función de influencia, $I(\theta)$, viene dada por

$$\begin{aligned} I(\theta) = & \left\{ g(\theta) \frac{E_{\pi_0}\{h(\theta)\}}{E_{\pi_0}[g(\theta)]E_{\pi_0^*}[h(\theta)]} \left(1 - \frac{E_{\pi_0^*}[g(\theta)]}{E_{\pi_0}[g(\theta)]} \right) \right. \\ & \left. + h(\theta) \frac{E_{\pi_0^*}[g(\theta)]}{E_{\pi_0}[g(\theta)]E_{\pi_0^*}[h(\theta)]} \left(1 - \frac{E_{\pi_0}[h(\theta)]}{E_{\pi_0^*}[g_2(\theta)]} \right) \right\} \frac{\pi_0^*(\theta)}{\pi_0(\theta)}. \end{aligned} \quad (2.12)$$

Demostración: La derivada en 0 de (2.9) puede reescribirse de la siguiente forma

$$\begin{aligned} P\dot{B}M(0)\delta &= \int_{\Theta} \left[\frac{\int_{\Theta} h(\theta)\pi_0(\theta)d\theta}{\int_{\Theta} g(\theta)\pi_0(\theta)d\theta \int_{\Theta} h(\theta)\pi_0^*(\theta)d\theta} \right] g_1(\theta) \frac{\delta(\theta)}{\pi_0(\theta)} \pi_0^*(\theta)d\theta \\ &- \int_{\Theta} \left[\frac{\int_{\Theta} g(\theta)\pi_0^*(\theta)d\theta \int_{\Theta} h(\theta)\pi_0(\theta)d\theta}{(\int_{\Theta} g(\theta)\pi_0(\theta)d\theta)^2 \int_{\Theta} h(\theta)\pi_0^*(\theta)d\theta} \right] g_1(\theta) \frac{\delta(\theta)}{\pi_0(\theta)} \pi_0^*(\theta)d\theta \\ &+ \int_{\Theta} \left[\frac{\int_{\Theta} g(\theta)\pi_0^*(\theta)d\theta}{\int_{\Theta} g(\theta)\pi_0(\theta)d\theta \int_{\Theta} h(\theta)\pi_0^*(\theta)d\theta} \right] g_2(\theta) \frac{\delta(\theta)}{\pi_0(\theta)} \pi_0^*(\theta)d\theta \\ &- \int_{\Theta} \left[\frac{\int_{\Theta} g(\theta)\pi_0^*(\theta)d\theta \int_{\Theta} h(\theta)\pi_0(\theta)d\theta}{\int_{\Theta} g(\theta)\pi_0(\theta)d\theta (\int_{\Theta} h(\theta)\pi_0^*(\theta)d\theta)^2} \right] g_2(\theta) \frac{\delta(\theta)}{\pi_0(\theta)} \pi_0^*(\theta)d\theta. \end{aligned}$$

Tomando ahora esperanzas y sacando factor común se tiene:

$$\begin{aligned} P\dot{B}M(0)\delta &= \int_{\Theta} \left\{ g(\theta) \frac{E_{\pi_0}\{h(\theta)\}}{E_{\pi_0}[g(\theta)]E_{\pi_0^*}[h(\theta)]} \left(1 - \frac{E_{\pi_0^*}[g(\theta)]}{E_{\pi_0}[g(\theta)]} \right) \right. \\ &+ h(\theta) \frac{E_{\pi_0^*}[g(\theta)]}{E_{\pi_0}[g(\theta)]E_{\pi_0^*}[h(\theta)]} \left(1 - \frac{E_{\pi_0}[h(\theta)]}{E_{\pi_0^*}[g_2(\theta)]} \right) \left. \right\} \frac{\pi_0^*(\theta)}{\pi_0(\theta)} \delta(\theta)d\theta. \end{aligned}$$

Finalmente, la norma de la derivada de Fréchet de (1.16) aparece a continuación

Teorema 2.7 *Bajo las condiciones de la proposición 2.3 se tiene:*

$$\|P\dot{B}M(0)\|_p = \max\{\|I^+(\theta)\|_q, \|I^-(\theta)\|_q\}. \quad (2.13)$$

Demostración: Es inmediato, siguiendo el mismo argumento que el utilizado en el Teorema 2.4 ■

La norma dada en el Teorema 2.7 dependerá obviamente de los valores que tome p . En nuestro caso tomaremos los mismos valores que Gustafson (1996b); esto es, $p = 1, 2, \infty$.

Esta elección no es arbitraria ya que el caso $p = \infty$ corresponde al caso de menor norma, mientras que el caso $p = 1$ al de norma mayor. El caso intermedio $p = 2$ involucra el problema en un espacio de Hilbert.

Como se comprobará en la Proposición 2.4, si $p = 1$, Γ es la clase de contaminación (Gustafson, 1996b) y cuando $p = \infty$ el estudio está conectado con el trabajo de Ruggeri y Wasserman (1993).

Por otra parte, parece necesario ampliar el resultado obtenido en el Teorema 2.4, para los casos particulares $p = 1, 2, \infty$. Para ello usaremos el siguiente corolario.

Corolario 2.1 *La norma de $P\dot{B}M(0)\delta$, $\|P\dot{B}M(0)\|_p$, $p = 1, 2, \infty$, viene dada por*

$$\begin{aligned} \|P\dot{B}M(0)\|_1 &= \max\left\{\sup_{\theta} |I^+(\theta)|, \sup_{\theta} |I^-(\theta)|\right\}, \\ \|P\dot{B}M(0)\|_2 &= \max\left\{\left(\int_{\Theta} I^+(\theta)^2 \pi_0(\theta) d\theta\right)^{1/2}, \left(\int_{\Theta} I^-(\theta)^2 \pi_0(\theta) d\theta\right)^{1/2}\right\}, \\ \|P\dot{B}M(0)\|_{\infty} &= \max\left\{\int_{\Theta} I^+(\theta) \pi_0(\theta) d\theta, \int_{\Theta} I^-(\theta) \pi_0(\theta) d\theta\right\}. \end{aligned} \quad (2.14)$$

Demostración: El resultado se sigue utilizando el Teorema 2.7 para los casos citados. ■

2.4.1. El caso particular de la norma del supremo esencial

Como se acaba de comentar, no existe ninguna restricción para los valores p que proporcionan la norma de la derivada de Fréchet, ya que sólo se les exige que pertenezcan al intervalo $[1, \infty]$. Sin embargo, existen tres valores que tienen una significación especial, teniendo en cuenta que, con respecto a una medida de probabilidad, si $\|\cdot\|_{q_1} \leq \|\cdot\|_{q_2}$ para todo q_1, q_2 tales que $1 \leq q_1 < q_2 \leq \infty$, entonces $L^\infty \subset L^2 \subset L^1$, y por tanto $\Gamma_{L^\infty} \subset \Gamma_{L^2} \subset \Gamma_{L^1}$. El caso $p = \infty$ corresponde a

la norma del supremo esencial, la menor de todas. El caso $p = 2$ sitúa nuestro problema en un espacio de Hilbert, mientras que el caso $p = 1$ corresponde al de mayor norma. Por tanto, parece que con el estudio de la norma de la derivada para estos tres valores de p sería suficiente. De hecho, serán estos tres valores los que se elegirán en las ilustraciones numéricas que desarrollaremos en el capítulo 4.

Por otro lado, obsérvese que si $I(\theta)$ es siempre positiva o siempre negativa, entonces $\|P\dot{B}M(0)\|_\infty$ puede calcularse analíticamente de una manera sencilla de la siguiente forma:

$$\begin{aligned} \|P\dot{B}M(0)\|_\infty &= \left| \frac{E_{\pi_0^x} [g(\theta)]}{E_{\pi_0} [g(\theta)]} \left\{ 1 - \frac{E_{\pi_0^x} [g(\theta)] E_{\pi_0} [h(\theta)]}{E_{\pi_0^x} [h(\theta)] E_{\pi_0} [g(\theta)]} \right\} \right| \\ &= \left| \frac{E_{\pi_0^x} [g(\theta)]}{E_{\pi_0} [g(\theta)]} (1 - PBM) \right|. \end{aligned} \quad (2.15)$$

Luego para estudiar la sensibilidad con respecto a la distribución a priori de una magnitud como en (1.16), quizás sólo sea necesario evaluar (2.15). Esto nos lleva a proponer la siguiente definición.

Definición 2.2 Sea $\alpha \in \mathbb{R}^+$, una magnitud como en (1.16) es α -robusta si

$$\left| \frac{E_{\pi_0^x} [g(\theta)]}{E_{\pi_0} [g(\theta)]} (1 - PBM) \right| < \alpha.$$

A partir de aquí tenemos los siguientes resultados.

Corolario 2.2 Suponiendo que la función de pérdida es cuadrática, y entonces $L(x, \hat{\theta}) = (x - \hat{\theta})^2$, la expresión (1.16) es α -robusta si

$$PBM |1 - PBM| < \alpha.$$

Demostración: Basta tener en cuenta que en este caso

$$\frac{E_{\pi_0^x} [g(\theta)]}{E_{\pi_0} [g(\theta)]} = PBM. \quad \blacksquare$$

Comentario 2.2 Se puede observar que este resultado se cumple independientemente del modelo asumido.

Algunos casos particulares, utilizados en la práctica, se exponen en los siguientes corolarios.

Corolario 2.3 *Bajo el modelo Poisson-gamma, esto es:*

$$f(x|\theta) = e^{-\theta} \theta^x / x!, \quad \theta > 0,$$

$$\pi_0(\theta) = b^a \theta^{a-1} e^{-b\theta} / \Gamma(a), \quad a > 0, b > 0,$$

y la función de pérdida $L(x, \hat{\theta}) = x(x - \hat{\theta})^2$ la magnitud (1.16) es α -robusta si

$$\left| \frac{(a + x + b + t)^2 + a + x}{(b + t)^2} \frac{b^2}{a + (a + b)^2} (1 - PBM) \right| < \alpha.$$

Demostración: Basta tener en cuenta que en este caso la distribución a posteriori es de nuevo una gamma con parámetros $a + x$ y $b + t$ y que $g(\theta) = \theta + 1$, para obtener:

$$\frac{E_{\pi_0^x} [g(\theta)]}{E_{\pi_0} [g(\theta)]} = \frac{(a + x + b + t)^2 + a + x}{(b + t)^2} \frac{b^2}{a + (a + b)^2}.$$

Corolario 2.4 *Bajo el modelo Poisson-gamma y la función de pérdida $L(x, \hat{\theta}) = e^{\alpha x} (x - \hat{\theta})^2$ con $\alpha > 0$, la magnitud (1.16) es α -robusta si*

$$\left| \frac{(a + x)(b - \alpha e^\alpha)}{a(b + t - \alpha e^\alpha)} (1 - PBM) \right| < \alpha.$$

Demostración: Basta tener en cuenta que en este caso $g(\theta) = \theta e^\alpha$, de donde:

$$\frac{E_{\pi_0^x} [g(\theta)]}{E_{\pi_0} [g(\theta)]} = \frac{(a + x)(b - \alpha e^\alpha)}{a(b + t - \alpha e^\alpha)}.$$

2.5. Conclusiones

En este capítulo hemos analizado la sensibilidad de la primas de seguro desde una perspectiva local. En este sentido, hemos llevado a cabo un estudio de la variación de las magnitudes a posteriori de interés perturbando infinitesimalmente la distribución a priori. La derivada del funcional en 0 de (1.12) ó (1.16) se puede expresar en términos de una función de influencia. La medida de robustez viene dada por el máximo de la norma asociada a los espacios L_p de la función de influencia.

Para terminar este capítulo, exponemos algunas conexiones entre nuestro trabajo y el realizado por Gustafson (1996a, 1996b).

En la siguiente proposición se probará que cuando δ viene dada en la forma $\delta = \varepsilon(q - \pi_0)$ con $q \in \mathcal{Q}$, $\varepsilon \in [0, 1]$, tenemos la ya conocida clase de contaminación (Gustafson, 1996, Martín *et al.*, 2003 y Sivaganesan y Berger, 1989; entre otros) y en este caso δ es una medida signada con $\delta(\Theta) = 0$.

Proposición 2.4 *La clase*

$$\Gamma_L^1(\Pi; \alpha) = \left\{ \pi / \left\| \frac{\delta}{\pi}, \Pi \right\|_1 \leq \alpha \right\}$$

coincide con la clase de contaminación

$$\pi = \pi_0 + \delta = \pi_0 + \varepsilon(q - \pi_0).$$

Demostración: Por la definición de $\left\| \frac{\delta}{\pi}, \Pi \right\|_p$, tenemos que

$$\left\| \frac{\delta}{\pi}, \Pi \right\|_1 = \int_{\Theta} \frac{\delta(\theta)}{\pi(\theta)} \pi(\theta) d\theta = \int_{\Theta} \delta(\theta) d\theta = 0 \leq \alpha.$$

Consideremos, sin pérdida de generalidad, la clase normalizada

$$\pi^* = \frac{\pi}{\left\| \frac{\delta}{\pi}, \Pi \right\|_1} = \frac{\pi}{1 + \int_{\Theta} \delta(\theta) d\theta} = \frac{\pi}{1 + \left\| \frac{\delta}{\pi}, \Pi \right\|_1}$$

que puede reescribirse de la siguiente manera:

$$\begin{aligned} \pi^* &= \frac{\pi}{\left\| \frac{\delta}{\pi}, \Pi \right\|_1} = \frac{\pi_0 + \delta}{1 + \left\| \frac{\delta}{\pi}, \Pi \right\|_1} \\ &= \frac{\pi_0}{1 + \left\| \frac{\delta}{\pi}, \Pi \right\|_1} + \frac{\delta}{1 + \left\| \frac{\delta}{\pi}, \Pi \right\|_1} = \left(1 - \frac{\left\| \frac{\delta}{\pi}, \Pi \right\|_1}{1 + \left\| \frac{\delta}{\pi}, \Pi \right\|_1} \right) \pi + \left(1 - \frac{\left\| \frac{\delta}{\pi}, \Pi \right\|_1}{1 + \left\| \frac{\delta}{\pi}, \Pi \right\|_1} \right) \delta \\ &= \left(1 - \frac{1}{\frac{\left\| \frac{\delta}{\pi}, \Pi \right\|_1}{\left\| \frac{\delta}{\pi}, \Pi \right\|_1} + 1} \right) \pi + \left(1 - \frac{1}{\frac{\left\| \frac{\delta}{\pi}, \Pi \right\|_1}{\left\| \frac{\delta}{\pi}, \Pi \right\|_1} + 1} \right) \delta \\ &= \left(1 - \left(1 + \frac{1}{\left\| \frac{\delta}{\pi}, \Pi \right\|_1} \right)^{-1} \right) \pi + \left(1 - \left(1 + \frac{1}{\left\| \frac{\delta}{\pi}, \Pi \right\|_1} \right)^{-1} \right) \delta \\ &= \left(1 - \left(1 + \frac{1}{\left\| \frac{\delta}{\pi}, \Pi \right\|_1} \right)^{-1} \right) \pi + \left(1 - \left(1 + \frac{1}{\left\| \frac{\delta}{\pi}, \Pi \right\|_1} \right)^{-1} \right) \frac{\left\| \frac{\delta}{\pi}, \Pi \right\|_1}{\left\| \frac{\delta}{\pi}, \Pi \right\|_1} \delta \end{aligned}$$

$$\begin{aligned}
&= \left(1 - \left(1 + \frac{1}{\left\|\frac{\delta}{\pi}, \Pi\right\|_1}\right)^{-1}\right) \pi + \left(\left\|\frac{\delta}{\pi}, \Pi\right\|_1 - \frac{\left\|\frac{\delta}{\pi}, \Pi\right\|_1^2}{\left\|\frac{\delta}{\pi}, \Pi\right\|_1 + 1}\right) \frac{\delta}{\left\|\frac{\delta}{\pi}, \Pi\right\|_1} \\
&= \left(1 - \left(1 + \frac{1}{\left\|\frac{\delta}{\pi}, \Pi\right\|_1}\right)^{-1}\right) \pi + \left(\frac{\left\|\frac{\delta}{\pi}, \Pi\right\|_1}{\left\|\frac{\delta}{\pi}, \Pi\right\|_1 + 1}\right) \frac{\delta}{\left\|\frac{\delta}{\pi}, \Pi\right\|_1} \\
&= \left(1 - \left(1 + \frac{1}{\left\|\frac{\delta}{\pi}, \Pi\right\|_1}\right)^{-1}\right) \pi + \left(1 + \frac{1}{\left\|\frac{\delta}{\pi}, \Pi\right\|_1}\right)^{-1} \frac{\delta}{\left\|\frac{\delta}{\pi}, \Pi\right\|_1}.
\end{aligned}$$

Puesto que $\left\|\frac{\delta}{\pi}, \Pi\right\|_1 \leq \alpha$ se tiene que

$$\frac{1}{1 + \frac{1}{\left\|\frac{\delta}{\pi}, \Pi\right\|_1}} \leq \frac{1}{1 + \frac{1}{\alpha}}.$$

Ahora, si llamamos $\varepsilon = \frac{1}{1 + \frac{1}{\left\|\frac{\delta}{\pi}, \Pi\right\|_1}}$ se tiene que

$$\varepsilon \leq \frac{1}{1 + \frac{1}{\alpha}} \Leftrightarrow \varepsilon \leq \frac{1}{\frac{\alpha+1}{\alpha}} \Leftrightarrow \varepsilon \leq \frac{\alpha}{\alpha+1}.$$

Por tanto, para todo ε que verifique la desigualdad anterior se tiene que π^* , normalizado, es la clase de contaminación $\pi = (1 - \varepsilon)\pi_0 + \varepsilon\delta$. ■

Finalmente haremos una breve reseña al caso de incorporar la simetría en las distribuciones a priori unimodales. Esto puede conseguirse a partir del Lema 2.2.1 en Sivaganesan y Berger (1989) intercambiando en (2.8)

$$\frac{1}{2} \int_{\theta_0}^{\theta_0+z} \left[g(\theta) - h(\theta) \frac{\rho_g}{\rho_h} \right] \frac{\pi_0^x(\theta)}{\pi_0(\theta)} d\theta$$

por

$$\frac{1}{2z} \int_{\theta_0-z}^{\theta_0+z} \left[g(\theta) - h(\theta) \frac{\rho_g}{\rho_h} \right] \frac{\pi_0^x(\theta)}{\pi_0(\theta)} d\theta.$$

Capítulo 3

Inferencia

La mezcla de distribuciones constituye una de las vías más utilizadas para obtener nuevas distribuciones de probabilidad. En este sentido y buscando una alternativa más flexible a la distribución de Poisson, especialmente bajo el fenómeno de sobredispersión (varianza mayor que la media), proponemos en este capítulo una reparametrización de las distribuciones binomial negativa y binomial que nos permitirá mezclar cada una de ellas con la distribución inversa gaussiana. Esta mezcla se realiza de una manera sencilla y nos permite obtener dos nuevas distribuciones de probabilidad discretas. El modelo resultante, como veremos, tiene unas propiedades atractivas que la hacen adecuado para aplicarlo no sólo en el mercado de seguros sino también en otros campos científicos. Para ambas distribuciones daremos expresiones recursivas para el cálculo de las probabilidades y para la distribución de la cantidad total reclamada. Como veremos, es posible obtener expresiones cerradas para estimar los parámetros.

3.1. Modelo binomial negativa–inversa gaussiana

En estadística actuarial es habitual considerar que dada una distribución de probabilidad discreta $\Pr \{X = x|\theta\}$ el parámetro θ se distribuya a su vez de acuerdo a cierta distribución de probabilidad $\pi(\theta)$. De esta manera, estamos asumiendo que dicho parámetro es aleatorio reflejando con ello la heterogeneidad de una población de individuos. Por otro lado, la distribución marginal $\Pr \{X = x\}$ viene dada por

$$\Pr \{X = x\} = \int_{\Theta} \Pr \{X = x|\theta\} \pi(\theta) d\theta$$

y refleja la distribución de X sobre toda la población, independientemente del parámetro θ .

Esta metodología constituye uno de los mecanismos habituales de generación de nuevas distribuciones de probabilidad, discretas y continuas, y que es utilizada en muchas disciplinas científicas, además de en la estadística actuarial. Así, en esta disciplina, los individuos lo constituyen los contratos de seguros dentro de una población de individuos que es la cartera de seguros.

