A Bayesian model for the claim number with incomplete count data

F.J. Vázquez–Polo and M. Martel

Dpt. of Quantitative Methods. University of Las Palmas de G.C., Spain
francisco.vazquezpolo@ulpgc.es

II Workshop on Pensions and Insurance
U. Barcelona, July 14–15, 2016

Motivación

- IBNR (Incurred But Not Reported).
  1. La "llegada" de reclamaciones está sujeta a retraso.
  2. Las compañías necesitan tener previsiones del número de reclamaciones (anual) para tener una imagen lo más precisa posible de su situación financiera (provisión de reservas).

  Métodos de resolución en la literatura: mínimos cuadrados, regresión, credibiilidad,... (Goovaerts et al., 1990).
  Verrall (1991) modelos loglineales y "chain ladder method".
  De Alba (1996), De Alba & Mendoza (1996), Mindelhall (2006) usan técnicas bayesianas.
- La distribución MBT (Modified Borel–Tanner).
  Gómez–Déniz, E., Vázquez–Polo, F.J. & García, V.J. (2015)The Modified Borel–Tanner (MBT) distribution and its regression model associated. *Statistical Journal–REVSTAT* (en prensa).

Planteamiento estadístico del problema

- Supongamos x_t (número de reclamaciones) un proceso homogéno $P(x_t|\lambda t)$ donde t denota un intervalo de tiempo y λ es desconocido con x_t no observable (al menos completamente).

  H1: Las x_t's son reportadas indepedientemente unas de otras
- Sin embargo, si disponemos de información muestral (observable) y_t que representa una parte (proporción) de x_t.

En definitiva:

x_t: "número de reclamaciones durante un tiempo t"

y_t: "número actual de reclamaciones presentadas"

- Las variables x_t siguen una distribución **MBT** con parámetro $(\lambda \cdot t)$:

$$f_1(x_t|\lambda t) = \frac{\Gamma(2x_t+1)}{\Gamma(x_t+2)\Gamma(x_t+1)} \frac{(\lambda t)^{x_t+1}}{(1+\lambda t)^{2x_t+1}}, \quad x_t = 0, 1, \dots \tag{1}$$

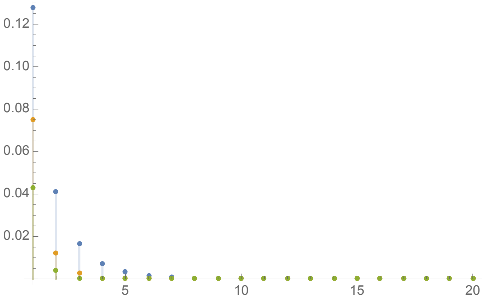

Figure: Distribuciones $MBT(\lambda),\ \lambda = 4, 10, 20.$

El modelo adoptado

- Las variables y_t siguen una distribución **Binomial** con parámetro (x_t, θ):

$$f_2(x_t|\lambda t) = \binom{x_t}{y_t} \cdot \theta^{y_t}(1-\theta)^{x_t - y_t}, \quad y_t = 0, \ldots, x_t. \qquad (2)$$

O1: Necesitamos una verosimilitud que relacione las observaciones y_t con los parámetros del problema: λ y θ.

El modelo

Para simplificar la notación consideramos $\lambda t \longleftrightarrow \lambda$.

$$f(y_t|\lambda,\theta) = \sum_{x_t=y_t}^{\infty} f_2(y_t|x_t,\lambda,\theta)\cdot f_1(x_t|\lambda,\theta)$$

H2: x_t es independiente de θ dado λ e y_t es independiente de λ dado (x_t,θ).