Por ejemplo, es bien sabido que la distribución binomial negativa (Klugman et al., 1998; Lemaire, 1979 y Simon, 1961) se obtiene mezclando una distribución de Poisson con parámetro $\theta > 0$ con una distribución gamma con parámetros $a > 0$ y $b > 0$ de la manera siguiente

$$\begin{aligned}\Pr\{X = x\} &= \int_{\Theta} \Pr\{X = x|\theta\} \pi(\theta) d\theta \\ &= \binom{r+x-1}{x} p^r (1-p)^x, \quad x = 0, 1, 2, \dots\end{aligned}$$

siendo

$$\begin{aligned}\Pr\{X = x|\theta\} &= \frac{e^{-\theta} \theta^x}{x!}, \quad x = 0, 1, 2, \dots, \\ \pi(\theta) &= \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}, \quad \theta > 0, \\ p &= \frac{b}{b+1}.\end{aligned}$$

Otras mezclas de distribuciones de probabilidad que han sido utilizadas en la disciplina actuarial son las siguientes:

- Mezcla de una distribución binomial negativa con Pareto tipo II: aparece en Klugman et al. (1998), Meng et al. (1999), Gómez y Vázquez (2003).
- Mezcla de una distribución de Poisson con la distribución inversa gaussiana: aparece en Gómez y Vázquez (2003) y Tremblay (1992).
- Mezcla de una distribución de Poisson generalizada con una gamma generalizada: aparece en Sarabia et al. (2004).
- Mezcla de una distribución binomial negativa con una gamma: aparece en Gómez et al. (2006a) en el contexto de un análisis bayesiano jerárquico.
- Mezcla de una distribución exponencial con una inversa gaussiana: aparece en Frangos y Karlis (2004) para modelizar cantidades de reclamaciones.

En esta sección introduciremos una nueva distribución de probabilidad obtenida a partir de una mezcla de dos distribuciones, las distribuciones binomial negativa y la inversa gaussiana, para dar lugar a una distribución, que hemos denominado como, binomial negativa inversa gaussiana, (*BNIG*). La nueva distribución

$BNIG$ tiene colas gruesas, parece unimodal y puede ser sesgada a la derecha o no dependiendo de los valores de los parámetros. Por otra parte, presenta la ventaja de admitir expresiones cerradas para estimar los parámetros del modelo mediante el método de los momentos, así como obtenerse una expresión recursiva para calcular su función de probabilidad, que permite obtener una fórmula recursiva para la distribución de la cantidad total reclamada, cuando se supone que el número de siniestros sigue la distribución $BNIG$. La variable aleatoria cantidad total reclamada juega un papel destacado en muchos problemas actuariales, especialmente en el modelo colectivo de riesgo y en teoría de la ruina. Finalmente, la distribución $BNIG$ se puede extender de una manera sencilla al caso multivariado que admite correlación estrictamente positiva.

3.1.1. Resultados básicos

Empezaremos esta sección recordando que una variable aleatoria Z sigue una distribución inversa gaussiana si su función de densidad viene dada por,

$$f(z; \mu, \psi) = \left(\frac{\psi}{2\pi z^3} \right)^{1/2} \exp \left\{ -\frac{\psi(z - \mu)^2}{2\mu^2 z} \right\}, \quad z > 0, \quad (3.1)$$

donde $\psi, \mu > 0$.

Si representamos la distribución inversa gaussiana por $Z \sim IG(\mu, \psi)$, no resulta difícil demostrar (ver Tweedie, 1957 y Chhikara y Folks, 1988) que la función generatriz de momentos de Z viene dada por

$$M_Z(t) = \mathbb{E}(e^{tZ}) = \exp \left[\frac{\psi}{\mu} \left(1 - \sqrt{1 - 2\mu^2 t / \psi} \right) \right]. \quad (3.2)$$

A continuación se define la distribución $BNIG$.

Definición 3.1 Una variable aleatoria X sigue una distribución binomial negativa-inversa gaussiana si admite la siguiente representación estocástica:

$$X | \theta \sim \mathcal{BN}(r, p = e^{-\theta}), \quad (3.3)$$

$$\theta \sim IG(\mu, \psi), \quad (3.4)$$

con $r, \mu, \psi > 0$, siendo $\mathcal{BN}(r, p = e^{-\theta})$ la binomial negativa de parámetros r y $e^{-\theta}$ dada por

$$\Pr(X = x) = \binom{r+x-1}{x} p^r (1-p)^x, \quad x = 0, 1, \dots$$

A partir de aquí denotaremos a esta distribución mediante $X \sim BNIG(r, \mu, \psi)$.

Obsérvese que, puesto que $\theta \in (0, +\infty)$ entonces $p \in (0, 1)$.

El siguiente Teorema establece las propiedades básicas de esta nueva distribución.

Teorema 3.1 Sea $X \sim \text{BNIG}(r, \mu, \psi)$ una distribución binomial negativa-inversa gaussiana definida en (3.3)-(3.4). Algunas propiedades básicas son:

(a) La función de probabilidad viene dada por,

$$\Pr(X = x) = \binom{r+x-1}{x} \sum_{j=0}^x (-1)^j \binom{x}{j} \exp \left\{ \frac{\psi}{\mu} \left[1 - \sqrt{1 + \frac{2(r+j)\mu^2}{\psi}} \right] \right\}, \quad (3.5)$$

con $x = 0, 1, 2, \dots$ y $r, \mu, \psi > 0$.

(b) El momento factorial de orden k viene dado por

$$\begin{aligned} \mu_{[k]}(X) &= \mathbb{E}[X(X-1)\cdots(X-k+1)] \\ &= \frac{\Gamma(r+k)}{\Gamma(r)} \sum_{j=0}^k (-1)^j \binom{k}{j} \exp \left\{ \frac{\psi}{\mu} \left[1 - \sqrt{1 - \frac{2(k-j)\mu^2}{\psi}} \right] \right\}, \end{aligned} \quad (3.6)$$

con $k = 1, 2, \dots$

(c) La media, momento de segundo y tercer orden y la varianza son,

$$\mathbb{E}(X) = r[M_\theta(1) - 1], \quad (3.7)$$

$$\mathbb{E}(X^2) = (r+r^2)M_\theta(2) - (r+2r^2)M_\theta(1) + r^2, \quad (3.8)$$

$$\mathbb{E}(X^3) = (r+r^2)M_\theta(2) - (r+2r^2)M_\theta(1) + r^2, \quad (3.9)$$

$$\text{Var}(X) = (r+r^2)M_\theta(2) - rM_\theta(1) - r^2M_\theta^2(1),$$

donde $M_\theta(u)$ es como en (3.2).

Demostración:

(a) Si $X|\theta \sim \text{BN}(r, e^{-\theta})$ y $\theta \sim \text{IG}(\mu, \psi)$, la función de probabilidad de X puede obtenerse usando la conocida fórmula,

$$\Pr(X = x) = \int_0^\infty \Pr(X = x|\theta) f(\theta; \mu, \psi) d\theta,$$

donde $f(\theta; \mu, \psi)$ es la función de probabilidad de una distribución inversa gaussiana definida en (3.1). Esta distribución puede expresarse en una forma cerrada a partir de:

$$\begin{aligned}
\Pr(X = x) &= \int_0^\infty \binom{r+x-1}{x} p^r q^x f(\theta; \mu, \psi) d\theta \\
&= \int_0^\infty \binom{r+x-1}{k} (e^{-\theta})^r (1 - e^{-\theta})^x f(\theta; \mu, \psi) d\theta.
\end{aligned}$$

Ahora haciendo uso del desarrollo del binomio de Newton podemos reescribir la última expresión de la forma siguiente

$$\begin{aligned}
\Pr(X = x) &= \int_0^\infty \binom{r+x-1}{x} (e^{-\theta})^r \sum_{j=0}^x \binom{x}{j} (-1)^j (e^{-\theta})^j f(\theta; \mu, \psi) d\theta \\
&= \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j \int_0^\infty (e^{-\theta})^r (e^{-\theta})^j f(\theta; \mu, \psi) d\theta \\
&= \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j \int_0^\infty (e^{-\theta})^{r+j} f(\theta; \mu, \psi) d\theta \\
&= \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j \int_0^\infty e^{-\theta(r+j)} f(\theta; \mu, \psi) d\theta \\
&= \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_\theta(-(r+j)).
\end{aligned}$$

finalmente, reemplazando $M_\theta(-(r+j))$ por la expresión dada en (3.2) para $t = -(r+j)$ se obtiene el resultado.

(b) El momento factorial de una distribución binomial negativa es (ver Balakrishnan y Nevzorov (2003)):

$$\mu_{[k]}(X|\theta) = \frac{\Gamma(r+k)}{\Gamma(r)} (e^\theta - 1)^k, \quad k = 1, 2, \dots$$

Si desarrollamos por el binomio de Newton el último factor tenemos,

$$(e^\theta - 1)^k = \sum_{j=0}^k (-1)^j \binom{k}{j} e^{\theta(k-j)}, \quad k = 1, 2, \dots$$

Finalmente, tomando esperanza matemática y combinando esta expresión con (3.2) obtenemos:

$$\begin{aligned}\mu_{|k|}(X) &= \mathbb{E}[X(X-1)\cdots(X-k+1)] = \int_0^\infty \mu_{|k|}(X|\theta) f(\theta; \mu, \psi) d\theta \\ &= \frac{\Gamma(r+k)}{\Gamma(r)} \int_0^\infty \sum_{j=0}^k (-1)^j \binom{k}{j} e^{\theta(k-j)} f(\theta; \mu, \psi) d\theta \\ &= \frac{\Gamma(r+k)}{\Gamma(r)} \sum_{j=0}^k (-1)^j \binom{k}{j} \exp \left\{ \frac{\psi}{\mu} \left[1 - \sqrt{1 - \frac{2(k-j)\mu^2}{\psi}} \right] \right\}.\end{aligned}$$

(c) Para calcular los momentos de primer y segundo orden usaremos la siguiente relación:

$$\mathbb{E}(X^r) = \mathbb{E}_\theta[\mathbb{E}(X^r|\theta)],$$

de donde:

$$\mathbb{E}(X^3) = (r^3 + 3r^2 + 2r)M_\theta(3) - (3r + 6r^2 + 3r^3)M_\theta(2) + (r + 3r^2 + 3r^3)M_\theta(1) - r^3,$$

Así el momento de primer orden es:

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}_\theta[\mathbb{E}(X|\theta)] \\ &= \mathbb{E}_\theta \left(r \frac{1-p}{p} \right) = \mathbb{E}_\theta \left(r \frac{1-e^{-\theta}}{e^{-\theta}} \right) \\ &= \int_0^\infty r \frac{1-e^{-\theta}}{e^{-\theta}} f(\theta; \mu, \psi) d\theta = \int_0^\infty r \left[\frac{1}{e^{-\theta}} - \frac{e^{-\theta}}{e^{-\theta}} \right] f(\theta; \mu, \psi) d\theta \\ &= r \int_0^\infty [e^\theta - 1] f(\theta; \mu, \psi) d\theta = r \int_0^\infty e^\theta f(\theta; \mu, \psi) d\theta - r \int_0^\infty f(\theta; \mu, \psi) d\theta \\ &= r \int_0^\infty e^\theta f(\theta; \mu, \psi) d\theta - r = r [M_\theta(1) - 1],\end{aligned}$$

el momento de segundo orden:

$$\begin{aligned}\mathbb{E}(X^2) &= \mathbb{E}_\theta[\mathbb{E}(X^2|\theta)] \\ &= \mathbb{E}_\theta \left(r \frac{1-p}{p^2} + r^2 \frac{(1-p)^2}{p^2} \right) = r \mathbb{E}_\theta \left(\frac{1-p}{p^2} \right) + r^2 \mathbb{E}_\theta \left(\frac{(1-p)^2}{p^2} \right)\end{aligned}$$

$$\begin{aligned}
&= r \int_0^\infty \frac{1 - e^{-\theta}}{e^{-2\theta}} f(\theta; \mu, \psi) d\theta + r^2 \int_0^\infty \frac{(1 - e^{-\theta})^2}{e^{-2\theta}} f(\theta; \mu, \psi) d\theta \\
&= r \int_0^\infty [e^{2\theta} - e^\theta] f(\theta; \mu, \psi) d\theta + r^2 \int_0^\infty [e^{2\theta} - 2e^\theta + 1] f(\theta; \mu, \psi) d\theta \\
&= r \int_0^\infty e^{2\theta} f(\theta; \mu, \psi) d\theta - r \int_0^\infty e^\theta f(\theta; \mu, \psi) d\theta \\
&+ r^2 \int_0^\infty e^{2\theta} f(\theta; \mu, \psi) d\theta - 2r^2 \int_0^\infty e^\theta f(\theta; \mu, \psi) d\theta + r^2 \\
&= rM_\theta(2) - rM_\theta(1) + r^2M_\theta(2) - 2r^2M_\theta(1) + r^2 \\
&= (r + r^2) M_\theta(2) - (r + 2r^2) M_\theta(1) + r^2,
\end{aligned}$$

y el de tercer orden:

$$\begin{aligned}
\mathbb{E}(X^3) &= \mathbb{E}_\theta[\mathbb{E}(X^3|\theta)] \\
&= \mathbb{E}_\theta \left(r \frac{1-p}{p^2} + r^2 \frac{(1-p)^2}{p^2} \right) = r \mathbb{E}_\theta \left(\frac{1-p}{p^2} \right) + r^2 \mathbb{E}_\theta \left(\frac{(1-p)^2}{p^2} \right) \\
&= r \int_0^\infty \frac{1 - e^{-\theta}}{e^{-2\theta}} f(\theta; \mu, \psi) d\theta + r^2 \int_0^\infty \frac{(1 - e^{-\theta})^2}{e^{-2\theta}} f(\theta; \mu, \psi) d\theta \\
&= r \int_0^\infty [e^{2\theta} - e^\theta] f(\theta; \mu, \psi) d\theta + r^2 \int_0^\infty [e^{2\theta} - 2e^\theta + 1] f(\theta; \mu, \psi) d\theta \\
&= r \int_0^\infty e^{2\theta} f(\theta; \mu, \psi) d\theta - r \int_0^\infty e^\theta f(\theta; \mu, \psi) d\theta \\
&+ r^2 \int_0^\infty e^{2\theta} f(\theta; \mu, \psi) d\theta - 2r^2 \int_0^\infty e^\theta f(\theta; \mu, \psi) d\theta + r^2 \\
&= rM_\theta(2) - rM_\theta(1) + r^2M_\theta(2) - 2r^2M_\theta(1) + r^2 \\
&= (r + r^2) M_\theta(2) - (r + 2r^2) M_\theta(1) + r^2.
\end{aligned}$$

Para calcular la varianza de la distribución \mathcal{BNIG} sólo basta tener en cuenta

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X],$$

o bien,

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X|\theta)] - \text{Var}[\mathbb{E}(X|\theta)].$$

■

En el siguiente teorema se prueba que la distribución \mathcal{BNTG} es sobredispersa, para ello se ha comparado los momentos de la nueva distribución con una distribución binomial negativa con la misma media.

Teorema 3.2 Sea $\theta \sim \mathcal{IG}(\mu, \psi)$ y $\tilde{X} \sim \mathcal{BN}(r, p = [\mathbb{E}(e^\theta)]^{-1})$. Consideremos una distribución binomial negativa-inversa gaussiana X definida en (3.3)-(3.4). Entonces:

(a) Se cumple que: $\mathbb{E}(\tilde{X}) = \mathbb{E}(X)$ y $\text{Var}(X) > \mathbb{E}(X)$.

(b) $\text{Var}(X) > \mathbb{E}(X)$.

Demostración: Obviamente $\mathbb{E}(e^\theta) > 1$, entonces $p = 1/\mathbb{E}(e^\theta)$ está bien definido. Ahora, $\mathbb{E}(X) = \mathbb{E}[\mathbb{E}(X|\theta)] = r[\mathbb{E}(e^\theta) - 1]$, y de $\text{Var}(X) = \mathbb{E}[\text{Var}(X|\theta)] + \text{Var}[\mathbb{E}(X|\theta)]$, obtenemos que $\text{Var}(X) = r[\mathbb{E}(e^{2\theta}) - \mathbb{E}(e^\theta)] + r^2 \text{Var}(e^\theta)$. De $\tilde{X} \sim \mathcal{BN}(r, p = [\mathbb{E}(e^\theta)]^{-1})$, tenemos que $\mathbb{E}(\tilde{X}) = r[\mathbb{E}(e^\theta) - 1]$ y $\text{Var}(\tilde{X}) = r[\mathbb{E}(e^\theta) - 1]\mathbb{E}(e^\theta)$. Ahora,

$$\begin{aligned} \text{Var}(X) - \text{Var}(\tilde{X}) &= \mathbb{E}_\theta[\text{Var}(X|\theta)] + \text{Var}_\theta[\mathbb{E}(X|\theta)] - \text{Var}(\tilde{X}) \\ &= \mathbb{E}[r(e^\theta - 1)e^\theta] + \text{Var}[r(e^\theta - 1)] - \text{Var}(\tilde{X}) \\ &= r[\mathbb{E}(e^{2\theta}) - (\mathbb{E}(e^\theta))^2] + r^2 \text{Var}(e^\theta) \\ &= (r + r^2) \text{Var}(e^\theta) > 0. \end{aligned}$$

Finalmente, (b) es consecuencia directa de (a). ■

La Figura 3.1.1 muestra varios ejemplos de la función de probabilidad de la distribución \mathcal{BNTG} para diferentes valores de r , μ y ψ . Como puede apreciarse, todas ellas parecen ser unimodales.

3.1.2. Relaciones recursivas para las probabilidades

En esta sección vamos a obtener una expresión recursiva para la función de probabilidad de una distribución \mathcal{BNTG} . Esta relación recursiva será el trampolín desde el que se podrá obtener de manera exacta y recursiva la función de densidad de probabilidad de la variable aleatoria cantidad total reclamada, que viene dada por

$$S = \sum_{n=0}^{\infty} \text{Pr}(N = n) f^{*n}(x)$$

la cual denotaremos mediante $g_s(x)$. Aquí $\text{Pr}(N = n)$ es la función de probabilidad de la variable aleatoria número de reclamaciones, $f(x)$ la función de densidad de probabilidad asociada a la cuantía de las reclamaciones y $f^{*n}(x)$ es la convolución n -ésima de $f(x)$.

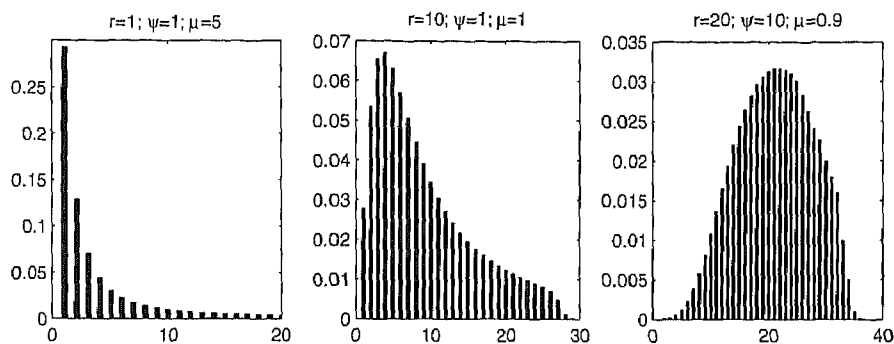


Figura 3.1: Algunos ejemplos de la función de masa de probabilidad de la distribución \mathcal{BNTG} para diferentes valores de r , ψ y μ

Como se mostrará más adelante, será posible obtener de manera exacta y recursiva la función de densidad de $g_s(x)$. Para el caso en que x sea continua la fórmula recursiva que se obtiene aparece expresada en términos de una ecuación integral similar a la que aparece en Panjer (1981).

Por simplicidad denotaremos a la distribución del número de reclamaciones como $f_r(x)$.