En tal caso,

$$f(y_t|\lambda,\theta) = \sum_{x_t=y_t}^{\infty} f_2(y_t|x_t,\theta)\cdot f_1(x_t|\lambda) =$$

$$= \frac{\theta^{y_t}\cdot\lambda^{y_t+1}}{y_t!(1+\lambda)^{y_t}}\cdot\sum_{x_t=y_t}^{\infty}\frac{(1-\theta)^{x_t-y_t}}{(x_t-y_t)!}\cdot\frac{\Gamma(2y_t+1)}{\Gamma(y_t+2)}\cdot\frac{\lambda^{x_t-y_t}}{(1+\lambda)^{2x_y-y_t+1}}$$

Verosimilitud para (λ, θ) observado y_t:

$$f(y_t|\lambda,\theta) = \frac{\theta^{y_t} \cdot \lambda^{y_t+1}}{(1+\lambda)^{2y_t+1}} \cdot \frac{\Gamma(2y_t+1)}{\Gamma(y_t)\Gamma(y_t+2)} \times {}_2F_1\left(y_t+\frac{1}{2}, y_t+1, y_t+2; 4\frac{\lambda}{(1+\lambda)^2}(1-\theta)\right),$$

$$y_t = x_t, \ldots (3)$$

OJO: Esta verosimilitud contiene productos $\lambda\theta$ lo cual indica que no distingue los pares (λ, θ).

El modelo

Ahora dada una a priori conjunta $\pi(\lambda, \theta)$, podemos encontrar las cantidades de interés del problema. Por ejemplo:

- **Media (a posteriori) de λ dado y_t:**

$$\mathbb{E}(\lambda|y_t) \propto \int_0^\infty \left(\int_0^1 f(y_t|\lambda, \theta)\pi(\lambda, \theta)\ d\theta \right)\ d\lambda$$

con la constante de proporcionalidad

$$\int_0^\infty \int_0^1 f(y_t|\lambda, \theta)\pi(\lambda, \theta)\ d\theta d\lambda.$$

- **Media (a posteriori) de x_t dado y_t:**

$$\mathbb{E}(x_t|y_t) \propto \int_0^\infty \left(\int_0^1 \left(\sum_{x_t=y_t}^\infty x_t f_2(y_t|x_t,\theta) f_1(x_t|\lambda) \right) \pi(\lambda,\theta)\; d\theta \right)\; d\lambda$$

con la constante de proporcionalidad

$$\int_0^\infty \int_0^1 f(y_t|\lambda,\theta)\pi(\lambda,\theta)\; d\theta d\lambda.$$

La densidad a priori $\pi(\lambda, \theta)$

- Independencia entre los parámetros hace al modelo "tratable".
- De H2, como en Moreno & Girón (1998), tenemos:
  - La independencia entre λ y θ implica, independencia entre x_t y θ.
  - El recíproco es cierto cuando la familia $\pi(\lambda|x_t)$ es completa.
- De la verosimilitud (3) no podemos deducir una a priori de referencia pues la verosimilitud no distingue (λ, θ).
- Cuando la información a priori sobre θ y λ puede ser especificada independientemente en término de familias conjugadas marginalmente, el análisis es analíticamente tratable.

La densidad a priori $\pi(\lambda, \theta)$

- La familia conjugada para θ es Beta(α_0, β_0) :

$$\begin{aligned} \pi(\theta|\alpha_0, \beta_0) &= \frac{1}{B(\alpha_0, \beta_0)} \cdot \theta^{\alpha_0 - 1} \cdot (1-\theta)^{\beta_0 - 1} \\ &\propto \theta^{\alpha_0 - 1} \cdot (1-\theta)^{\beta_0 - 1}, \qquad 0 < \theta < 1, \end{aligned} \tag{4}$$

con $B(\alpha_0, \beta_0) = \dfrac{\Gamma(\alpha_0)\Gamma(\beta_0)}{\Gamma(\alpha_0 + \beta_0)}$.

$$\mathbb{E}(\theta) = \frac{\alpha_0}{\alpha_0 + \beta_0}.$$

$$\text{Moda}(\theta) = \frac{\alpha_0}{\alpha_0 + \beta_0 - 2}.$$

- Para λ tenemos que su modelo poblacional es

$$\ell(\boldsymbol{x}|\lambda) \propto \frac{\lambda^{n\bar{x}+n}}{(1+\lambda)^{2n\bar{x}+n}} = \left(\frac{\lambda}{1+\lambda}\right)^n \cdot \exp\left\{n\bar{x}\log\frac{\lambda}{(1+\lambda)^2}\right\}$$