Teorema 3.3 Las probabilidades de una variable aleatoria que sigue una distribución \mathcal{BNTG} puede obtenerse de forma recursiva mediante:

$$f_r(x) = \frac{r+x-1}{x} \left[f_r(x-1) - \frac{r}{r+x-1} f_{r+1}(x-1) \right], \quad x = 1, 2, \dots \quad (3.10)$$

Demostración: Para la distribución binomial negativa con función de probabilidad

$$f_r(x|\theta) = \binom{r+x-1}{x} e^{-\theta r} (1 - e^{-\theta})^x, \quad x = 0, 1, \dots$$

tenemos la siguiente recursión

$$\frac{f_r(x|\theta)}{f_r(x-1|\theta)} = \frac{r+x-1}{x} (1 - e^{-\theta}), \quad x = 1, 2, \dots \quad (3.11)$$

Usando la definición de una distribución \mathcal{BNTG} y (3.11) tenemos

$$\begin{aligned} f_r(x) &= \int_0^\infty f_r(x|\theta) f(\theta) d\theta \\ &= \int_0^\infty \frac{r+x-1}{x} (1 - e^{-\theta}) f_r(x-1|\theta) f(\theta) d\theta \end{aligned}$$

$$= \frac{r+x-1}{x} \left[f_r(x-1) - \int_0^\infty e^{-\theta} f_r(x-1|\theta) f(\theta) d\theta \right].$$

Ahora:

$$\begin{aligned} \int_0^\infty e^{-\theta} f_r(x-1|\theta) f(\theta) d\theta &= \int_0^\infty \binom{r+x-2}{x-1} e^{-\theta(r+1)} (1-e^{-\theta})^{x-1} f(\theta) d\theta \\ &= \int_0^\infty \frac{r}{r+x-1} \binom{r+1+x-2}{x-1} \\ &\times e^{-\theta(r+1)} (1-e^{-\theta})^{x-1} f(\theta) d\theta \\ &= \frac{r}{r+x-1} f_{r+1}(x-1), \end{aligned}$$

de donde resulta (3.10). ■

Desafortunadamente, no parece posible una fórmula recursiva para $f_{r+1}(x)$ expresada en términos de $f_r(x)$.

El siguiente resultado proporciona una fórmula exacta y recursiva, en términos de una ecuación integral, para la función de densidad de la cantidad total reclamada.

Teorema 3.4 *Si la variable aleatoria cantidad total reclamada es continua con función de densidad de probabilidad $f(x)$, $x > 0$, entonces la función de densidad de la cantidad total reclamada cuando la distribución del número de reclamaciones es \mathcal{BNIG} verifica la siguiente ecuación integral:*

$$g_s(x; r) = p_r(0) + \int_0^x \frac{ry + x - y}{x} g_s(x-y; r) f(y) dy - \int_0^x \frac{ry}{x} g_s(x-y; r+1) f(y) dy. \quad (3.12)$$

donde $p_r(k)$ es la distribución del número de reclamaciones.

Demostración: La cantidad total reclamada viene dada por

$$g_s(x; r) = \sum_{k=0}^{\infty} p_r(k) f^{k*}(x) = p_r(0) f^{0*}(x) + \sum_{k=1}^{\infty} p_r(k) f^{k*}(x),$$

donde f^{k*} es la k -ésima convolución de $f(x)$.

Usando (3.10) tenemos:

$$p_r(k) = \left(\frac{r}{k} + 1 - \frac{1}{k} \right) p_r(k-1) - \frac{r}{k} p_{r+1}(k-1), \quad k = 1, 2, \dots$$

Entonces

$$\sum_{k=1}^{\infty} p_r(k) f^{k*}(x) = \sum_{k=1}^{\infty} \frac{r}{k} p_r(k-1) f^{k*}(x) + \sum_{k=1}^{\infty} p_r(k-1) f^{k*}(x)$$

$$- \sum_{k=1}^{\infty} \frac{1}{k} p_r(k-1) f^{k*}(x) - \sum_{k=1}^{\infty} \frac{r}{k} p_{r+1}(k-1) f^{k*}(x).$$

Ahora, después de algunos cálculos y usando las identidades

$$f^{k*}(x) = \int_0^x f^{(k-1)*}(x-y) f(y) dy, \quad k = 1, 2, \dots, \quad (3.13)$$

$$\frac{f^{k*}(x)}{k} = \int_0^x \frac{y}{x} f^{(k-1)*}(x-y) f(y) dy, \quad k = 1, 2, \dots \quad (3.14)$$

obtenemos el resultado. ■

Es sencillo comprobar que (3.13) y (3.14) se satisfacen intercambiando el símbolo integral por el de suma, lo que permite obtener una expresión exacta y recursiva para la distribución total reclamada para el caso en que la variable aleatoria cantidad total reclamada sea discreta. Esto se pone de manifiesto en el siguiente resultado.

Corolario 3.1 *Si la distribución de la cantidad reclamada es discreta y positiva, la distribución de la cantidad total reclamada satisface la siguiente fórmula recursiva:*

$$g_s(x; r) = p_r(0) + \sum_{y=0}^x \frac{ry + x - y}{x} g_s(x - y; r) f(y) - \sum_{y=0}^x \frac{ry}{x} g_s(x - y; r + 1) f(y).$$

Demostración: Basta sustituir en (3.12) el símbolo integral por el de suma. ■

3.1.3. Distribución a posteriori

Puede obtenerse de una manera sencilla la distribución a posteriori del parámetro utilizando el siguiente resultado.

Teorema 3.5 *Sea $\text{Pr}(\mathbf{x}|\theta)$ la función de verosimilitud de una distribución $\text{BN}(r, p = e^{-\theta})$ cuyo parámetro θ sigue una distribución a priori $\text{IG}(\mu, \psi)$ que denotaremos por $f(\theta)$ entonces la distribución a posteriori del modelo BNIG viene expresada por*

$$f(\theta|\mathbf{x}) = \frac{\sum_{j=0}^{\mathbf{x}} (-1)^j \binom{\mathbf{x}}{j} \sqrt{\frac{\psi}{2\pi\theta^3}} \exp \left\{ -\theta(j + rt) - \frac{\psi(\theta - \mu)^2}{2\mu^2\theta} \right\}}{\sum_{j=0}^{\mathbf{x}} (-1)^j \binom{\mathbf{x}}{j} \exp \left\{ \frac{\psi}{\mu} \left[1 - \sqrt{1 + \frac{2(rt + j)\mu^2}{\psi}} \right] \right\}},$$

donde $\mathbf{x} = \sum_{i=1}^t x_i = t\bar{x}$

Demostración: Dividendo la distribución conjunta entre la distribución marginal tenemos:

$$\begin{aligned}
 f(\theta|\mathbf{x}) &= \frac{\Pr(\mathbf{x}|\theta)f(\theta)}{\int_0^{\infty} \Pr(\mathbf{x}|\theta)f(\theta)d\theta} = \frac{\prod_{i=1}^l \binom{r+x_i-1}{x_i} [e^{-\theta}]^r [1-e^{-\theta}]^{x_i} f(\theta)}{\int_{\Theta} \prod_{i=1}^l \binom{r+x_i-1}{x_i} [e^{-\theta}]^r [1-e^{-\theta}]^{x_i} f(\theta)d\theta} \\
 &= \frac{(1-e^{-\theta})^{\mathbf{x}} e^{-\theta r t} f(\theta)}{\int_{\Theta} (1-e^{-\theta})^{\mathbf{x}} e^{-\theta r t} f(\theta)d\theta} = \frac{\sum_{j=0}^{\mathbf{x}} (-1)^j \binom{\mathbf{x}}{j} [e^{-\theta}]^j e^{-\theta r t} f(\theta)}{\int_{\Theta} \sum_{j=0}^{\mathbf{x}} (-1)^j \binom{\mathbf{x}}{j} [e^{-\theta}]^j e^{-\theta r t} f(\theta)d\theta} \\
 &= \frac{\sum_{j=0}^{\mathbf{x}} (-1)^j \binom{\mathbf{x}}{j} e^{-\theta(j+rt)} f(\theta)}{\sum_{j=0}^{\mathbf{x}} (-1)^j \binom{\mathbf{x}}{j} \int_{\Theta} e^{-\theta(j+rt)} f(\theta)d\theta} = \frac{\sum_{j=0}^{\mathbf{x}} (-1)^j \binom{\mathbf{x}}{j} e^{-\theta(j+rt)} f(\theta)}{\sum_{j=0}^{\mathbf{x}} (-1)^j \binom{\mathbf{x}}{j} M_{\theta}(-(j+rt))}.
 \end{aligned}$$

y sustituyendo en la última expresión (3.1) y (3.2) se obtiene el resultado deseado. ■

3.1.4. Versión multivariada

Ahora propondremos la siguiente definición de una distribución multivariada \mathcal{BNIG} , la cual es una extensión natural de la definición (3.3)–(3.4). La versión multivariada de la distribución \mathcal{BNIG} puede considerarse como la mezcla de distribuciones $\mathcal{BN}(r_i, p = e^{-\theta})$, $i = 1, 2, \dots, d$ independientes con una distribución inversa gaussiana.

Definición 3.2 Una distribución multivariada $\mathcal{BNIG}(X_1, \dots, X_d)$ se define por la siguiente representación estocástica:

$$\begin{aligned}
 X_i|\theta &\sim \mathcal{BN}(r_i, e^{-\theta}), \quad i = 1, \dots, d \text{ independientes,} \\
 \theta &\sim \mathcal{IG}(\mu, \psi).
 \end{aligned}$$

La función de distribución conjunta es $(x_1, x_2, \dots, x_d = 0, 1, 2, \dots)$:

$$\Pr(X_1 = x_1, X_2 = x_2, \dots, X_d = x_d)$$

$$= \prod_{i=1}^d \binom{r_i + x_i - 1}{x_i} \sum_{j=0}^{\tilde{x}} (-1)^j \binom{\tilde{x}}{j} \exp \left\{ \frac{\psi}{\mu} \left[1 - \sqrt{1 + \frac{2(\tilde{r} + j)\mu^2}{\psi}} \right] \right\}, \quad (3.15)$$

donde $\mu, \psi, r_1, \dots, r_d > 0$ y

$$\begin{aligned} \tilde{r} &= r_1 + \dots + r_d, \\ \tilde{x} &= x_1 + \dots + x_d. \end{aligned} \quad (3.16)$$

Esta última fórmula admite una expresión más sencilla que puede resultar conveniente para calcular las probabilidades. Siendo la distribución univariada de la \mathcal{BNIG} definida como $\tilde{Z} \sim \mathcal{BNIG}(\tilde{r}, \mu, \psi)$, donde \tilde{r} viene dada en (3.16), entonces una expresión alternativa para (3.15) es

$$\Pr(X_1 = x_1, X_2 = x_2, \dots, X_d = x_d) = \frac{\prod_{i=1}^d \binom{r_i + x_i - 1}{x_i}}{\binom{\tilde{r} + \tilde{x} - 1}{\tilde{x}}} \cdot \Pr(\tilde{Z} = \tilde{x}),$$

con $d \geq 2$.

Obviamente, las marginales son $X_i \sim \mathcal{BNIG}(r_i, \mu, \psi)$, $i = 1, 2, \dots, d$ y cualquier subvector (X_1, \dots, X_s) con $s < d$ es nuevamente una distribución multivariada \mathcal{BNIG} de dimensión s .

El siguiente resultado establece las características más importantes de esta nueva distribución de probabilidad.

Teorema 3.6 *Los momentos de primer y segundo orden y la covarianza de la distribución \mathcal{BNIG} multivariada vienen dados por:*

$$\mathbb{E}(X_i) = r_i [M_\theta(1) - 1], \quad i = 1, 2, \dots, \quad (3.17)$$

$$\text{Var}(X) = (r_i + r_i^2)M_\theta(2) - r_i M_\theta(1) - r_i^2 M_\theta^2(1), \quad i = 1, 2, \dots, \quad (3.18)$$

$$\text{Cov}(X_i, X_j) = r_i r_j [M_\theta(2) - M_\theta^2(1)], \quad i \neq j. \quad (3.19)$$

Demstración: Las expresiones (3.17) y (3.18) se obtienen de manera inmediata utilizando esperanzas condicionadas. Para obtener (3.19) se requiere previamente el cálculo de:

$$\mathbb{E}(X_i X_j)$$

$$= \mathbb{E}_\theta[\mathbb{E}(X_i X_j | \theta)]$$

$$\begin{aligned}
&= \mathbb{E}_\theta \left(r_i r_j \frac{1-p}{p} \frac{1-p}{p} \right) = r_i r_j \mathbb{E}_\theta \left(\frac{(1-p)^2}{p^2} \right) \\
&= r_i r_j \int_0^\infty \frac{(1-e^{-\theta})^2}{e^{-2\theta}} f(\theta; \mu, \psi) d\theta = r_i r_j \int_0^\infty \frac{1-e^{-\theta} + e^{-2\theta}}{e^{-2\theta}} f(\theta; \mu, \psi) d\theta \\
&= r_i r_j \int_0^\infty e^{2\theta} f(\theta; \mu, \psi) d\theta - r_i r_j \int_0^\infty e^\theta f(\theta; \mu, \psi) d\theta + r_i r_j \int_0^\infty f(\theta; \mu, \psi) d\theta \\
&= r_i r_j M_\theta(2) - r_i r_j M_\theta(1) + r_i r_j \\
&= r_i r_j (M_\theta(2) - 2M_\theta(1) + 1), \quad i \neq j. \tag{3.20}
\end{aligned}$$

puesto que

$$\text{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i) \mathbb{E}(X_j)$$

resulta inmediato (3.19) combinando (3.17) y (3.20) ■

Corolario 3.2 *La distribución BNIG multivariada tiene correlación estrictamente positiva.*

Demostración: Puesto que $M_\theta(2) - M_\theta^2(1) = \text{Var}(e^\theta)$, resulta de manera inmediata que $\text{Cov}(X_i, X_j) > 0$, $i \neq j$ y por tanto, $\rho(X_i, X_j) > 0$. ■

3.1.5. Regresión

El problema que consideraremos en esta sección es utilizar los valores de una variable, digamos X_1 , con el objetivo de explicar o predecir los valores de otra, digamos X_2 . Para ello será necesario calcular la distribución condicional de X_2 dado que $X_1 = x_1$ y la esperanza condicional. Suponiendo que $X_i \sim \mathcal{BN}(r_i, e^{-\theta})$, $i = 1, 2$ y $\theta \sim \mathcal{IG}(\mu, \psi)$, la distribución condicional de X_2 dado X_1 viene dada por

$$\Pr(X_2 = x_2 | X_1 = x_1) = \frac{\binom{r_1 + x_1 - 1}{x_1} \binom{r_2 + x_2 - 1}{x_2}}{\binom{r_1 + r_2 + x_1 + x_2 - 1}{x_1 + x_2}} \frac{\Pr(X_1 + X_2 = x_1 + x_2 | r_1, r_2)}{\Pr(X_1 = x_1 | r_1)}, \tag{3.21}$$

La esperanza condicional, esto es la regresión de X_2 dado X_1 se muestra en el siguiente resultado.

Teorema 3.7 *La esperanza condicional de $X_2 | X_1$ viene dada por*

$$\mathbb{E}(X_2 | X_1 = x_1) = \frac{r_2(x_1 + 1)}{r_1 - 1} \frac{\Pr(x_1 + 1 | r_1 - 1)}{\Pr(x_1 | r_1)}.$$

Demostración: De (3.21) tenemos que

$$\begin{aligned}
 \mathbb{E}(X_2|X_1 = x_1) &= \sum_{x_2} x_2 \Pr(X_2 = x_2|X_1 = x_1) \\
 &= \sum_{x_2} \frac{\binom{r_1 + x_1 - 1}{x_1} \binom{r_2 + x_2 - 1}{x_2}}{\binom{r_1 + r_2 + x_1 + x_2 - 1}{x_1 + x_2}} \frac{\Pr(X_1 + X_2 = x_1 + x_2)}{\Pr(X_1 = x_1)} \\
 &= \frac{r_2(x_1 + 1)}{r_1 - 1} \\
 &\quad \times \sum_{x_2} \frac{\binom{(r_1 - 1) + (x_1 + 1) - 1}{x_1 + 1} \binom{(r_2 - 1) + (x_2 + 1) - 1}{x_2 - 1}}{\binom{(r_1 - 1) + (r_2 + 1) + (x_1 + 1) + (x_2 - 1) - 1}{(x_1 + 1) + (x_2 - 1)}} \\
 &\quad \times \frac{\Pr(X_1 + X_2 = x_1 + 1 + x_2 - 1)}{\Pr(X_1 = x_1)} \\
 &= \frac{r_2(x_1 + 1)}{r_1 - 1} \frac{\Pr(X_1 + 1|r_1 - 1)}{\Pr(X_1|r_1)}.
 \end{aligned}$$

■

Ahora, usando (3.5) y después de algunos cálculos tenemos que la esperanza condicional puede reescribirse como:

$$\mathbb{E}(X_2|X_1 = x_1) = r_2(x_1 + 1) \frac{\sum_{j=0}^{x_1+1} \frac{(-1)^j}{(x_1+1-j)!j!} \exp\left\{-\sqrt{\psi\left[\frac{\psi}{\mu^2} + 2(r_1 + j - 1)}\right]}\right\}}{\sum_{j=0}^{x_1} \frac{(-1)^j}{(x_1-j)!j!} \exp\left\{-\sqrt{\psi\left[\frac{\psi}{\mu^2} + 2(r_1 + j)}\right]}\right\}}. \quad (3.22)$$

Un simple vistazo muestra que la regresión (3.22) no es una función lineal en x_1 . Por otro lado, merece destacarse que la ecuación (3.22) es el estimador Bayes de $\mathbb{E}(X|\theta) = r(1-p)/p = r(e^\theta - 1)$ dado $X_1 = x_1$, y por tanto la prima Bayes bajo el principio de prima neta.

3.1.6. Estimación de parámetros

En esta sección se propondrán dos métodos de estimación de parámetros para los casos univariado y multivariado, el método de los momentos y el método de máxima verosimilitud.

Caso Univariado

Método de los momentos

El método de los momentos se basa en igualar los momentos teóricos y los muestrales. En este caso, necesitaremos los momentos de primer, segundo y tercer orden de la distribución \mathcal{BNIG} , los cuales vienen dados por (3.7), (3.8) y (3.9) e igualando dichas expresiones a los respectivos momentos muestrales se tiene:

$$\bar{x}_1 = r \left[e^{\frac{\psi}{\mu} \left(1 - \sqrt{1 - \frac{2\mu^2}{\psi}} \right)} - 1 \right], \quad (3.23)$$

$$\bar{x}_2 = (r + r^2) e^{\frac{\psi}{\mu} \left(1 - \sqrt{1 - \frac{4\mu^2}{\psi}} \right)} - (r + 2r^2) \left(\frac{\bar{x}_1}{r} + 1 \right) + r^2, \quad (3.24)$$

$$\begin{aligned} \bar{x}_3 &= (3r^3 + 3r^2 + r) \left(\frac{\bar{x}_1}{r} + 1 \right) \\ &- (3r^3 + 6r^2 + 3r) \left[\frac{\bar{x}_2 + (1 + 2r)(\bar{x}_1 + r) - r^2}{r + r^2} \right] \\ &+ (r^3 + 3r^2 + 2r) e^{\frac{\psi}{\mu} \left[1 - \sqrt{1 - \frac{6\mu^2}{\psi}} \right]} - r^3, \end{aligned} \quad (3.25)$$

donde hemos denotado mediante \bar{x}_i , ($i = 1, 2, 3$) los momentos muestrales de primer, segundo y tercer orden respectivamente.

Despejando el parámetro ψ de las ecuaciones (3.23) y (3.24) obtenemos

$$\psi = \frac{\mu \log^2 \left(\frac{\bar{x}_1}{r} + 1 \right)}{2 \log \left(\frac{\bar{x}_1}{r} + 1 \right) - 2\mu} \quad (3.26)$$

y

$$\psi = \frac{\mu \log^2 \left[\frac{\bar{x}_2 + (1 + 2r)(\bar{x}_1 + r) - r^2}{r + r^2} \right]}{2 \log \left[\frac{\bar{x}_2 + (1 + 2r)(\bar{x}_1 + r) - r^2}{r + r^2} \right] - 4\mu} \quad (3.27)$$

respectivamente.