Tenemos por tanto un estadístico suficiente $\bar{x}$ y una descomposición del tipo $h(\lambda)$ y $\psi(\lambda)$ con:

$$h(\lambda) = \frac{\lambda}{1+\lambda}, \qquad \psi(\lambda) = \log\frac{\lambda}{(1+\lambda)^2}$$

- La familia conjugada para λ es por tanto

$$\pi(\lambda) \propto h(\lambda)^{p_0} \exp\{q_0 \psi(\lambda)\} :$$

$$\begin{aligned} \pi(\lambda|p_0, q_0) &= \frac{1}{B(p_0, q_0)} \cdot \frac{\lambda^{p_0-1}}{(1+\lambda)^{p_0+q_0}} \\ &\propto \frac{\lambda^{p_0-1}}{(1+\lambda)^{p_0+q_0}} \qquad \lambda > 0, \end{aligned} \tag{5}$$

con $q_0 > 1$.

Beta Generalizada de segundo tipo (GBII)

BetaPrima, BP(p_0, q_0) :

$$\mathbb{E}(\lambda) = \frac{p_0}{q_0 - 1}, \qquad q_0 > 1.$$

$$\text{Moda}(\lambda) = \frac{p_0 - 1}{q_0 + 1}, \qquad p_0 \geq 1.$$

La densidad a priori $\pi(\lambda, \theta)$

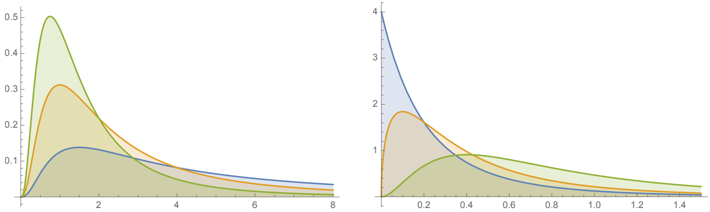

Figure: Distribuciones $BP(4, q),\ q = 1, 2, 3$ (panel izquierdo) y $BP(p, 4),\ q = 1, 1.5, 3$ (panel derecho)

La a posteriori para λ observadas $\boldsymbol{x}$ es:

$$\begin{aligned} \pi(\lambda|\boldsymbol{x}) &\propto f(\boldsymbol{x}|\lambda)\pi(\lambda) \\ &\propto \frac{\lambda^{n\bar{x}+n}}{(1+\lambda)^{2n\bar{x}+n}} \cdot \frac{\lambda^{p_0-1}}{(1+\lambda)^{p_0+q_0}} \\ &\propto \frac{\lambda^{n\bar{x}+n+p_0-1}}{(1+\lambda)^{2n\bar{x}+n+p_0+q_0}} \sim \mathrm{BP}(n\bar{x}+n+p_0, n\bar{x}+q_0) \end{aligned}$$

Por tanto, $\boxed{\pi(\lambda,\theta) \propto \theta^{\alpha_0-1} \cdot (1-\theta)^{\beta_0-1} \cdot \frac{\lambda^{p_0-1}}{(1+\lambda)^{p_0+q_0}}}$

La densidad a posteriori $\pi(\lambda, \theta|y_t)$

La a posteriori para λ es:

$$\begin{aligned}\pi(\lambda, \theta|y_t) &\propto f(y_t|\lambda, \theta)\pi(\lambda, \theta)\\ &\propto \theta^{\alpha_0 - 1} \cdot (1-\theta)^{\beta_0 - 1} \cdot \frac{\lambda^{p_0 - 1}}{(1+\lambda)^{p_0+q_0}} \cdot \frac{\theta^{y_t}\lambda^{y_t+1}}{(1+\lambda)^{2y_t+1}} \times\\ &\quad {}_2F_1\left(y_t + \frac{1}{2}, y_t+1, y_t+2; 4\frac{\lambda}{(1+\lambda)^2}(1-\theta)\right), \text{(6)}\end{aligned}$$