Igualando (3.26) y (3.27) y después de algunos cálculos, tenemos:

$$\mu = \frac{A + B}{C + D}, \quad (3.28)$$

donde

$$A = \log \left[\frac{\bar{x}_2 + (1 + 2r)(\bar{x}_1 + r) - r^2}{r + r^2} \right] \log^2 \left(\frac{\bar{x}_1}{r} + 1 \right),$$

$$\begin{aligned} \mathcal{B} &= \log^2 \left[\frac{\bar{x}_2 + (1 + 2r)(\bar{x}_1 + r) - r^2}{r + r^2} \right] \log\left(\frac{\bar{x}_1}{r} + 1\right) 2 \log^2\left(\frac{\bar{x}_1}{r} + 1\right), \\ \mathcal{C} &= 2 \log^2\left(\frac{\bar{x}_1}{r} + 1\right), \\ \mathcal{D} &= \log^2 \left[\frac{\bar{x}_2 + (1 + 2r)(\bar{x}_1 + r) - r^2}{r + r^2} \right]. \end{aligned}$$

Sustituyendo (3.26) ó (3.27) y (3.28) en la ecuación (3.25) obtenemos una ecuación que depende sólo del parámetro r y que por tanto puede resolverse numéricamente, proporcionando el estimador \hat{r} . Llevando este valor a (3.28) se obtiene el estimador $\hat{\mu}$. Finalmente llevando \hat{r} y $\hat{\mu}$ a (3.26) o a (3.27) conseguimos el estimador $\hat{\psi}$.

Método de máxima verosimilitud

La estimación de los parámetros del modelo \mathcal{BNIG} puede también realizarse por el método de máxima verosimilitud. Denotando los parámetros de la distribución \mathcal{BNIG} por (r, μ, ψ) , las frecuencias observadas por f_0, \dots, f_n , con $f = \sum_{i=1}^n f_i$, y las correspondientes probabilidades de la ecuación (3.5) mediante p_0, \dots, p_n , tenemos que la función logaritmo verosimilitud viene dada por:

$$\begin{aligned} \log L(r, \mu, \psi) &= \sum_{x=0}^n f_x \log p_x \\ &= \sum_{x=0}^n f_x \log \left[\binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_{\theta}(-(r+j)) \right]. \end{aligned}$$

Los estimadores máximo verosímiles de r , ψ y μ se obtienen resolviendo el siguiente sistema de ecuaciones:

$$\begin{aligned} \frac{\partial \log L(r, \mu, \psi)}{\partial r} &= \sum_{x=0}^n \frac{f_x}{p_x} \frac{\partial}{\partial r} \left[\binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_{\theta}(-(r+j)) \right] \\ &= \sum_{x=0}^n \frac{f_x}{p_x} \sum_{j=0}^x \binom{x}{j} (-1)^j \binom{r+x-1}{x} \frac{\partial}{\partial r} M_{\theta}(-(r+j)) = 0, \\ & \hspace{15em} (3.29) \\ \frac{\partial \log L(r, \mu, \psi)}{\partial \mu} &= \sum_{x=0}^n \frac{f_x}{p_x} \frac{\partial}{\partial \mu} \left[\binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_{\theta}(-(r+j)) \right] \end{aligned}$$

$$= \sum_{x=0}^n \frac{f_x}{p_x} \sum_{j=0}^x \binom{x}{j} (-1)^j \binom{r+x-1}{x} \frac{\partial}{\partial \psi} M_{\theta}(-(r+j)) = 0, \quad (3.30)$$

$$\begin{aligned} \frac{\partial \log L(r, \mu, \psi)}{\partial \psi} &= \sum_{x=0}^n \frac{f_x}{p_x} \frac{\partial}{\partial \psi} \left[\binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_{\theta}(-(r+j)) \right] \\ &= \sum_{x=0}^n \frac{f_x}{p_x} \sum_{j=0}^x \binom{x}{j} (-1)^j \binom{r+x-1}{x} \frac{\partial}{\partial \psi} M_{\theta}(-(r+j)) = 0, \end{aligned} \quad (3.31)$$

para ello se precisarán de las siguientes derivadas:

$$\begin{aligned} \frac{\partial M_{\theta}(t)}{\partial \mu} &= \left[-\frac{\psi}{\mu^2} \left\{ 1 - \sqrt{1 + \frac{2\mu^2 t}{\psi}} \right\} - 2t / \sqrt{1 + \frac{2\mu^2 t}{\psi}} \right] \\ &\times e^{\frac{\psi}{\mu} \left(1 - \sqrt{1 + \frac{2\mu^2 t}{\psi}} \right)}, \end{aligned} \quad (3.32)$$

$$\begin{aligned} \frac{\partial M_{\theta}(t)}{\partial \psi} &= \left[\frac{1}{\mu} \left(1 - \sqrt{1 + \frac{2\mu^2 t}{\psi}} \right) + \mu t / \left(\psi \sqrt{1 + \frac{2\mu^2 t}{\psi}} \right) \right] \\ &\times e^{\frac{\psi}{\mu} \left(1 - \sqrt{1 + \frac{2\mu^2 t}{\psi}} \right)}, \end{aligned} \quad (3.33)$$

$$\frac{\partial M_{\theta}(t)}{\partial r} = -\frac{\mu}{\sqrt{1 + \frac{2\mu^2 t}{\psi}}} e^{\frac{\psi}{\mu} \left(1 - \sqrt{1 + \frac{2\mu^2 t}{\psi}} \right)}, \quad (3.34)$$

$$\frac{\partial}{\partial r} \left\{ \frac{\Gamma(r+k)}{\Gamma(r)\Gamma(k+1)} \right\} = \frac{\Gamma(r)\Gamma'(r+k) - \Gamma(r+k)\Gamma'(r)}{\Gamma(k+1) [\Gamma(r)]^2}. \quad (3.35)$$

Resulta obvio que, a pesar de los avances informáticos, el sistema de ecuaciones anterior no resulta sencillo de resolver. En el capítulo de ilustraciones numéricas y como aplicación a un ejemplo, se propondrá su resolución a través del método *scoring* (veáse Klugman et al, 1998).

Método de la proporción de ceros

Aunque el método de los momentos proporciona una vía sencilla, en nuestro caso, para estimar los parámetros del modelo, existe otra alternativa que resulta adecuada cuando la proporción de ceros observadas es alta, el método de la proporción de ceros. Este método (Alanko y Duffy, 1996) es similar al de los momentos utilizando, en nuestra distribución, tres ecuaciones dadas por la proporción de ceros, p_0 , $\mathbb{E}(X)$ y $\mathbb{E}(X^2)$.

De esta manera, utilizamos las ecuaciones (3.23), (3.24) y

$$p_0 = \Pr(X = 0) = e^{-\frac{\psi}{\mu}} \left[1 - \sqrt{1 + \frac{2\mu^2 r}{\psi}} \right]. \quad (3.36)$$

De nuevo, despejando el parámetro ψ en las ecuaciones (3.36) y (3.23) tenemos

$$\psi = \frac{\mu \log^2 p_0}{2\mu r + 2 \log p_0} \quad (3.37)$$

y

$$\psi = \frac{\mu \log^2 \left(\frac{\bar{x}_1}{r} + 1 \right)}{2 \log \left(\frac{\bar{x}_1}{r} + 1 \right) - 2\mu}, \quad (3.38)$$

respectivamente.

Igualando (3.37) y (3.38) y después de algunos cálculos, obtenemos el parámetro μ , dado por:

$$\mu = \frac{\log^2 p_0 \log \left(\frac{\bar{x}_1}{r} + 1 \right) - \log p_0 \log^2 \left(\frac{\bar{x}_1}{r} + 1 \right)}{r \log^2 \left(\frac{\bar{x}_1}{r} + 1 \right) + \log^2 p_0}. \quad (3.39)$$

Nuevamente, sustituyendo (3.37) o (3.38) y (3.39) en la ecuación (3.24) obtenemos una expresión que sólo depende del parámetro r y que por tanto se puede resolver numéricamente para obtener el estimador \hat{r} . Llevando este valor a (3.39) se obtiene $\hat{\mu}$, y finalmente incorporando \hat{r} y $\hat{\mu}$ en (3.37) o en (3.38) se obtiene $\hat{\psi}$.

Caso Bivariado

Ahora mostraremos el método de estimación por el método de los momentos y de máxima verosimilitud para el caso bivariado, exteniéndose de manera natural para el caso trivariado, etc. Usando las expresiones (3.17), (3.18) y (3.19) en la forma:

$$\begin{aligned} \mathbb{E}(X_1) &= \bar{x}_1, \\ \mathbb{E}(X_2) &= \bar{x}_2, \\ \text{Var}(X_1) + \text{Var}(X_2) &= s_1^2 + s_2^2, \\ \text{Cov}(X_1, X_2) &= s_{12}, \end{aligned}$$

donde \bar{x}_i , s_i , ($i = 1, 2$) y s_{12} son las medias muestrales, varianzas y covarianzas, respectivamente y después de algunos cálculos se obtienen las siguientes expresiones:

$$s_1^2 + s_2^2 = r_1 (A - B - r_1 B^2), \quad A = \left(\frac{\bar{x}_1}{r_1} + 1 \right)^2 + \frac{s_{12} \bar{x}_1}{r_1^2 \bar{x}_2}, \quad B = \frac{\bar{x}_1}{r_1} + 1, \quad (3.40)$$

$$\hat{\mu} = \frac{\log B \log^2 A - \log A \log^2 B}{\log^2 A - 2 \log^2 B}, \quad (3.41)$$

$$\hat{\psi} = \frac{\hat{\mu} \log^2 A}{2 \log A - 4 \hat{\mu}} = \frac{\hat{\mu} \log^2 B}{2 \log B - 2 \hat{\mu}}, \quad (3.42)$$

$$\hat{r}_2 = \frac{\bar{x}_2}{\bar{x}_1} \hat{r}_1. \quad (3.43)$$

La expresión (3.40) puede resolverse numéricamente para obtener \hat{r}_1 . El resto de los parámetros se obtiene a partir de (3.41), (3.42) y (3.43) en un procedimiento secuencial.

La estimación por el método de máxima verosimilitud es similar al caso univariado aunque, obviamente, se van a necesitar derivadas más complejas. Denotando los parámetros de la distribución por (r_1, r_2, μ, ψ) , las frecuencias observadas por f_0, \dots, f_{N_1, N_2} , $f = \sum_{i=0}^{N_1 N_2} f_i$, y por p_0, \dots, p_{N_1, N_2} las probabilidades correspondientes a la ecuación (3.15), tenemos que el logaritmo de la función de verosimilitud para el caso bivariado viene dado por

$$\begin{aligned} \log L(r_1, r_2, \psi, \mu) &= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} f_{x_1 x_2} \log p_{x_1 x_2} \\ &= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} f_{x_1 x_2} \log \prod_{i=1}^2 \binom{r_i + x_i - 1}{x_i} \\ &\quad \times \sum_{j=0}^{x_1+x_2} (-1)^j \binom{x_1+x_2}{j} M_\theta(-(r_1+r_2+j)). \end{aligned}$$

Tomando derivadas parciales respecto a r_1, r_2, μ y ψ e igualando a 0 obtenemos el sistema de ecuaciones:

$$\begin{aligned} \frac{\partial \log L(\psi, \mu)}{\partial r_1} &= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} \frac{f_{x_1 x_2}}{p_{x_1 x_2}} \frac{\partial}{\partial r_1} \left[\prod_{i=1}^2 \binom{r_i + x_i - 1}{x_i} \right. \\ &\quad \left. \times \sum_{j=0}^{x_1+x_2} (-1)^j \binom{x_1+x_2}{j} M_\theta(-(r_1+r_2+j)) \right] \end{aligned}$$

$$\begin{aligned}
&= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} \frac{f_{x_1 x_2}}{p_{x_1 x_2}} \prod_{i=1}^2 \binom{r_i + x_i - 1}{x_i} \\
&\times \sum_{j=0}^{x_1+x_2} (-1)^j \binom{x_1+x_2}{j} \frac{\partial}{\partial r_2} M_\theta(-(r_1+r_2+j)) = 0, \quad (3.44)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \log L(\psi, \mu)}{\partial r_2} &= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} \frac{f_{x_1 x_2}}{p_{x_1 x_2}} \frac{\partial}{\partial r_1} \left[\prod_{i=1}^2 \binom{r_i + x_i - 1}{x_i} \right. \\
&\times \left. \sum_{j=0}^{x_1+x_2} (-1)^j \binom{x_1+x_2}{j} M_\theta(-(r_1+r_2+j)) \right] \\
&= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} \frac{f_{x_1 x_2}}{p_{x_1 x_2}} \prod_{i=1}^2 \binom{r_i + x_i - 1}{x_i} \\
&\times \sum_{j=0}^{x_1+x_2} (-1)^j \binom{x_1+x_2}{j} \frac{\partial}{\partial r_2} M_\theta(-(r_1+r_2+j)) = 0, \quad (3.45)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \log L(\psi, \mu)}{\partial \mu} &= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} \frac{f_{x_1 x_2}}{p_{x_1 x_2}} \frac{\partial}{\partial \mu} \left[\prod_{i=1}^2 \binom{r_i + x_i - 1}{x_i} \right. \\
&\times \left. \sum_{j=0}^{x_1+x_2} (-1)^j \binom{x_1+x_2}{j} M_\theta(-(r_1+r_2+j)) \right] \\
&= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} \frac{f_{x_1 x_2}}{p_{x_1 x_2}} \prod_{i=1}^2 \binom{r_i + x_i - 1}{x_i} \\
&\times \sum_{j=0}^{x_1+x_2} (-1)^j \binom{x_1+x_2}{j} \frac{\partial}{\partial \mu} M_\theta(-(r_1+r_2+j)) = 0, \quad (3.46)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \log L(\psi, \mu)}{\partial \psi} &= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} \frac{f_{x_1 x_2}}{p_{x_1 x_2}} \frac{\partial}{\partial \psi} \left[\prod_{i=1}^2 \binom{r_i + x_i - 1}{x_i} \right. \\
&\times \left. \sum_{j=0}^{x_1+x_2} (-1)^j \binom{x_1+x_2}{j} M_\theta(-(r_1+r_2+j)) \right] \\
&= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} \frac{f_{x_1 x_2}}{p_{x_1 x_2}} \prod_{i=1}^2 \binom{r_i + x_i - 1}{x_i} \\
&\times \sum_{j=0}^{x_1+x_2} (-1)^j \binom{x_1+x_2}{j} \frac{\partial}{\partial \psi} M_\theta(-(r_1+r_2+j)) = 0, \quad (3.47)
\end{aligned}$$

donde las derivadas que se precisan vienen dadas de nuevo por (3.32), (3.33), (3.34), mientras que

$$\frac{\partial}{\partial r_i} \left\{ \frac{\Gamma(r_i + k)}{\Gamma(r)\Gamma(k+1)} \right\} = \frac{\Gamma(r_i)\Gamma'(r_i + k) - \Gamma(r_i + k)\Gamma'(r)}{\Gamma(k+1)[\Gamma(r_i)]^2}, \quad i = 1, 2.$$

Por supuesto que al igual que ocurría con el caso univariado la resolución de este sistema no es sencilla pero, como veremos en el capítulo de ilustraciones numéricas, el método *scoring* se mostrará como un procedimiento adecuado para su resolución.

3.2. Modelo binomial–inversa gaussiana

Un desarrollo paralelo al expuesto en las secciones anteriores para la distribución binomial negativa se llevará a cabo en esta sección para la distribución binomial. El caso univariado para el que si X sigue una distribución binomial con parámetros n y $p = 1 - e^{-\theta}$ y θ sigue una distribución inversa gaussiana aparece desarrollado en Alanko y Duffy (1996). De hecho este trabajo constituyó la fuente inspiración del modelo *BNTG*. Nos limitaremos a exponer expresiones recursivas para el caso univariado y a desarrollar un modelo multivariado similar al de la sección anterior. El nuevo modelo binomial inversa gaussiana lo denotaremos de aquí en adelante *BIG*.

3.2.1. Versión multivariada del modelo *BIG*

Definición 3.3 Una distribución multivariada binomial–inversa gaussiana (X_1, \dots, X_d) viene definida por la siguiente representación estocástica:

$$\begin{aligned} X_i | \theta &\sim \mathcal{B}(N_i, 1 - e^{-\theta}), \quad i = 1, \dots, d \text{ independientes,} \\ \theta &\sim \mathcal{IG}(\mu, \psi). \end{aligned}$$

Teorema 3.8 La función de probabilidad conjunta es $(x_1, x_2, \dots, x_d = 0, 1, 2, \dots)$:

$$\begin{aligned} &\Pr(X_1 = x_1, X_2 = x_2, \dots, X_d = x_d) \\ &= \prod_{i=1}^d \binom{N_i}{x_i} \sum_{j=0}^{\tilde{x}} (-1)^{\tilde{x}-j} \binom{\tilde{x}}{j} \exp \left\{ \frac{\psi}{\mu} \left[1 - \sqrt{1 + \frac{2(\tilde{N} - j)\mu^2}{\psi}} \right] \right\}, \end{aligned} \quad (3.48)$$

donde $\theta, N_1, \dots, N_p > 0$ y

$$\tilde{N} = N_1 + \dots + N_d, \quad (3.49)$$

$$\tilde{x} = x_1 + \dots + x_d.$$

Demostración: Si $X_i|\theta \sim \mathcal{B}(N_i, 1-e^{-\theta})$, $i = 1, 2, \dots, d$ y $\theta \sim \mathcal{IG}(\mu, \psi)$, la función de probabilidad conjunta de (X_1, \dots, X_d) puede obtenerse usando la conocida fórmula,

$$\begin{aligned} & \Pr(X_1 = x_1, X_2 = x_2, \dots, X_d = x_d) \\ &= \int_0^\infty \Pr(X_1 = x_1, X_2 = x_2, \dots, X_d = x_d|\theta) f(\theta; \mu, \psi) d\theta, \end{aligned}$$

donde $f(\theta; \mu, \psi)$ es la función de probabilidad de una distribución inversa gaussiana definida en (3.1). Como puede observarse esta distribución puede expresarse también en una forma cerrada.

$$\begin{aligned} & \Pr(X_1 = x_1, X_2 = x_2, \dots, X_d = x_d) \\ &= \int_0^\infty \prod_{i=1}^d \binom{N_i}{x_i} p^{x_i} q^{N_i-x_i} f(\theta; \mu, \psi) d\theta \\ &= \int_0^\infty \prod_{i=1}^d \binom{N_i}{x_i} (1-e^{-\theta})^{x_i} (e^{-\theta})^{N_i-x_i} f(\theta; \mu, \psi) d\theta. \end{aligned}$$

Ahora usando el desarrollo del binomio de Newton podemos reescribir la última expresión de la forma siguiente

$$\begin{aligned} & \Pr(X_1 = x_1, X_2 = x_2, \dots, X_d = x_d) \\ &= \int_0^\infty \prod_{i=1}^d \binom{N_i}{x_i} \sum_{j=0}^{\tilde{x}} \binom{\tilde{x}}{j} (-1)^j (e^{-\theta})^j (e^{-\theta})^{\tilde{N}-\tilde{x}} f(\theta; \mu, \psi) d\theta \\ &= \prod_{i=1}^d \binom{N_i}{x_i} \sum_{j=0}^{\tilde{x}} \binom{\tilde{x}}{j} (-1)^j \int_0^\infty (e^{-\theta})^{\tilde{N}-\tilde{x}+j} f(\theta; \mu, \psi) d\theta \\ &= \prod_{i=1}^d \binom{N_i}{x_i} \int_0^\infty \sum_{j=0}^{\tilde{x}} \binom{\tilde{x}}{j} (-1)^{\tilde{x}-j} (e^{-\theta})^{\tilde{N}-j} f(\theta; \mu, \psi) d\theta \end{aligned}$$

$$= \prod_{i=1}^d \binom{N_i}{x_i} \sum_{j=0}^{\tilde{x}} \binom{\tilde{x}}{j} (-1)^{\tilde{x}-j} M_\theta(-(\tilde{N} - j)),$$

y sustituyendo en la expresión anterior (3.2), obtenemos el resultado. \blacksquare

Esta última expresión, al igual que en el caso anterior \mathcal{BNIG} puede obtenerse de una forma alternativa para calcular las probabilidades. Si $\tilde{Z} \sim \mathcal{B}(\tilde{N})$, donde \tilde{N} viene dada como en (3.49), entonces una expresión alternativa para (3.48) es

$$\Pr(X_1 = x_1, X_2 = x_2, \dots, X_d = x_d) = \frac{\prod_{i=1}^d \binom{N_i}{x_i}}{\binom{\tilde{N}}{\tilde{x}}} \cdot \Pr(\tilde{Z} = \tilde{x}). \quad (3.50)$$

con $d \geq 2$.

Las distribuciones marginales son obviamente $X_i \sim \mathcal{B}(N_i, 1 - e^{-\theta})$, $i = 1, 2, \dots, d$ y cualquier subvector (X_1, \dots, X_s) con $s < d$ es nuevamente una distribución multivariada de dimensión s .

Algunas propiedades aparecen en el siguiente resultado.

Teorema 3.9 Sea $X \sim \mathcal{BIG}(\mu, \psi)$ la versión univariada de la distribución binomial-inversa gaussiana puede obtenerse como un caso particular de (3.48) cuando $d = 1$.

Esta distribución coincide con la obtenida por Alanko y Duffy (1996):

(a) La función de probabilidad viene dada por,

$$\Pr(X = x) = \binom{N}{x} \sum_{j=0}^x (-1)^{x-j} \binom{x}{j} \exp \left\{ \frac{\psi}{\mu} \left[1 - \sqrt{1 + \frac{2(N-j)\mu^2}{\psi}} \right] \right\}, \quad (3.51)$$

con $x = 0, 1, 2, \dots, N$ y μ y $\psi > 0$.

(b) Los momentos de la distribución \mathcal{BIG} vienen dados por:

$$\mathbb{E}(X_i) = N_i [1 - M_\theta(-1)], \quad i = 1, 2, \dots, \quad (3.52)$$

$$\text{Var}(X_i) = (N_i^2 - N_i)M_\theta(-2) + N_i M_\theta(-1) - N_i^2 M_\theta^2(-1), \quad i = 1, 2, \dots, \quad (3.53)$$

$$\mathbb{E}(X_i X_j) = N_i N_j [1 - 2M_\theta(-1) + M_\theta(-2)], \quad i, j = 1, 2, \dots, i \neq j \quad (3.54)$$

$$\text{Cov}(X_i, X_j) = N_i N_j [M_\theta(-2) - M_\theta^2(-1)], \quad i \neq j. \quad (3.55)$$

Demostración:

- (a) Es trivial tomando $d = 1$ en el Teorema 3.8.
- (b) Las expresiones (3.52) y (3.53) pueden obtenerse fácilmente usando esperanzas condicionadas.

Para obtener (3.54) hemos de tener en cuenta:

$$\begin{aligned}
 \mathbb{E}(X_i X_j) &= \mathbb{E}_\theta[\mathbb{E}(X_i X_j | \theta)] \\
 &= \mathbb{E}_\theta(N_i N_j p^2) = N_i N_j \mathbb{E}_\theta(p^2) \\
 &= N_i N_j \int_0^\infty (1 - e^{-\theta})^2 f(\theta; \mu, \psi) d\theta \\
 &= N_i N_j \int_0^\infty (1 - e^{-\theta} + e^{-2\theta}) f(\theta; \mu, \psi) d\theta \\
 &= N_i N_j \int_0^\infty f(\theta; \mu, \psi) d\theta - 2N_i N_j \int_0^\infty e^{-\theta} f(\theta; \mu, \psi) d\theta \\
 &+ N_i N_j \int_0^\infty e^{-2\theta} f(\theta; \mu, \psi) d\theta \\
 &= N_i N_j - 2N_i N_j M_\theta(-1) + N_i N_j M_\theta(-2) \\
 &= N_i N_j (1 - 2M_\theta(-1) + M_\theta(-2)), \quad i \neq j, \tag{3.56}
 \end{aligned}$$

como la covarianza se define por

$$\text{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i) \mathbb{E}(X_j),$$

el resultado (3.55) se obtiene combinando (3.54) y (3.52). ■

Corolario 3.3 *La extensión multivariada de la distribución BIG tiene correlación estrictamente positiva.*

Demostración: Puesto que $M_\theta(2) - M_\theta^2(1) = \text{Var}(e^{-\theta})$, resulta de manera inmediata que $\text{Cov}(X_i, X_j) > 0$, $i \neq j$ y por tanto, $\rho(X_i, X_j) > 0$. ■

3.2.2. Relación recursiva para el cálculo de probabilidades

Si en (3.48) escribimos $d = 1$ se obtiene la distribución BIG univariada que aparece en Alanko y Duffy (1996). En este caso:

$$\Pr(X = x) = \binom{N}{x} \sum_{j=0}^x (-1)^{x-j} \binom{x}{j} \exp \left\{ \frac{\psi}{\mu} \left[1 - \sqrt{1 + \frac{2(N-j)\mu^2}{\psi}} \right] \right\},$$

a la que haremos referencia en el siguiente Teorema.

Teorema 3.10 *Las probabilidades de una variable aleatoria que siga una distribución BIG pueden obtenerse de forma recursiva como:*

$$f_N(x) = \frac{N}{x} f_{N-1}(x-1) - \frac{N-x+1}{x} f_N(x-1), \quad x = 1, 2, \dots, N. \quad (3.57)$$

Demostración: Para la distribución binomial con función de probabilidad

$$f(x|\theta) = \binom{N}{x} (1 - e^{-\theta})^x (e^{-\theta})^{N-x}, \quad x = 0, 1, \dots, N$$

tenemos la siguiente fórmula recursiva

$$\frac{f(x|\theta)}{f(x-1|\theta)} = \frac{N-x+1}{x} \frac{(1 - e^{-\theta})}{(e^{-\theta})}, \quad x = 1, 2, \dots, N. \quad (3.58)$$

Usando la definición de la distribución BIG y (3.58) obtenemos

$$\begin{aligned} f_N(x) &= \int_0^\infty f(x|\theta) f(\theta) d\theta \\ &= \int_0^\infty \frac{N-x+1}{x} \frac{(1 - e^{-\theta})}{(e^{-\theta})} f(x-1|\theta) f(\theta) d\theta \\ &= \frac{N-x+1}{x} \int_0^\infty (e^\theta - 1) f(x-1|\theta) f(\theta) d\theta \\ &= \frac{N-x+1}{x} \int_0^\infty e^\theta f(x-1|\theta) f(\theta) d\theta - \frac{N-x+1}{x} f_N(x-1). \end{aligned}$$

Ahora se tiene

$$\begin{aligned} &\frac{N-x+1}{x} \int_0^\infty e^\theta f(x-1|\theta) f(\theta) d\theta \\ &= \frac{N-x+1}{x} \int_0^\infty e^\theta \binom{N}{x-1} (1 - e^{-\theta})^{x-1} (e^{-\theta})^{N-x+1} f(\theta) d\theta \\ &= \frac{N-x+1}{x} \int_0^\infty \frac{1}{e^{-\theta}} \binom{N}{x-1} (1 - e^{-\theta})^{x-1} (e^{-\theta})^{N-x+1} f(\theta) d\theta \\ &= \frac{N-x+1}{x} \int_0^\infty \binom{N}{x-1} (1 - e^{-\theta})^{x-1} (e^{-\theta})^{N-x} f(\theta) d\theta \\ &= \frac{N-x+1}{x} \int_0^\infty \binom{N}{x-1} (1 - e^{-\theta})^{x-1} (e^{-\theta})^{N-1-(x-1)} f(\theta) d\theta \end{aligned}$$

$$\begin{aligned}
&= \binom{N}{x} \int_0^\infty (1 - e^{-\theta})^{x-1} (e^{-\theta})^{N-1-(x-1)} f(\theta) d\theta \\
&= \frac{\binom{N}{x}}{\binom{N-1}{x-1}} \int_0^\infty \binom{N-1}{x-1} (1 - e^{-\theta})^{x-1} (e^{-\theta})^{N-1-(x-1)} f(\theta) d\theta \\
&= \frac{N}{x} f_{N-1}(x-1),
\end{aligned}$$

y por lo tanto obtenemos (3.57). ■

El resultado principal se propone en el siguiente Teorema.

Teorema 3.11 *Si la variable aleatoria cantidad reclamada es continua con función de densidad de probabilidad $f(x)$, $x > 0$, entonces la función de densidad de la cantidad total reclamada cuando la distribución del número de reclamaciones es la BIG verifica la siguiente ecuación integral:*

$$\begin{aligned}
g_s(x; N) &= p_N(0) + \int_0^x \frac{Ny}{x} g_s(x-y; N-1) f(y) dy \\
&\quad - \int_0^x \frac{Ny-x+y}{x} g_s(x-y; N) f(y) dy. \tag{3.59}
\end{aligned}$$

donde $p_N(k)$ es la distribución del número de reclamaciones.

Demostración: Tenemos que la cantidad total reclamada viene dada por

$$g_s(x; N) = \sum_{k=0}^{\infty} p_N(k) f^{k*}(x) = p_N(0) f^{0*}(x) + \sum_{k=1}^{\infty} p_N(k) f^{k*}(x),$$

donde $f^{k*}(x)$ denota la k -ésima convolución de $f(x)$.

Ahora, usando (3.57) tenemos:

$$p_N(k) = \frac{N}{k} p_{N-1}(k-1) - \frac{N-k+1}{k} p_N(k-1), \quad k = 1, 2, \dots$$

Entonces, se tiene que

$$\begin{aligned}
\sum_{k=1}^{\infty} p_N(k) f^{k*}(x) &= \sum_{k=1}^{\infty} \frac{N}{k} p_{N-1}(k-1) f^{k*}(x) - \sum_{k=1}^{\infty} \frac{N}{k} p_N(k-1) f^{k*}(x) \\
&\quad + \sum_{k=1}^{\infty} p_N(k-1) f^{k*}(x) - \sum_{k=1}^{\infty} \frac{1}{k} p_N(k-1) f^{k*}(x). \tag{3.60}
\end{aligned}$$

Ahora, después de algunos cálculos y usando las identidades siguientes

$$f^{k*}(x) = \int_0^x f^{(k-1)*}(x-y)f(y)dy, \quad k = 1, 2, \dots$$

$$\frac{f^{k*}(x)}{k} = \int_0^x \frac{y}{x} f^{(k-1)*}(x-y)f(y)dy, \quad k = 1, 2, \dots$$

obtenemos el resultado. ■

La ecuación integral (3.59) debe resolverse numéricamente. Existen varios algoritmos e implementaciones para resolver la ecuación integral de Volterra de segunda especie que aparece, sin embargo se necesitarán de modificaciones para usarlas en (3.59).

Es sencillo comprobar que (3.59) y (3.60) se satisfacen intercambiando el símbolo integral por el de suma, lo que permite obtener una expresión exacta y recursiva para la distribución de la cantidad total reclamada para el caso en que la variable aleatoria cantidad reclamada sea discreta. Esto se pone de manifiesto en el siguiente resultado.

Corolario 3.4 Si la distribución de la cuantía de las reclamaciones es discreta y definida sobre los enteros positivos, la distribución de la cuantía agregada es:

$$g_s(x; N) = p_N(0) + \sum_{y=0}^x \frac{Ny}{x} g_s(x-y; N-1)f(y)$$

$$- \sum_{y=0}^x \frac{Ny-x+y}{x} g_s(x-y; N)f(y).$$

Demostración: Basta sustituir en (3.59) el símbolo integral por el de suma. ■

3.2.3. Estimación de parámetros

En esta sección estudiaremos métodos para estimar los parámetros de la distribución $\mathcal{B}IG$ para el caso biviado.

Método de los momentos

Seguiremos un procedimiento parecido al usado para estimar los parámetros de la distribución bivariada $\mathcal{B}NIG$. En este caso sólo hay dos parámetros que estimar, μ y ψ . Usando las expresiones (3.52) y (3.55) obtenemos el sistema de ecuaciones:

$$\mathbb{E}(X_1) + \mathbb{E}(X_2) = \bar{x}_1 + \bar{x}_2,$$

$$\text{Cov}(X_1, X_2) = s_{12},$$

donde \bar{x}_i , ($i = 1, 2$) y s_{12} son las medias muestrales y la covarianza muestral respectivamente, y después de cálculos considerables obtenemos la fórmulas siguientes

$$\bar{x}_1 + \bar{x}_2 = (N_1 + N_2)[1 - A], \quad s_{12} = N_1 N_2 [B - A^2], \quad (3.61)$$

$$\hat{\mu} = \frac{2\hat{\psi} \log A}{\log^2 A - 2\hat{\psi}} = \frac{2\hat{\psi} \log B}{\log^2 B - 4\hat{\psi}}, \quad (3.62)$$

$$\hat{\psi} = \frac{\log A \log^2 B - \log B \log^2 A}{4 \log A - 2 \log B}. \quad (3.63)$$

siendo $A = \exp \left\{ \frac{\psi}{\mu} \left[1 - \sqrt{1 + \frac{2\mu^2}{\psi}} \right] \right\}$ y $B = \exp \left\{ \frac{\psi}{\mu} \left[1 - \sqrt{1 + \frac{4\mu^2}{\psi}} \right] \right\}$.

De (3.61) se obtienen los valores de A y B ; (3.63) puede resolverse numéricamente para obtener el estimador $\hat{\psi}$. El estimador del parámetro $\hat{\mu}$ se calcula usando (3.62) en un procedimiento secuencial.

Método de máxima verosimilitud

La estimación de parámetros de la distribución BIG en (3.48) puede realizarse también por el método de máxima verosimilitud. Denotando los parámetros de la distribución IG por (ψ, μ) , las frecuencias observadas por $f_0, \dots, f_{N_1 N_2}$, $f = \sum_{i=0}^{N_1 N_2} f_i$, y por $p_0, \dots, p_{N_1 N_2}$ las probabilidades correspondientes a la ecuación (3.48), tenemos que el logaritmo de la función de verosimilitud para el caso bivariado viene dado por

$$\begin{aligned} \log L(\psi, \mu) &= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} f_{x_1 x_2} \log p_{x_1 x_2} = \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} f_{x_1 x_2} \log \prod_{i=1}^2 \binom{N_i}{x_i} \\ &\times \sum_{j=0}^{x_1+x_2} (-1)^{x_1+x_2-j} \binom{x_1+x_2}{j} M_\theta(-(N_1 + N_2 - j)). \end{aligned}$$

Tomando derivadas parciales respecto a ψ y μ e igualando a 0 obtenemos el sistema de ecuaciones:

$$\begin{aligned} \frac{\partial \log L(\psi, \mu)}{\partial \mu} &= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} \frac{f_{x_1 x_2}}{p_{x_1 x_2}} \frac{\partial}{\partial \mu} \left[\prod_{i=1}^2 \binom{N_i}{x_i} \right. \\ &\times \left. \sum_{j=0}^{x_1+x_2} (-1)^{x_1+x_2-j} \binom{x_1+x_2}{j} M_\theta(-(N_1 + N_2 - j)) \right] \\ &= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} \frac{f_{x_1 x_2}}{p_{x_1 x_2}} \prod_{i=1}^2 \binom{N_i}{x_i} \end{aligned}$$

$$\begin{aligned}
& \times \sum_{j=0}^{x_1+x_2} (-1)^{x_1+x_2-j} \binom{x_1+x_2}{j} \frac{\partial}{\partial \mu} M_\theta(-(N_1+N_2-j)) = 0, \\
& \qquad \qquad \qquad (3.64) \\
\frac{\partial \log L(\psi, \mu)}{\partial \psi} &= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} \frac{f_{x_1 x_2}}{p_{x_1 x_2}} \frac{\partial}{\partial \psi} \left[\prod_{i=1}^2 \binom{N_i}{x_i} \right. \\
& \times \left. \sum_{j=0}^{x_1+x_2} (-1)^{x_1+x_2-j} \binom{x_1+x_2}{j} M_\theta(-(N_1+N_2-j)) \right] \\
&= \sum_{x_1=0}^{N_1} \sum_{x_2=0}^{N_2} \frac{f_{x_1 x_2}}{p_{x_1 x_2}} \prod_{i=1}^2 \binom{N_i}{x_i} \\
& \times \sum_{j=0}^{x_1+x_2} (-1)^{x_1+x_2-j} \binom{x_1+x_2}{j} \frac{\partial}{\partial \psi} M_\theta(-(N_1+N_2-j)) = 0, \\
& \qquad \qquad \qquad (3.65)
\end{aligned}$$

donde de nuevo se precisa de las derivadas (3.32) y (3.33).

3.3. Conclusiones

En este capítulo se ha introducido una nueva distribución de probabilidad, la *BNIG* en su versión univariada y multivariada, además de la extensión multivariada de la *BIG*. Ambas distribuciones parecen adecuadas para utilizarse en escenarios actuariales en las que se de el fenómeno de sobredispersión, es decir, que la varianza del número observado de la distribución de frecuencias sea mayor que la media. Ambas distribuciones, además de presentar expresiones cerradas para la función de probabilidad tienen la ventaja añadida de admitir relaciones recursivas útiles en el modelo colectivo compuesto, en los que se trabajan con las distribuciones número de reclamaciones y la cuantía de las mismas. La utilidad se pone de manifiesto en que la función de densidad de la cantidad total reclamada puede obtenerse de manera exacta y recursiva mediante una expresión similar a la fórmula de recursión de Panjer (Panjer, 1981).

El modelo *BNIG* se muestra mucho más versátil que el modelo *BIG*, y es que la versión bivariada de este último presenta la desventaja de ser simétrica, esto es, $f(x_1, x_2) = f(x_2, x_1)$, lo que no resulta adecuado para modelizar datos que no satisfagan esta relación.

Capítulo 4

Ilustraciones numéricas

En este capítulo se muestran algunos ejemplos con los que se pretende ilustrar los modelos y métodos desarrollados en los capítulos precedentes. En primer lugar se analizará la sensibilidad, desde un punto de vista local, de primas de seguros y primas bonus-malus. En segundo lugar, analizaremos el buen comportamiento de la distribución \mathcal{BNIG} tanto en el ajuste de la distribución empírica de frecuencias, en el caso univariado y en el caso bivariado, así como en el cálculo de primas bonus-malus, comparando con los modelos Poisson-gamma y con el Poisson-inversa gaussiana. Finalmente se llevará a cabo el estudio de sensibilidad local de este modelo para compararlo con el tradicional Poisson-gamma.

4.1. Robustez local. Modelo Poisson-gamma

Nos ocupamos en esta sección de estudiar la sensibilidad, desde un punto de vista local o infinitesimal, de las primas de seguros que se obtienen bajo el tradicional modelo Poisson-gamma (\mathcal{PG}) utilizando los principios de cálculo de prima neta y de varianza.

Bajo el modelo \mathcal{PG} se asume que el número de reclamaciones de un póliza de seguros o un asegurado perteneciente a una cartera de seguros está representado por una variable aleatoria X que sigue una distribución de Poisson con media $\theta > 0$. A su vez, se supone que este parámetro θ es desconocido y aleatorio, y que sigue en el colectivo, la cartera de seguros, una cierta distribución de probabilidad $\pi_0(\theta)$ para la que se asume la densidad de probabilidad gamma con parámetros $a > 0, b > 0$, que representaremos mediante $\mathcal{G}(a, b)$. Un modelo de esta naturaleza

ha sido utilizado en la literatura actuarial en numerosas ocasiones, destacando a Bühlmann (1967), Eichenauer et al. (1988), Gómez et al. (2000, 2002, 2006a), Heilmann (1989), Lemaire (1979, 1985), Ríos et al. (1999), Simon (1961); entre otros. El modelo asumido es, pues:

$$\begin{aligned} X &\sim \mathcal{P}(\theta), \quad \theta > 0, x = 0, 1, \dots, \\ \theta &\sim \mathcal{G}(a, b), \quad a > 0, b > 0, \end{aligned}$$

esto es:

$$\begin{aligned} f(x|\theta) &= e^{-\theta} \theta^x / x!, \\ \pi_0(\theta) &= b^a \theta^{a-1} e^{-b\theta} / \Gamma(a). \end{aligned}$$

Suponemos ahora que observamos el número de reclamaciones en que incurre el asegurado durante t períodos de tiempo (usualmente t medido en años) y que éstos son x_1, x_2, \dots, x_t . Asumiendo que $X_i|\theta$ son i.i.d. se tiene que la distribución a posteriori del parámetro θ viene dada por

$$\pi(\theta|x_1, \dots, x_t) \propto \theta^{a+x} e^{-(b+t)\theta}, \quad \mathbf{x} = \sum_{i=1}^t x_i,$$

es decir, la distribución a posteriori es de nuevo gamma con los parámetros actualizados

$$a \rightarrow a + \mathbf{x}, \tag{4.1}$$

$$b \rightarrow b + t. \tag{4.2}$$

Utilizando la expresión (1.8) se deduce que la prima de riesgo bajo el principio de varianza es $P(\theta) = \theta + 1$, mientras que de (1.11) se deduce que la prima colectiva es

$$PC = \frac{a(a+1) + 2ab + b^2}{b(a+b)}. \tag{4.3}$$

Finalmente, la prima Bayes de varianza se obtiene a partir de (4.3) reemplazando los parámetros a y b por los parámetros actualizados (4.1) y (4.2), resultando

$$PB = \frac{(a+x)((a+x)+1) + 2(a+x)(b+t) + (b+t)^2}{(b+t)((a+x) + (b+t))}.$$

Consideraremos ahora la cartera de seguros en Lemaire (1979), dada en la Tabla 4.1 que corresponde a una cartera de seguro de automóviles.