(1) La distribución posteriori para x_t es:

$$\pi(x_t|y_t) \propto \pi(x_t) \cdot \pi(y_t|x_t).$$

$$\begin{aligned}
\pi(y_t|x_t) &= \int \pi(y_t|x_t, \theta)\pi(\theta||x_t)\, d\theta \\
&= \binom{x_t}{y_t} \cdot \frac{B(\alpha_0 + y_t, \beta_0 + x_t - y_t)}{B(\alpha_0, \beta_0)} \\
&= \binom{x_t}{y_t} \cdot \frac{\Gamma(\alpha_0 + y_t)\Gamma(\beta_0 + x_t - y_t)}{\Gamma(\alpha_0 + \beta_0 + x_t)} \cdot \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0) + \Gamma(\beta_0)}.
\end{aligned}$$

(2) La distribución marginal para x_t es:

$$\pi(x_t) = \int_0^\infty f_1(x_t|\lambda) \cdot \pi(\lambda)\, d\lambda.$$

$$\begin{aligned} \pi(x_t) &= \int_0^\infty f_1(x_t|\lambda) \cdot \pi(\lambda)\, d\lambda \\ &= \frac{\Gamma(2x_t+1)}{\Gamma(x_t+2)\Gamma(x_t+1)} \cdot \frac{B(x_t+p_0+1, x_t+q_0)}{B(p_0, q_0)} \end{aligned}$$

Multiplicando ambas expresiones anteriores se obtiene:

$$\pi(x_t|y_t) \quad \propto \quad \binom{x_t}{y_t} \frac{B(\alpha_0 + y_t, \beta_0 + x_t - y_t)}{B(\alpha_0, \beta_0)} \frac{\Gamma(2x_t + 1)}{\Gamma(x_t + 2)\Gamma(x_t + 1)} \times \frac{B(x_t + p_0 + 1, x_t + q_0)}{B(p_0, q_0)} \tag{7}$$

donde la constante de proporcionalidad se deduce de

$$\sum_{x_t = y_t}^{\infty} \pi(x_t|y_t) = 1.$$

Un ejemplo práctico

Datos tomados de Doray (1996), representando el número de reclamaciones reportadas hasta Septiembre 1987.

Number of accidents reported by September 30, 1987

	...	1985	1986	1987	1980–86
I	...	170	178	202	185.86
II	...	177	130	156	165.00
III	...	120	154	138	154.71
IV	...	102	134	153	138.29
V	...	156	213	198	190.57
VI	...	195	201	178	198.43
VII	...	186	201	127	202.29
VIII	...	184	203	142	189.57
IX	...	167	219	93	197.57
X	...	167	205	–	209.00
XI	...	167	193	–	195.29
XII	...	260	162	–	197.57

Un ejemplo práctico

Número medio y total de reclamaciones al año

1980	1981	1982	1983	1984	1985	1986	1987
198.8	223.1	168.2	182.0	171.7	170.9	182.7	–
2386	2677	2018	2184	2060	2051	2193	$y_t = 1387$

Los cálculos han sido realizados con `Mathematica v.10.4`
(códigos disponibles)

Un ejemplo práctico

Distribuciones a priori utilizadas:

- Informativa: Beta$(235.6, 87.5)$
  Construida con los datos históricos
  $\theta_{1980} = 0.685, \theta_{1981} = 0.759, \ldots, \theta_{1986} = 0.745$
- No informativa: BP$(p_0 = 1, q_0 = 1)$ que permite el análisis "conjugado".

predictiva.pdf

Figure: Distribución a posteriori $\pi(x_t|y_t)$ para los datos de Doray (1996).

Valor estimado global: $y_{1987}^{\text{total}} = 1907.$
Intervalo bayesiano al 95% : $[1780, 2055]$

Conclusiones

- El problema IBNR puede resolverse de forma sencilla mediante un modelo bayesiano con datos incompletos.
- La distribución MBT es adecuada para dicho análisis.
- Las a priori utilzadas muestran buenas propiedades estadísticas y tratables analíticamente.
- Los datos faltantes no son más que las estimaciones obtenidas de la distribución predictiva de x_t.

Muchas gracias por la atención!!!