La media del número de reclamaciones es 0,1011, mientras que la varianza es 0,1074. Teniendo en cuenta ahora que bajo el modelo \mathcal{PG} asumido la distribución

Tabla 4.1: Frecuencias observadas (Lemaire, 1979)

Nº Siniestros	Observados
0	96978
1	9240
2	704
3	43
4	9
Más de 4	0
Total	106974

predictiva es una binomial negativa con parámetros a y $b/(b+1)$, y utilizando el método de los momentos se deduce que los estimadores de los parámetros a y b son $\hat{a} = 1,613$, $\hat{b} = 16,138$.

La figura 4.1 muestra las primas Bayes bajo el principio de varianza para $x = 1, 2, \dots, 6$ y $t = 1, 5$ y 10 . En dicha figura aparecen las normas obtenidas para $p = \infty$ y $q = 1$ y para el supremo esencial (esssup) correspondientes a $p = 1$ y $q = \infty$.

Estos son los casos extremos de menor y mayor norma, respectivamente. Se ha comparado las normas que resultan utilizando la clase de todas las distribuciones de probabilidad y la clase de las distribuciones unimodales con la misma moda que $\pi_0(\theta) = \mathcal{G}(a, b)$, con moda $\theta_0 = (a-1)/b$. Como era previsible, las normas se reducen considerablemente cuando se utiliza la clase de distribuciones unimodales.

Cuanto mayor es la prima mayor es el valor de la norma para la clase de todas las distribuciones de probabilidad, Γ^A . El efecto parece ser el mismo cuando se utiliza la clase de las distribuciones unimodales, Γ^U . El resultado es coherente con el hecho de que los valores más pequeños de las normas corresponden a situaciones en que existe menos conflicto entre los datos y la distribución a priori considerada. En el ejemplo, el parámetro θ representa el número medio de reclamaciones y es un hecho comprobado empíricamente que esta cantidad no es grande en la mayoría de los ramos de seguros, incluido el automovilístico.

Como dato curioso, y que merecería ser objeto de investigación hay que destacar que el hecho de incorporar la propiedad de unimodalidad no parece tener efecto en el entorno del valor $x = 2$. Finalmente, puesto que los valores de las normas no resultan excesivamente grandes podemos asegurar que el modelo \mathcal{PG} bajo el principio de varianza parece robusto.

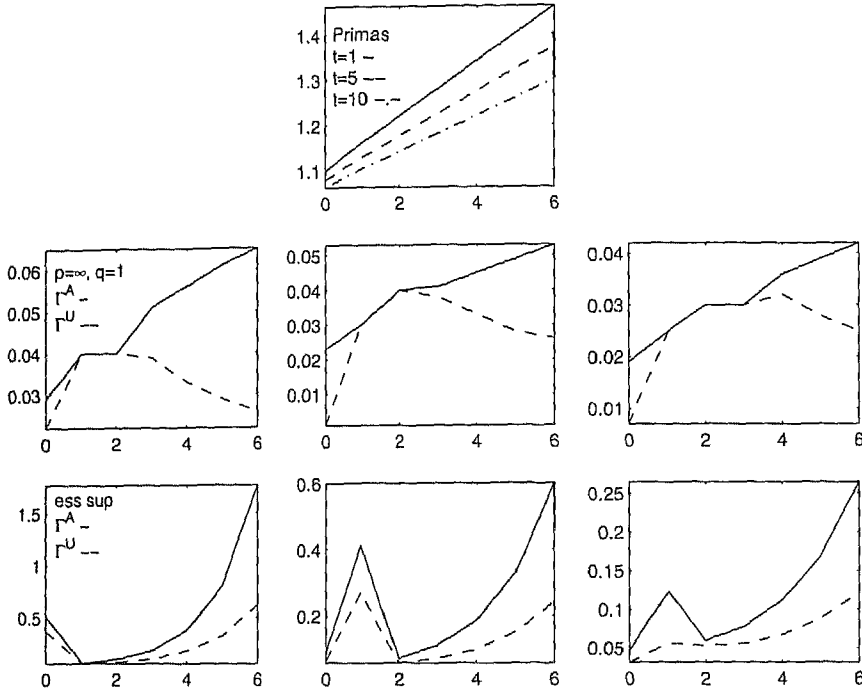


Figura 4.1: Primas y normas para $p = \infty$, $q = 1$ y supremo esencial

4.2. Robustez local de la PBM

Procedemos a desarrollar en esta sección un estudio similar al efectuado en la sección anterior para la prima bonus-malus. Utilizaremos, como es habitual en los SBM los principios de prima neta y de varianza. De (1.17) y (1.19) es fácil deducir que las PBM bajo el modelo $\mathcal{P}\mathcal{G}$ y los principios de prima neta y de varianza vienen dadas por

$$PBM = \frac{a + x b}{b + t a}, \tag{4.4}$$

$$PBM = \frac{(a + x)((a + x) + 1) + 2(a + x)(b + t) + (b + t)^2}{(b + t)((a + x) + (b + t))} \frac{b(a + b)}{a(a + 1) + 2ab + b^2}, \tag{4.5}$$

respectivamente.

Ahora hemos utilizado la cartera de seguro de automóviles en Bühlmann (1970, p.107) dada en la Tabla 4.2 y que aparece también en Klugman, et al. (1988), pp.

245 y Denuit (1997), pp. 240. que ha sido utilizada ampliamente en el ramo de

Tabla 4.2: Frecuencias observadas (Bühlmann, 1970)

Nº Siniestros	Observados
0	103704
1	14075
2	1766
3	255
4	45
5	6
6	2
Más de 7	0
Total	119853

seguro de automóviles para calcular primas bonus-malus. Los datos corresponden a una cartera de seguro de automóviles en Suiza y en el año 1961. La media y la varianza del número de reclamaciones son 0,155 y 0,179, respectivamente, de donde de nuevo utilizando el hecho de que la distribución predictiva es binomial negativa se deducen los estimadores de los parámetros del modelo $\hat{a} = 0,766595$, $\hat{b} = 3,40513$. Utilizando (4.4) y (4.5) se obtienen las PBM bajo los principios de prima neta y de varianza que aparecen en la Tabla 4.3.

Tabla 4.3: PBM bajo los principios de prima neta (en negrita) y de varianza y el modelo \mathcal{PG}

$t \setminus x$	0	1	2	3	4
1	0.77 0.94	1.78 1.14	2.78 1.34	3.79 1.53	4.80 1.71
2	0.62 0.91	1.45 1.07	2.27 1.23	3.09 1.38	3.91 1.53
3	0.53 0.88	1.22 1.02	1.91 1.15	2.61 1.28	3.30 1.41

Dichas primas aparecen expresadas en tanto por uno. Las primas obtenidas para el principio de prima neta aparecen en negrita en la primera fila de la celda y bajo el principio de varianza en la segunda fila de cada celda. Puede observarse

que cuando el asegurado cambia de clase, es decir cambia su valor de \mathbf{x} , la prima cargada para el asegurado es mayor para el principio de prima neta que para el principio de varianza. Sin embargo cuando el valor de t aumenta, la bonificación es mayor bajo el principio de prima neta que para el principio de varianza. Este asunto, aparentemente trivial, tiene suma importancia en el ramo de seguro de automóviles, pues empíricamente se ha comprobado que los asegurados prefieren pagar un poco más en el período actual conociendo que si experimentan reclamación su penalización no sea excesivamente alta. Este problema se le conoce en la literatura actuarial como el de sobrecargas (Lemaire, 1979; Sarabia et al., 2004; entre otros). Por tanto, podemos señalar que el principio de varianza presenta menores sobrecargas que el principio de prima neta bajo el modelo \mathcal{PG} , lo que lo hace ser mejor candidato a utilizarse en los SBM que el principio de prima neta.

Pasemos ahora a realizar el estudio de sensibilidad de este modelo. Como ya vimos en el capítulo 2, la función de influencia bajo el principio de prima neta es

$$I(\theta) = g(\theta) \frac{E_{\pi_0}[h(\theta)]}{E_{\pi_0}[g(\theta)]E_{\pi_0^*}[h(\theta)]} [1 - PBM] \frac{\pi_0^x(\theta)}{\pi_0(\theta)},$$

la cual es siempre positiva o siempre negativa dependiendo de los valores de (\mathbf{x}, t) .

En este caso la norma de la derivada de la prima bonus-malus en 0 en el caso $p = \infty$ viene dada en el Teorema 2.7.

Usando el Corolario 2.1, hemos calculado las normas de las PBM. Estos resultados se muestran para el principio de prima neta en la Tabla 4.4.

Tabla 4.4: Normas de la PBM bajo el principio de prima neta y el modelo \mathcal{PG}

$t \setminus \mathbf{x}$	0	1	2	3	4
1	0.175	1.391	4.992	10.627	18.294
	0.221	2.616	14.498	48.405	131.839
	0.451	13.152	186.536	1619.7	11572.3
2	0.233	0.655	2.895	6.485	11.426
	0.273	1.025	6.262	19.707	49.198
	0.430	2.728	29.223	163.822	734.878
3	0.249	0.275	1.762	4.210	7.621
	0.283	0.390	3.217	10.076	24.111
	0.414	0.815	9.988	47.186	171.784

Bajo el principio de varianza la función de influencia es también siempre positiva o negativa como puede ser visto representando la gráfica de $I(\theta)$. En la Tabla 4.5 se muestran las normas de la derivada en 0 de las PBM bajo el principio de varianza.

Tabla 4.5: Normas de la PBM bajo el principio de varianza y el modelo \mathcal{PG}

$t \setminus x$	0	1	2	3	4
1	0.050	0.191	0.608	1.234	2.104
	0.052	0.350	1.931	6.635	18.872
	0.085	2.465	41.511	411.057	3281.12
2	0.075	0.084	0.350	0.741	1.274
	0.076	0.123	0.768	2.428	6.193
	0.081	0.344	4.280	26.272	127.714
3	0.090	0.025	0.211	0.478	0.837
	0.092	0.033	0.376	1.174	2.828
	0.102	0.070	1.229	6.212	23.951

Los datos aparecen en el siguiente orden de arriba a abajo: las normas para $p = \infty$, $p = 2$ y $p = 1$, respectivamente. Todas las normas $p = 1$, $p = 2$ y $p = \infty$ aumentan con el valor de x y disminuyen con el valor de t , mostrando gran dependencia con el tamaño muestral. Se puede observar que la diferencia entre las normas $p = 1$ y $p = 2$ aumentan con el valor de x . Efectos similares pueden observarse para la diferencia entre las normas para $p = 1$ y $p = \infty$ y para la diferencia entre las normas para $p = 2$ y $p = \infty$.

Observamos que los valores más grandes de $\|P\dot{B}M(0)\|_p$ aparecen para $x \geq 2$, reflejando el hecho de que cuando la distribución a priori está en conflicto con los datos no se garantizan resultados robustos. Por tanto, creemos necesario elegir una distribución a priori que de más peso a la región en dónde el tamaño de la muestra esté en conflicto con la media de la distribución a priori. Esto se puede conseguir usando una distribución a priori que pueda construirse como una suma convexa de dos o más distribuciones a priori.

Otra conclusión interesante que puede inferirse de los Tablas 4.4 y 4.5 es que el principio de varianza es más robusto que el principio de prima neta al tomar la norma de la derivada como medida de robustez.

Ahora, las normas $\|P\dot{B}M(0)\|_\infty$ que aparecen en la primera fila de cada celda en la Tabla 4.4 para el principio de prima neta y en la Tabla 4.5 para el principio de varianza puede calcularse usando los Corolarios 2.2 y 2.4, respectivamente, sin

hacer uso del Corolario 2.1.

Además, si un decisor experto considera que una PBM para la que

$$\|PBM(0)\|_{\infty} > 12$$

no es robusta podría decidir que todas las primas calculadas bajo el principio de prima neta son 11-robustas, excepto las primas obtenidas para $(x, t) = (4, j)$, $j = 1, 2$. Puede observarse que bajo el principio de varianza todas las primas son 2.2-robustas.

4.3. Aplicaciones de la distribución *BNTG*

En esta sección analizaremos el comportamiento de la distribución *BNTG* (Gómez et al., 2005c, 2005d, 2007) como una alternativa competitiva al modelo Poisson-inversa gaussiana (Hougaard et al. (1997), Tremblay (1992) y Willmott (1987, 1988)). Los ejemplos que se llevarán a cabo son los siguientes:

1. Ajustes de la distribución *BNTG* en el caso univariado.
2. Ajustes de la distribución *BNTG* en el caso bivariado.
3. Análisis de regresión.
4. Cálculo de PBM.
5. Análisis de robustez local de las PBM.

4.3.1. Ajustes para el caso univariado

Los datos de este primer ejemplo se refieren al número de pólizas de seguros de automóviles privados en Suiza que aparecieron en la Tabla 4.2 (ver Bühlmann (1970), Klugman, et al. (1988), pp. 245 y Denuit (1997), pp. 240). Estos datos aparecen de nuevo en la Tabla 4.6 (primera y segunda columnas). Como puede observarse, estos datos son fuertemente sesgados a la derecha y sobredispersos puesto que, como ya se señaló anteriormente, la media y la varianza del número de reclamaciones son 0,155 y 0,179, respectivamente. Por lo tanto, parece razonable que estos datos deban ajustarse por medio de una distribución sobredispersa. Por esta razón, creemos que la distribución *BNTG* es adecuada para el ajuste.

Utilizando las ecuaciones apropiadas presentadas en la sección de estimación del capítulo anterior se han estimado los parámetros del modelo por el método de la proporción de ceros, el de los momentos y el de máxima verosimilitud. Puede observarse que la estimación por el método de los momentos (MM), el método de la proporción de ceros (PM) y el método de máxima verosimilitud (MLE) nos proporcionan un ajuste casi perfecto. El método de los momentos proporciona dos

Tabla 4.6: Siniestros observados y ajustados

Nº Siniestros	Obs.	%	Ajust. (1)	Ajust. (2)	Ajust. (3)	Ajust. (4)
0	103704	86.52	103711	103708	103704	103704
1	14075	11.74	14051.1	14060.9	14071.5	14074.7
2	1766	1.47	1789.67	1779.25	1773.31	1770.8
3	255	0	251.32	254.2	253.26	251.95
4	45	0	40.17	41.08	41.45	41.76
5	6	0	7.31	7.32	7.61	7.98
6	2	0	1.50	1.40	1.53	1.72
> 6	0	0	0	0	0	0
Total	119853	100	119853	119853	119853	119853
χ^2			0.5516	0.1655	0.0877	0.0825
g.l.			2	2	2	2
p - valor			0.7604	0.9210	0.9573	0.9598
Estimadores			MM (1)	MM (2)	PM	MLE
$\hat{\mu}$			2.085	28.12	12.5	3.7381
$\hat{\psi}$			0.2253	0.0059	0.0122	0.075
$\hat{\mu}$			0.0709	0.0054	0.0148	0.04022
$L_{\text{máx}}$						-54609.8

soluciones, y ambas producen un ajuste similar. Obviamente, el método de máxima verosimilitud produce el mejor ajuste.

Comparando estos resultados con los ajustes producidos por otros modelos, y que pueden verse en Klugman et al. (1988) y Denuit (1997), puede afirmarse que la distribución *BNIG* nos da un mejor ajuste para estos datos si tomamos el test- χ^2 como criterio de comparación. En este caso, y como es habitual cuando se utiliza este test, se han agrupado en una sola clase las clases para $x \geq 4$. Esta elección también se confirma por sus respectivos p -valores. Finalmente, el máximo del logaritmo de la función de verosimilitud es mayor en la distribución binomial negativa-inversa gaussiana que en las distribuciones Poisson y binomial negativa e igual a la distribución Poisson-inversa gaussiana (ver Klugman, et al. (1988), pp. 245).

La Tabla 4.7 muestra datos similares del número de siniestros en una cartera de seguros de automóviles (Klugman et al. (1988), pp. 244; Simon (1961)).

Las observaciones se muestran en la primera y segunda columna. Estos datos

son también sobredispersos y sesgados a la derecha, con una cola más gruesa que la del ejemplo anterior.

La cuarta y quinta columnas muestran los valores ajustados por el método de los momentos (MM) y el de máxima verosimilitud (MLE), respectivamente. Como puede observarse, ambos métodos de estimación producen un ajuste casi idéntico a los datos. Los valores estimados de los parámetros \hat{r} , $\hat{\psi}$ y $\hat{\mu}$ se muestran en las tres últimas filas. Debemos advertir que el método de la proporción de ceros no puede usarse en este caso puesto que la proporción de ceros en la muestra no es alta (33.22% del total de observaciones). Comparando estos resultados con los valores de los ajustes en Klugman et al. (1988), pp. 244, observamos que el ajuste por medio de la distribución *BNTG* produce una significativa mejora que la que se produce al ajustar los datos a través de las distribuciones Poisson y binomial negativa teniendo en cuenta el bajo valor del estadístico χ^2 . Sin embargo, la distribución Polya-Aeppli produce un mejor ajuste para estos datos que los que se consiguieron para la distribución *BNTG* si se toma el test de la χ^2 como criterio de comparación.

Tabla 4.7: Siniestros observados y ajustados

N° Siniestros	Observados	%	Ajustados (1)	Ajustados (2)
0	99	33.22	95.31	95.34
1	65	21.81	76.08	76.40
2	57	19.12	50.62	50.78
3	35	11.74	31.44	31.44
4	20	6.71	18.83	18.75
5	10	3.35	11.03	10.93
6	4	1.34	6.36	6.27
7	0	0.00	3.36	3.55
8	3	1.00	2.05	2.00
9	4	1.34	1.15	1.11
10	0	0.00	0.64	0.62
11	1	0.00	0.36	0.34
12	0	0.00	0.20	0.19
Total	298	100	298	298
χ^2			4.0085	3.99546
g.l.			4	4
p - valor			0.4143	0.4160
Estimadores			MM	MLE
\hat{r}			1.5	1.51787
$\hat{\psi}$			2224.82	3059.91
$\hat{\mu}$			0.760099	0.75091
$L_{\text{máx}}$				-528,786

4.3.2. Ajuste para el caso bivariado

Para mostrar cómo se comporta la distribución *BNTG* en el caso bivariado, hemos analizado datos referentes a absentismo laboral que aparecen en Arbous y Sichel (1954) y Stein et al. (1987). Estos datos se refieren al número de ausencias al trabajo (absentismo) de 248 trabajadores correspondientes a los años 1947 y 1948 en una empresa norteamericana. Los datos de la muestra se adaptan a las hipótesis de regresión lineal de la distribución *BNTG* puesto que presentan correlación positiva, como se puede comprobar. Hemos estimado los cuatro parámetros de la distribución usando el método de los momentos (MM) y de máxima verosimilitud (MLE). Por razones computacionales, hemos agrupado los datos originales en la forma que se muestra en la Tabla 4.8 y en la Tabla 4.9. El método de los momentos nos proporciona un buen ajuste, como puede observarse en la Tabla 4.9. Esta Tabla compara las frecuencias observadas con las frecuencias esperadas (en paréntesis) para cada celda de la distribución muestral. Al usar esta agrupación de los datos muestrales hemos obtenido un estadístico de bondad de ajuste de χ^2 con 21 grados de libertad de 18.86, con un p -valor de 0.5994.

Obviamente, se consigue un mejor ajuste de los datos usando el método de máxima verosimilitud (MLE), como puede observarse en la Tabla 4.9. Nuevamente, las frecuencias esperadas muestran que el ajuste es muy bueno. Podemos observar que el ajuste de la distribución *BNTG* por el método de máxima verosimilitud (MLE) nos da una significativa mejora sobre el método de los momentos (MM) al juzgar por su bajo valor del estadístico χ^2 , con un p -valor de 0.7781. Como puede observarse, hemos obtenido un valor del estadístico χ^2 de 15,64 y el máximo de el logaritmo de la función de verosimilitud es $L_{\max} = -1181,42$.

Finalmente, merece destacarse que este modelo ha mejorado apreciablemente los resultados obtenidos por la extensión multivariada de la distribución de Sichel en Stein et al. (1987): $L_{\max} = -1194,349$ y $\chi^2 = 16,9$, respectivamente. Por lo tanto, creemos que la distribución bivariada *BNTG* puede ser una buena alternativa, o por lo menos un modelo competitivo, para modelar datos de esta naturaleza, según nuestro criterio.

Todos los cálculos han sido realizados usando el programa MATHEMATICA. En el caso univariado se ha usado el método Newton–Raphson para estimar los parámetros de la distribución por el método de máxima verosimilitud, tomando sus estimadores por el método de los momentos como valores iniciales. En el caso bivalente, los estimadores máximo verosimilitud se calcularon usando el método *scoring* (Klugman et al. (1988), pp. 58–59 and pp. 217) con los estimadores hallados por el método de los momentos como valores iniciales. Este método presenta la ventaja de no precisarse de la segunda derivada para aproximar la solución, aunque la convergencia al valor óptimo es muy lenta.

Tabla 4.8: Frecuencias observadas y esperadas de los datos de absentismo (Método de los momentos)

1947	1948							
	0	1	2	3	4-5	6-8	9-13	14-27
0	7 (5.33)	7 (6.08)	10 → (16.31)					
1	9 (6.39)	6 (7.65)	8 (6.57)	11 → (16.9)				
2	9 (5.43)	9 (6.82)	6 (6.11)	5 (4.83)	7 → (12.89)			
3		← 10 (9.33)		← 11 (9.05)	11 → (12.14)			
4-5		← 11 (11.57)	6 (6.61)	6 (5.68)	10 (8.17)	13 → (11.48)		
6-8					← 27 (28.71)	7 → (5.82)		
9-13						← 26 (19.01)	3 (1.84)	
14-27	13 → (8.78)							
$\chi^2 = 18.86$; g.l. = 19; p -valor = 0,5994								
Estimadores $\hat{r}_1 = 2,85167$ $\hat{r}_2 = 2,71239$ $\hat{\psi} = 5,68473$ $\hat{\mu} = 0,899465$								

4.3.3. Regresión

La ecuación (3.22), así como los estimadores de máxima verosimilitud se utilizaron ahora para predecir el número medio esperado de absentismo laboral en 1948 dado el número de absentismo observado en 1947. Arbous y Sichel (1954) (véase también Johnson, Kotz y Balakrishnan (1997)) introducen una distribución bivalente con función de probabilidad dada por

$$\Pr(X_1 = x_1, X_2 = x_2) = \left(\frac{\phi}{\theta + 2\phi} \right)^\theta \frac{(\theta - 1 + x_1 + x_2)!}{(\theta - 1)!x_1!x_2!} \left(\frac{\phi}{\theta + 2\phi} \right)^{x_1 + x_2},$$

$$x_1, x_2 = 0, 1, \dots$$

La ecuación de regresión de X_2 dado X_1 es,

$$E(X_2|X_1 = x_1) = \frac{\phi}{\theta + \phi}(\theta + x_1), \quad (4.6)$$

Tabla 4.9: Frecuencias observadas y esperadas de los datos de absentismo (Método de máxima verosimilitud)

1947	1948							
	0	1	2	3	4-5	6-8	9-13	14-27
0	7 (7.86)	7 (7.55)	10 → (6.44)					
1	9 (9.38)	6 (10.07)	8 (7.36)	11 → (11.71)				
2	9 (7.41)	9 (8.83)	6 (7.11)	5 (4.93)	7 → (8.87)			
3		← 10 (11.56)	9 (5.76)	2 (4.32)	11 → (9.26)			
4-5		← 11 (12.63)	6 (7.36)	6 (6.09)	10 (8.04)	13 → (8.62)		
6-8					← 27 (22.61)	7 → (5.02)		
9-13	29 → (22.37)							
14-27	13 → (11.12)							
$\chi^2 = 15,64$; g.l. = 20; p -valor = 0,7781 $L_{\max} = -1181,42$								
Estimadores $\hat{r}_1 = 5,500$ $\hat{r}_2 = 4,420$ $\hat{\psi} = 1,527$ $\hat{\mu} = 0,573$								

que resulta lineal en x_1 . Los valores obtenidos utilizando (3.22) (y denotados como Esperados (1)) se han comparado con los valores obtenidos bajo la línea de regresión (4.6) y que son denotados ahora como Eperados (2). El absentismo predicho en 1948 bajo ambos modelos aparece en la Figura 4.2. La diferencia entre los valores predichos y los verdaderos valores se muestran a la derecha. Como conclusión puede destacarse que ambos modelos predicen en términos generales valores similares.

Luego, el modelo de regresión derivado del modelo \mathcal{BNIG} parece tener también un buen comportamiento.

4.3.4. Cálculo de PBM bajo el modelo \mathcal{BNIG}

A continuación desarrollaremos un ejemplo para el cálculo de las primas bonus malus para el modelo \mathcal{BNIG} bajo el principio de prima neta y de varianza. Aunque la distribución a priori \mathcal{IG} no es conjugada con la verosimilitud \mathcal{BN} , la distribución

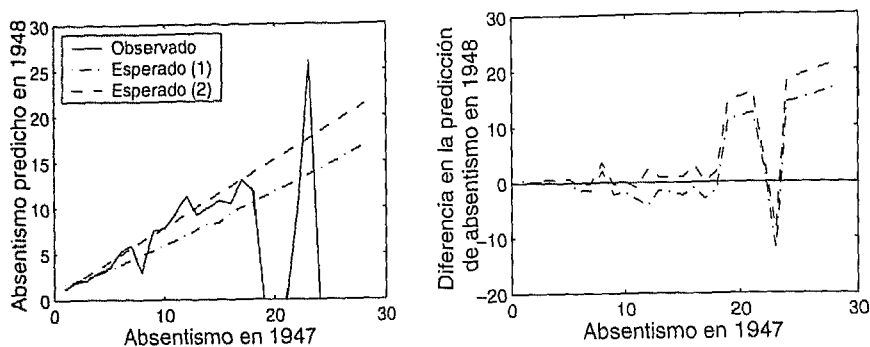


Figura 4.2: Absentismo predicho en 1948 y diferencias entre nuestro modelo y el de Arbous y Sichel (1954)

a posteriori puede obtenerse fácilmente a partir del Teorema 3.5.

Usando los datos de la Tabla 4.6 hemos calculado las PBM bajo el principio de prima neta y de varianza para el modelo $BNIG$, que se muestran en la Tabla 4.10.

Una cuestión interesante que merece ser estudiada en detalle llegado a este punto es el problema de las sobrecargas, mencionado anteriormente. Para ello vamos a calcular las primas bonus-malus de los modelos PG y Poisson-inversa gaussiana (PIG) bajo los principios de varianza y prima neta respectivamente. Este último modelo es utilizado habitualmente en el escenario de los SBM (Gómez y Vázquez, 2003; Tremblay, 1992; entre otros).

Las PBM para el modelo PG para los principios de prima neta y varianza se muestran en la Tabla 4.11. Se ha tomado como estimadores de los parámetros los obtenidos por el método de máxima verosimilitud que aparecen en Klugman et al. (1998). Las estimaciones de los parámetros son $\hat{a} = 0,766$ y $\hat{b} = 3,405$.

Las PBM para el modelo PIG para los principios de prima neta y varianza se muestran en la Tabla 4.12. En este caso hemos usado la estimación de los parámetros por máxima verosimilitud que aparece en Tremblay (1992), que son $\hat{\mu} = 0,155$ y $\hat{\psi} = 0,155$.

Puesto que no es objeto de esta Memoria el modelo PIG , se ha prescindido de los detalles del mismo, aunque apuntamos que consiste en la mezcla de la distribución de Poisson con la inversa gaussiana. El cálculo de primas bajo este modelo resulta atractivo en cuanto la distribución a posteriori, considerando como verosimilitud la Poisson y distribución a priori la inversa gaussiana, es la generalizada inversa gaussiana. Luego el modelo es conjugado.

Como ya se comentó antes, en Lemaire (1979) se enfatiza en el problema de las sobrecargas que aparecen en los SBM. Este problema consiste en estudiar cómo

Tabla 4.10: PBM bajo los principios de prima neta (en negrita) y de varianza bajo el modelo *BNI*G

$t \setminus x$	0	1	2	3	4
1	0.92 0.94	1.38 0.99	1.99 1.03	2.71 1.08	3.50 1.14
2	0.86 0.98	1.26 1.02	1.79 1.06	2.42 1.11	3.10 1.17
3	0.81 0.98	1.17 1.01	1.63 1.05	2.18 1.09	2.79 1.14

Tabla 4.11: PBM bajo los principios de prima neta (en negrita) y de varianza bajo el modelo *PG*

$t \setminus x$	0	1	2	3	4
1	0.77 0.94	1.78 1.14	2.78 1.34	3.79 1.53	4.80 1.72
2	0.63 0.91	1.45 1.07	2.27 1.23	3.09 1.38	3.91 1.53
3	0.53 0.88	1.22 1.02	1.91 1.15	2.61 1.28	3.30 1.41

varía la prima cuando los asegurados cambian de clase, especialmente de la clase $x = 0$ a $x = 1$. Como puede observarse a partir de la Tabla 4.13, bajo el principio de prima neta y el principio de varianza, el modelo *BNI*G produce menores sobrecargas para asegurados en la clase $x = 0$, siendo éstas menores para el principio de varianza. El problema de las sobrecargas, como ya se señaló, surge porque los asegurados de la clase bonus $x = 0$ prefieren pagar una prima ligeramente mayor como contrapartida a no sufrir una penalización muy elevada en caso de experimentar reclamación. Por tanto, el modelo *BNI*G es más competitivo en este terreno que los tradicionales *PG* y *PI*G puesto que presenta menores sobrecargas.

Tabla 4.12: PBM bajo los principios de prima neta (en negrita) y de varianza bajo el modelo \mathcal{PIG}

$t \setminus x$	0	1	2	3	4
1	0.87	1.63	2.75	4.09	5.52
	0.97	1.09	1.26	1.46	1.66
2	0.78	1.40	2.29	3.35	4.50
	0.96	1.05	1.18	1.34	1.50
3	0.71	1.23	1.97	2.85	3.80
	0.95	1.02	1.13	1.26	1.40

Tabla 4.13: Porcentaje de incremento de la prima cuando el asegurado pasa de $x = 0$ a $x = 1$ bajo el principio de prima neta (en negrita) y el de varianza

t	\mathcal{PG}	\mathcal{PIG}	\mathcal{BNIG}
0	78.13	63.64	38.56
	14.61	9.61	3.13
1	87.82	60.51	37.07
	13.64	8.05	2.79
2	94.49	14.61	57.50
	12.42	6.89	2.51
3	99.32	54.79	34.46
	11.28	6.02	2.29

4.3.5. Análisis de robustez

Se ha procedido a calcular finalmente las normas de las PBM bajo el modelo \mathcal{BNIG} obteniéndose los datos que aparecen en las Tablas 4.14 y 4.15. Comparando con la Tabla 4.4 que nos daba las normas de las PBM bajo el tradicional modelo \mathcal{PG} se deduce que el nuevo modelo es mucho más robusto que el tradicional \mathcal{PG} .

4.4. Conclusiones

Resulta interesante destacar como conclusiones finales de este capítulo que:

- Aunque las conclusiones, en cuanto a robustez local se refiere, no difieren

Tabla 4.14: Normas de la PBM bajo principio de prima neta del modelo *BNTG*

$t \setminus x$	0	1	2	3	4
1	0.069	0.534	1.983	4.662	8.808
	0.080	0.923	5.77	23.430	75.838
	0.238	11.740	270.685	3515.43	32440
2	0.117	0.340	1.430	3.441	6.560
	0.128	0.510	3.305	12.684	38.568
	0.227	2.442	40.977	377.686	2578.18
3	0.151	0.204	1.048	2.605	5.024
	0.159	0.277	2.204	7.658	22.297
	0.230	0.841	13.263	100.645	569.86

Tabla 4.15: Normas de la PBM bajo principio de varianza del modelo *BNTG*

$t \setminus x$	0	1	2	3	4
1	0.007	0.041	0.129	0.272	0.492
	0.008	0.053	0.279	1.096	3.683
	0.009	0.563	17.287	291.005	3400.36
2	0.013	0.027	0.096	0.205	0.368
	0.014	0.031	0.168	0.601	1.826
	0.015	0.092	1.688	17.868	141.93
3	0.017	0.016	0.073	0.159	0.285
	0.018	0.018	0.110	0.373	1.056
	0.036	0.036	0.520	4.175	25.889

de las realizadas anteriorente bajo el análisis de robustez global, el análisis local resulta muchísimo más flexible, permitiendo el estudio sin la necesidad de introducir clases refinadas de distribuciones a priori. Esto es, el análisis local se adapta mejor a estimadores Bayesianos que tengan una formulación relativamente complicada, como así ocurre con el cociente de magnitudes a posteriori o con los estimadores bonus-malus.

- La distribución *BNTG* produce estimadores por momentos –más sencillos de obtener que los de máxima verosimilitud– que generan buenos ajustes para

las distribuciones empíricas de frecuencias que se han analizado.

El estudio de regresión es coherente con modelos alternativos, las PBM generan menores sobrecargas que con los tradicionales modelos \mathcal{PG} y \mathcal{PIG} , y además son muchísimo más robustas.

Capítulo 5

Conclusiones y líneas abiertas

Exponemos finalmente en este capítulo las conclusiones fundamentales de esta Memoria, así como algunas líneas de investigación abiertas y que bajo nuestro punto de vista merecerían ser estudiadas.

5.1. Conclusiones

En la presente Memoria se ha pretendido abordar desde una nueva perspectiva el problema de la robustez para primas de seguros basada en la sensibilidad o robustez local. La búsqueda de nuevos modelos de los cuales puedan derivarse primas que sean robustas, nos ha conducido a nuevas distribuciones de probabilidad discretas con las que se obtienen buenos ajustes de distribuciones empíricas de frecuencias, relacionadas éstas con el número de siniestros en que incurre un asegurado.

El cálculo de la prima colectiva y de la prima individual de un asegurado basada en su propia experiencia de siniestralidad (prima Bayes) requiere de una distribución estructura o distribución a priori. Es sabido que en muchas ocasiones el actuario tiene dificultad para especificar una única distribución a priori, bien porque no está seguro sobre la misma, porque no la conoce perfectamente o porque la decisión sobre la misma deba ser tomada por más de un decisor que no tienen porqué coincidir en su decisión final. Por este motivo se introduce el análisis de sensibilidad o de robustez bayesiano.

Este análisis tiene por objetivo medir la variación que con respecto a la distribución a priori se produce cuando el investigador manifiesta imprecisión en la

especificación de una única distribución a priori.

Desde el punto de vista bayesiano el análisis de sensibilidad puede realizarse desde dos puntos de vista, global y local. En el primer caso el estudio consiste en estudiar la variación de la magnitud a posteriori de interés (la prima, en nuestro caso) utilizando clases de distribuciones de probabilidad, y en el segundo perturbando infinitesimalmente la distribución a priori.

En esta Memoria nos hemos decantado por el estudio de sensibilidad local de la magnitud a posteriori. Creemos que esta técnica presenta la ventaja de permitir el análisis de robustez cuando las características del problema hacen poco tratable el estudio de sensibilidad global.

Por otra parte, la modelización inicial que se asuma para el problema debe ser coherente con los datos que se dispone. En este sentido, es muy frecuente trabajar en Estadística Actuarial con modelos que sean sobredispersos, esto es $\text{Var}(X) > \mathbb{E}(X)$. Por esta razón se ha dedicado también parte de esta Memoria a buscar nuevos modelos que sean sobredispersos, obteniéndose las distribuciones binomial negativa–inversa gaussiana y binomial–inversa gaussiana, en sus versiones univariadas y multivariadas.

Los siguientes comentarios recogen las conclusiones más importantes que pueden extraerse cuando analizamos la sensibilidad local de los funcionales (1.12) y (1.16) al perturbar infinitesimal y linealmente la distribución a priori mediante una medida signada. Así, merece ser destacado que las expresiones (1.12) y (1.16) se pueden contemplar como funcionales dependientes de la distribución a priori inicial. El análisis de sensibilidad local consiste básicamente, en calcular la derivada del funcional con respecto a una medida signada, cuantificando la magnitud de la perturbación por medio de la norma de la derivada. La derivada funcional utilizada es la de Fréchet. Para medir la norma de la derivada nos hemos decantado en esta Memoria por usar los espacios L_p con norma asociada la definida en el capítulo 1.

Algunas de las conclusiones obtenidas en el análisis de robustez local son:

- Si en la expresión (1.12) tanto gf como hf son integrables entonces existe la derivada en 0 del funcional, pudiéndose expresar ésta en términos de una función de influencia.
- La medida de robustez del funcional (1.12) viene dada por el máximo de la norma asociada a los espacios L_p de la función de influencia. Dependiendo de los valores de p y q obtenemos cantidades diferentes para el valor de la norma.

- Puede reducirse la clase de distribuciones utilizada incorporando la propiedad de unimodalidad, reduciéndose con ello los valores de las normas.
- Si en la expresión (1.16) tanto gf como hf son integrables entonces existe la derivada en 0 del funcional, pudiéndose expresar ésta también en términos de una función de influencia.
- La medida de robustez del funcional (1.16) viene dada por el máximo de la norma asociada a los espacios L_p de la función de influencia. Dependiendo de los valores de p y q obtenemos también diferentes medidas de robustez.
- Existen tres valores de p que tienen una especial relevancia. El caso $p = 1$ es el caso de mayor norma. El caso $p = 2$ situa el problema de la medida de robustez en un espacio de Hilbert. Un significado especial tiene el caso $p = \infty$, correspondiente a la norma del supremo esencial, y que generará la menor norma.
- Cuando la función de influencia es siempre positiva o siempre negativa, la norma del funcional (1.16) para el caso del supremo esencial puede obtenerse de una manera sencilla para el principio de prima neta, independientemente del modelo asumido.

En el capítulo 3 de esta Memoria estudiamos un problema de inferencia. Los resultados más importantes obtenidos fueron:

- Se obtuvo un nuevo modelo que consiste básicamente en la mezcla o composición de las distribuciones binomial negativa con parámetros $r > 0$ y $p = e^{-\theta}$, $\theta > 0$, e inversa gaussiana. Esta nueva distribución depende de tres parámetros en el caso univariado.
- La nueva distribución obtenida es sobredispersa y sus parámetros pueden estimarse de una manera relativamente sencilla por el método de los momentos y el de máxima verosimilitud.
- Para esta distribución se obtuvo una fórmula recursiva para el cálculo de la función de probabilidad que nos permite expresar la distribución de la cantidad total reclamada en términos de una ecuación integro diferencial, en la misma línea que el trabajo de Panjer (1981).
- Esta distribución admite una extensión multivariada para la que tampoco resulta difícil estimar los parámetros del modelo. Además la versión multivariada de la nueva distribución presenta correlación estrictamente positiva.

- En este mismo capítulo se ha estudiado asimismo la extensión multivariada del modelo resultante de mezclar las distribuciones binomial con parámetro n y $p = 1 - e^{-\theta}$, $\theta > 0$, e inversa gaussiana. Esta extensión presenta correlación estrictamente positiva, pero tiene la desventaja de no ser sobredisperso, además de ser menos flexible que el modelo multivariado binomial negativa-inversa gaussiana ya que su función de probabilidad es simétrica. Así, en el caso bivariado se tiene que $f(x, y) = f(y, x)$, lo que no resulta adecuado para modelar datos que no satisfagan esta relación.

A continuación se exponen una serie de puntos que creemos pueden ser objeto de futuros trabajos de investigación.

5.2. Sobre la distribución perturbadora

En esta Memoria hemos escogido como clase de perturbación

$$\Gamma = \{\pi : \pi = \pi_0 + \delta\},$$

es decir, la clase de las distribuciones que resulta de perturbar la distribución a priori inicial linealmente mediante una medida δ . Un trabajo que merecería ser objeto de investigación sería la incorporación de perturbaciones no lineales sobre la distribución a priori, siguiendo la línea desarrollada por Gustafson (1996a). Es decir, considerar

$$\Gamma = \{\pi : \pi = k\pi_0^\varepsilon \delta^{1-\varepsilon}, \varepsilon \in [0, 1]\},$$

siendo k una constante de normalización.

Este tipo de perturbación no ha sido contemplado nunca bajo la óptica del análisis bayesiano robusto global, y desde un punto de vista local sólo para el estudio de variación de la media a posteriori (Gustafson (1996a)).

5.3. Sobre el comportamiento asintótico de las normas

Un estudio que merece ser estudiado es el comportamiento asintótico de las normas de las derivadas de (1.12) y de (1.16). En este sentido sería conveniente partir del trabajo de Gustafson y Wasserman (1995) en el que se estudia el comportamiento asintótico de las normas de la derivada del estimador media a posteriori. Se muestra en él que cuando el tamaño muestral aumenta, la medida de sensibilidad tiende a infinito. La importancia de este estudio radica en que el comportamiento asintótico de las PBM es un aspecto de suma importancia en la teoría de los SBM.

5.4. Sobre los principios de cálculo de primas

Los principios de cálculo de primas que se han utilizado en esta Memoria no son los únicos existentes aunque si los más utilizados en la literatura actuarial. Otros principios que merecen destacarse son el de desviación típica, el de pérdida máxima, el de valor esperado, etc. (ver Gerber (1979) y Sarabia et al. (2006)).

Un estudio similar al desarrollado en esta Memoria en el Capítulo 2, puede llevarse a cabo de una manera similar para los principios de cálculo de prima exponencial y Esscher. La importancia de ambos principios estriba en el hecho de que para el cálculo de la prima se requiere, en ambos casos, de la especificación de la constante de aversión al riesgo, llamada también medida de Arrow-Pratt. Es indudable que la utilización de diferentes valores para este parámetro alterará también el valor de las normas.

5.5. Sobre la incorporación de condiciones de momentos

En estadística actuarial la media es un concepto fundamental. Hemos visto como la PBM se obtiene a partir de momentos de determinadas funciones. Es seguro que un investigador con una buena formación estadística no debe tener dificultad en especificar, por ejemplo la media de la distribución a priori, utilizando para ello su experiencia profesional. Un estudio importante, aunque también difícil, en este contexto sería el considerar clases con condiciones sobre momentos para llevar a cabo posteriormente el análisis de robustez local, con el simple objetivo de reducir el valor de las normas. En el análisis global los estudios de esta naturaleza son escasos, destacando Betrò et al. (1995), Gómez y Vázquez (2005a), Gómez et al. (2002), Goutis (1994) y Moreno et al. (2003).

Un investigador con una buena formación en estadística no tendría ninguna dificultad en especificar el valor de la media del parámetro de riesgo y sus valores numéricos de acuerdo a su experiencia previa. También se podría especificar condiciones sobre los momentos usando la metodología que aparece en Betrò et al. (1996) y Moreno et al. (2003).

5.6. Realizar mezclas con diferentes distribuciones

El trabajo realizado en el capítulo 3 de esta Memoria puede extenderse al estudio de mezclas de diferentes distribuciones de probabilidad. En este sentido, podría mezclarse la distribución binomial negativa con la gamma, aunque en este caso los métodos de estimación no resultan fáciles de aplicar. Alanko y Duffi (1996)

realizaron un estudio similar mezclando las distribuciones binomial y gamma. El estudio multidimensional del modelo binomial-gamma sería conveniente de analizarse puesto que el problema de la estimación de los parámetros no presenta serios problemas.

Finalmente, es bien sabido (Balakrishnan y Nevzorov, 2003) que dos casos particulares de la distribución inversa Gaussiana, $\mathcal{IG}(\mu, \psi)$ se obtiene de: (1) Hacer $\mu \rightarrow \infty$ y (2) Hacer $\mu^2 = \psi$. Luego del modelo \mathcal{BNIG} expuesto en esta Memoria pueden derivarse modelos más sencillos mediante estas simples relaciones que proporcionarían un modelo con un parámetro menos a estimar.

Apéndice: Método *Scoring*

Cuando los datos están agrupados en clases, existe un método efectivo para maximizar la función de verosimilitud. Este método es similar al de Newton–Raphson y se le conoce en la literatura estadística como método *scoring*. A continuación, y siguiendo la línea argumental de Klugman et al. (1998) exponemos de forma resumida dicho método al caso univariado en primer lugar, y la extensión natural al caso multivariado a continuación.

Caso univariado

Presentamos aquí, un método efectivo, aunque a veces lento, de obtención de estimadores por el método de máxima verosimilitud, similar al de Newton–Raphson.

Supongamos, de entrada, que el parámetro a estimar es único, esto es, un escalar. La función de verosimilitud viene dada por:

$$l(\theta) = \sum_{j=1}^k n_j \log [P_j(\theta)],$$

entonces, el estimador de máxima verosimilitud satisface

$$dl/d\theta = l'(\theta) = \sum_{j=1}^k n_j P_j' / P_j(\theta) = 0.$$

La iteración por el método de Newton–Raphson queda

$$\theta^* = \theta - l'(\theta)/l''(\theta),$$

con

$$l''(\theta) = \sum_{j=1}^k n_j \frac{P_j(\theta)P_j''(\theta) - P_j'(\theta)^2}{P_j(\theta)^2},$$

donde θ^* es el nuevo valor del parámetro.

El principal problema con esta aproximación es que la segunda derivada generalmente es difícil de obtener. La modificación propuesta por el método de *scoring* es tratar a n_j como una variable aleatoria y reemplazarla por su esperanza en la expresión para la segunda derivada. Se supone que tiene una distribución binomial negativa con esperanza $nP_j(\theta)$. Ahora tenemos,

$$\begin{aligned} E[l''(\theta)] &= n \sum_{j=1}^k \frac{P_j(\theta)P_j''(\theta) - P_j'(\theta)^2}{P_j(\theta)} = n \sum_{j=1}^k P_j''(\theta) - n \sum_{j=1}^k \frac{P_j'(\theta)^2}{P_j(\theta)} \\ &= n \frac{d^2}{d\theta^2} \sum_{j=1}^k P_j(\theta) - n \sum_{j=1}^k \frac{P_j'(\theta)^2}{P_j(\theta)} \\ &= n \sum_{j=1}^k \frac{P_j'(\theta)^2}{P_j(\theta)}. \end{aligned}$$

Llevando esto a

$$\theta^* = \theta - l'(\theta)/l''(\theta),$$

se obtiene la fórmula de iteración por el método *scoring* que resulta

$$\theta^* = \theta + \frac{\sum_{j=1}^k n_j \frac{P_j'(\theta)}{P_j(\theta)}}{n \sum_{j=1}^k \frac{P_j'(\theta)^2}{P_j(\theta)}}.$$

Caso multivariado

El método se extiende fácilmente al caso multivariado, y las fórmulas que resultan son:

$$\begin{aligned} \theta^* &= \theta + [I(\theta)]^{-1} S(\theta), \\ S(\theta)_r &= \sum_{j=1}^k n_j \frac{\partial P_j(\theta)/\partial \theta_r}{P_j(\theta)}, \\ I(\theta)_{rs} &= n \sum_{j=1}^k \frac{[\partial P_j(\theta)/\partial \theta_r][\partial P_j(\theta)/\partial \theta_s]}{P_j(\theta)}, \end{aligned}$$

donde $S(\theta)_r$ indica el r -ésimo elemento del vector $S(\theta)$ e $I(\theta)_{rs}$ indica el (r, s) -ésimo elemento de la matriz $I(\theta)$.

Bibliografía

- [1] Arbous, A.G. y Sichel, H.S. (1954). New techniques for the analysis of absenteeism data. *Biometrika*, **41**, 1/2, 77-90.
- [2] Alanko, T. y Duffy, J.C (1996). Compound binomial distribution for modelling consumption data. *The Statistician*, **45**, 3, 269 – 286.
- [3] Balakrishnan, N. y Nevzorov, V. (2003). A primer on statistical distributions. John Wiley and Sons, New York.
- [4] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.
- [5] Berger, J. (1994). An overview of robust Bayesian analysis (with discussion). *Test*, **3**, 5–125.
- [6] Berger, J. y Berliner, L.M. (1986). Robust Bayes and empirical Bayes analysis with ε -contaminated priors. *Annals of Statistics* **14**, 461–486.
- [7] Betrò, B., Ruggeri, F. y Męczarski, M. (1994). Robust Bayesian analysis under generalized moments conditions. *Journal of Statistical Planning and Inference*, **41**, 257 – 266.
- [8] Bühlmann, H. (1967). Experience rating and credibility. *Astin Bulletin*, **4**, 3, 199–207.
- [9] Bühlmann, H. (1970). *Mathematical Methods in Risk Theory*. Springer.
- [10] Calderín-Ojeda, E., Gómez-Déniz, E. y Cabrera-Ortega, I. (2007). Bayesian local robustness under weighted squared-error loss function incorporating unimodality. *Statistics & Probability Letters*, **77**, 69–74.
- [11] Chhikara, R. y Folks, T. (1988). The inverse Gaussian distribution. Theory, methodology, and applications.
- [12] Denuit, M. (1997). A new distribution of Poisson-type for the numbers of clamis. *Astin Bulletin*, **27**, 2, 229 – 242.

- [13] Diaconis, P. y Freedman, D. (1986). On the consistency of Bayes estimates. *Annals of Statistics*, **14**, 1-67.
- [14] Eichenauer, J.; Lehn, J. y Rettig, S. (1988). A gamma-minimax result in credibility theory, *Insurance: Mathematics and Economics*, **7**, 1, pp. 49-57.
- [15] Frangos, N. y Karlis, D. (2004). Modelling losses using an exponential-inverse Gaussian distribution. *Insurance: Mathematics and Economics*, **35**, 53-67.
- [16] Gerber, H.U. (1979). An introduction to mathematical risk theory. Huebner Foundation Monograph 8.
- [17] Gómez-Déniz, E.; Hernández, A. y Vázquez, F. (1999). The Esscher premium principle in risk theory: a Bayesian sensitivity study. *Insurance: Mathematics and Economics*, **25**, 3, 387-395
- [18] Gómez-Déniz, E; Hernández, A. y Vázquez, F.J. (2000). Robust Bayesian Premium Principles in Actuarial Science. *Journal of the Royal Statistical Society (The Statistician, Series D)*. **49**, 2, 241-252.
- [19] Gómez-Déniz, E, Hernández, A., Pérez, J.M. y Vázquez-Polo, F., (2002). Measuring Sensitivity in a Bonus-Malus System. *Insurance: Mathematics and Economics* **31**, 105-113.
- [20] Gómez-Déniz, E y Vázquez, F. (2003). Robustness in Bayesian Model for Bonus-Malus Systems. In: *Intelligent and Other Computation Techniques in Insurance. Theory and Applications*. World Scientific.
- [21] Gómez-Déniz, E. Calderín-Ojeda, E. y Cabrera-Ortega, I. (2004a). Bayesian local robustness in L_p spaces. Applications in insurance premiums. Comunicación oral en el VIII congreso internacional de Insurance Mathematics and Economics. Roma (Italia).
- [22] Gómez-Déniz, E. Calderín-Ojeda, E. y Cabrera-Ortega, I. (2004b). Using L_p spaces to measure uncertainty in insurance premiums. Comunicación oral en la XVIII Reunión Asepelt-España. León (España).
- [23] Gómez-Déniz, E. Calderín-Ojeda, E. y Cabrera-Ortega, I. (2004c). Utilización de espacios normados para el análisis bayesiano local de primas de seguros. Comunicación oral en el congreso nacional SEIO. Cádiz (España).
- [24] Gómez-Déniz, E. Calderín-Ojeda, E. y Cabrera-Ortega, I. (2004d). Studying Bayesian local sensitivity of BMS using L_p spaces. Comunicación en el XVI congreso Compstat. Praga (República Checa).
- [25] Gómez-Déniz, E. y Vázquez, F.J. (2005a). Modelling uncertainty in insurance bonus-malus premiums by using a Bayesian robustness approach. *Journal of Applied Statistics*, **32**, 7, 771-784.

- [26] Gómez-Déniz, Bermúdez, L. y Morillo, I. (2005b). Computing bonus-malus premiums under partial prior information. *British Actuarial Journal*, **11**, II, 361-374.
- [27] Gómez-Déniz, E., Calderín-Ojeda, E. y Sarabia, J. M. (2005c) Using a new distribution of probability to model the number of claims in an automobile insurance portfolio. Comunicación oral. XIX Reunión ASEPELT-ESPAÑA. Badajoz (España)
- [28] Gómez-Déniz, E., Sarabia, J. M. y Calderín-Ojeda, E. (2005d) Univariate and Multivariate versions of the Negative Binomial-Inverse Gaussian distribution. Applications in Actuarial Statistics and other fields. Comunicación oral. 9º Congreso Internacional de Insurance Mathematics and Economics. Quebec (Canadá)
- [29] Gómez-Déniz, E. Calderín-Ojeda, E. y Cabrera-Ortega, I. (2005e). Un sencillo mecanismo para decidir si una prima bonus-malus es o no robusta. Comunicación oral. 1ª Reunión de investigación en Seguros y Gestión de riesgos. Barcelona (España).
- [30] Gómez, Pérez y J.M. y Vázquez-Polo, F. (2006a). A note on computing bonus-malus insurance premiums using a hierarchical Bayesian framework. *Test*, **15**, 2, 345-359.
- [31] Gómez-Déniz, E.; Calderín-Ojeda, E. y Cabrera-Ortega, I. (2006b). A simple method to study sensitivity of BMP's. *Communications in Statistics. Theory and Methods*, **35**, 583-591.
- [32] Gómez-Déniz, E, Sarabia, J.M. y Calderín-Ojeda, E. (2007). Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. *Insurance: Mathematics and Economics* (en prensa).
- [33] Goutis, C. (1994). Ranges of posterior measures for some classes of priors with specified moments. *Int. Statist. Review*, **62**, 245-357.
- [34] Griffel, D.H. (1984). *Applied Functional Analysis*. John Wiley & Sons.
- [35] Gustafson, P. (1996a). Local sensitivity of posterior expectations. *The Annals of Statistics*. **24**, 1, 174-195.
- [36] Gustafson, P. (1996b). Local sensitivity of inferences to Prior Marginals. *Journal of the American Statistical Association*. **91**, 434, 774-781.
- [37] Gustafson, P. y Wasserman, L. (1995). Local sensitivity diagnostics for Bayesian inference. *The Annals of Statistics*. **23**, 6, 2153-2167.

- [38] Heilmann, W. (1989). Decision theoretic foundations of credibility theory. *Insurance: Mathematics and Economics*, **6**, 145-149.
- [39] Hougaard, P., Lee, M.-T., y Whitmore, G.A. (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson process. *Biometrics*, **53**, 1225-1238.
- [40] Johnson, N.L., Kotz, S. y Balakrishnan, N. (1997). *Discrete Multivariate Distribution*. John Wiley and Sons, New York.
- [41] Kokonendji, C. y Khoudar, M. (2004). On strict arcsine distribution. *Communications in Statistics. Theory and Methods*, **33**, 5, 993-1006.
- [42] Klugman, S. (1987). Credibility for classification ratemaking via the hierarchical normal linear model. *Proceedings of the Casualty Actuarial Society*, **74**, 272-321.
- [43] Klugman, S.; Panjer, H. y Willmot, G. (1998). *Loss Models. From Data to Decisions*. John Wiley and Sons, New York.
- [44] Lemaire, J. (1979). How to define a bonus-malus system with an exponential utility function. *Astin Bulletin*, **10**, 274-282.
- [45] Lemaire, J. (1985). *Automobile Insurance (Actuarial Models)*. Kluwer-Nijhoff Publishing. Boston/Dordrecht/Lancaster.
- [46] Lemaire, J. (1995). *Bonus-Malus systems in automobile insurance*. Kluwer Academic Publishers, London.
- [47] Makov, U.E. (1994). Loss robustness via Fisher-weighted squared-error loss function. *Insurance: Mathematics and Economics*, **16**, 1-6.
- [48] Martín, J; Ríos, D. y Ruggeri, F. (2003). Joint sensitivity in Bayesian Decision Theory. *Test*, **12**, 1, 173-194.
- [49] Meng, S.; Wei, Y. y Whitmore, G.A. (1999). Accounting for individual overdispersion in a Bonus-Malus system. *Astin Bulletin*, **29**, 2, 327-337.
- [50] Milne, R.D. (1980). *Applied Functional Analysis. An Introductory Treatment*. Courier Dover Publications.
- [51] Moreno, E. y Cano, J. (1991). Robust Bayesian analysis for ϵ -contaminations partially known. *Journal of the Royal Statistical Society. Series B*, **53**, 143-155.
- [52] Moreno, E.; Bertolino, F. y Racugno, W. (2003). Bayesian inference under partial prior information. *Scandinavian Journal of Statistics*, **30**, 565-580.

- [53] Panjer, H.J. (1981). Recursive evaluation of a family of compound distributions. *Astin Bulletin*, **12**, 1, 22–26.
- [54] Peña, D. y Zamar, R. (1997). A simple diagnostic tool for local prior sensitivity. *Statistics & Probability Letters*, **36**, 205–212.
- [55] Pérez, J.; Sarabia, J.; Gómez-Déniz, E. y Vázquez, F.J. (2006). Applying a Bayesian hierarchical model in actuarial science: inference and ratemaking. En *Distribution Models Theory*. World Scientific, pp. 233–241.
- [56] Ríos, S.; Martín, J.; Ríos, D. y Ruggeri, F. (1999). Bayesian Forecasting for Accident Proneness Evaluation. *Scandinavian Actuarial Journal*, **2**, 134–156.
- [57] Ríos, D. y Ruggeri, F. (2000). *Robust Bayesian Analysis*. Lectures Notes in Statistics. Springer, New York.
- [58] Ruggeri, F. y Wasserman, L. (1993). Infinitesimal sensitivity of posterior distributions. *Canadian Journal of Statistics*, **21**, 195–203.
- [59] Sarabia, J.M.; Gómez-Déniz, E. y Vázquez-Polo, F.J., (2004). On the use of conditional specification models in claim count distributions: an application to bonus-malus systems. *Astin Bulletin*, **34**, 1, 85–98.
- [60] Sarabia, J.M.; Gómez-Déniz, E. y Vázquez-Polo, F.J., (2006). *Estadística Actuarial. Teoría y Aplicaciones*. Pearson. Prentice Hall.
- [61] Simon, L. (1961). Fitting Negative Binomial Distribution by the Method of Maximum Likelihood. *Proceedings of the Casualty Actuarial Society*, **XLVIII**, 45–53 (with discussion).
- [62] Sivaganesan, S. (2000). Global and Local Robustness Approaches: Uses and Limitations. In *Robust Bayesian Analysis* (D. Ríos and F. Ruggeri, eds.), pages 89–108. Springer-Verlag, New York.
- [63] Sivaganesan, S. y Berger, O. (1989) Ranges of posterior measures for priors with unimodal contaminations. *The Annals of Statistics*, **17**, 2 pp. 868–889.
- [64] Stein, G.; Zucchini, W. y Juritz, J. (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association*, **82**, 339, 938–944.
- [65] Tremblay, L. (1992). Using the Poisson inverse gaussian in bonus-malus systems. *Astin Bulletin*, **22**, 1, 97–106.
- [66] Tweedie, M.C.K. (1957). Statistical Properties of Inverse Gaussian Distributions I. *Annals of Mathematical Statistics*, **28**, 2, 362–377.

- [67] Willmot, G.E. (1987). The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuarial Journal*, 113-127.
- [68] Willmot, G.E. (1988). Parameter orthogonality for a family of discrete distribution. *Journal of the American Statistical Association*, 83, 517-521.