ACT-11302: Cálculo Actuarial III **ITAM**

Frecuencia de siniestros (Distribuciones)

Prof: Juan Carlos Martínez Ovando *24 de septiembre de 2015*

En estadística actuarial se modela la pérdica inducida por reclamaciones de pólizas de seguros. Para este fecto, tanto el número de siniestros como la severidad (monto) de los mismos son relevantes. El número de siniestros (frecuencia) mide el número de reclamaciones para un bloque de pólizas de seguros en un periodo de tiempo finito (mensual o anual).

Se supone que la frecuencia de siniestros, X, es una variable aleatoria discreta con soporte en los enteros positivos, $\mathscr{X} = \{0, 1, 2, \ldots\}$. En esta sesión revisaremos algunas distriuciones paramétricas que son comúnmente empleadas para modelar la frecuencia de siniestros.

Note que aunque el bloque de pólizas de seguros está formado por un número finito, la modelación de frecuencias supone, en ocasiones, un soporte numerable. Esto no es un problema en la práctica, pues muchas de estas distribuciones en verdad son una aproximación razonable para una distribución de frecuencia de siniestros con un número finito de pólizas.

Ls distribuciones que revisamos en esta sesión son:

- Binomial (Bin).
- Binomial negativa (BinN).
- Poisson (Po).
- Geométrica (Geo).

1. Distribución binomial

La variable aleatoria X es modelada con la distribución binomial, $\mathrm{Bin}(x;n,\theta)$, donde n el número de pólizas en el bloque de seguros, y $0<\theta<1$ es la probabilidad de ocurrencia de siniestro de cada póliza individual, si

$$p_X(x) = Pr(X=x) \propto \theta^x(1-\theta)^{n-x}, \tag{1}$$

para $x \in \{0,1,2,\ldots,n\}$. Aquí se supone que n es finito y conocido.

Note que los supuestos fundamentales para el uso de la distribución binomial para la modelación de frecuencias de siniestros son: i) la ocurrencia de siniestros entre pólizas es la misma (i.e. los siniestros comparten condiciones homogéneas de exposición al riesgo de siniestro), y ii) la ocurrencia de siniestros es independiente entre pólizas.

Esta distribución tiene las siguientes propiedades:

$$\begin{aligned}\mathbb{E}(X) &= n\theta.\\ \mathbb{V}(X) &= n\theta(1-\theta).\\ M_X(t) &= \left(\theta e^t + (1-\theta)\right)^n.\end{aligned}$$

La distribución binomial es simétrica si $\theta = 1/2$. En el caso $\theta < 1/2$, la distribución es sesgada a la derecha (positivamente sesgada); mientras que en el caso $\theta > 1/2$, la distribución es sesgada a la izquierda (negativamente sesgada).

El estimador máximo verosimil para θ, bajo el supuesto de independencia estocástica entre los siniestros de las pólizas, es $\hat{\theta} = 1/n\sum_{i=1}^{n}\mathbb{I}(\text{siniestro}_i)$. La distribución inicial conjugada para θ, bajo el enfoque bayesiano de inferencia, es una distribución beta, $\mathrm{Be}(\theta|\alpha_0,\beta_0)$, con $\alpha_0,\beta_0>0$. Los valores de los hiperparámetros α_0 y β_0 son previamente especificados por el modelador, empleando información adicional relevante (e.g. datos de frecuencia de siniestros de un bloque de pólizas para un periodo de tiempo distinto y coberturas semejantes). La distribución final para θ será así beta, pero con parámetros $\alpha_1 = \alpha_0 + \sum_{i=1}^{n} \mathbb{I}(\text{siniestro}_i)$ y $\beta_1 = \beta_0 + \sum_{i=1}^{n} \mathbb{I}(\text{no siniestro}_i)$.

2. Distribución binomial negativa

La distribución binomial negativa, $\text{BinN}(x; r, \theta)$ modela el número de pólizas siniestradas, x, antes de observar r pólizas no siniestradas, donde la probabilidad de siniestro es $0 < \theta < 1$. La distribución está dada por:

$$p_X(x) = Pr(X = x) \propto \theta^x (1-\theta)^r, \tag{2}$$

para $x \in \mathscr{X}$.

Al igual que en el modelo binomial, se supone que las pólias comparten condiciones homogéneas de exposición al riesgo de ser siniestradas.

Esta distribución tiene las siguientes propiedades:

$$\begin{aligned} \mathbb{E}(X) &= r\frac{(1-\theta)}{\theta}. \\ \mathbb{V}(X) &= r\frac{(1-\theta)}{\theta^2}. \\ M_X(t) &= \left(\frac{\theta}{1-(1-\theta)e^t}\right)^r. \end{aligned}$$

Usualmente, el parámetro r se supone fijo y conocido. En la práctica, r es desconocido, por tanto es estimable. Así, el estimador máximo verosímil para θ sería, $\hat{\theta} = \frac{\sum_i \mathbb{I}(\text{siniestro}_i)}{N_r + \sum_i \mathbb{I}(\text{siniestro}_i)}$, donde N_r es el estimador de r el cuál se obtiene empleando métodos numéricos usando el algoritmo de Brent[1]

Bajo el enfoque bayesiano de inferencia, la distribución inicial conjugada para θ es $\text{Be}(\theta|\alpha_0, \beta_0)$.

[1] Brent, R.P. (2002) *Algorithms for minimization without derivatives.* Dover Publications.

Condicional en r, la distribución final para θ es $\text{Be}(x|\alpha_1, \beta_1)$, con $\alpha_1 = \alpha_0 + \sum_i \mathbb{I}(\text{siniestro}_i)$ y $\beta_1 = \beta_0 + r$. Cuando se extiende la incertidumbre a r tambi se requiere definir una prior para r, la cual puede ser $\text{Po}(r|\lambda_0)$. La distribución final conjunta para (r, θ), en este caso, se calcula empleando métodos numéricos también (e.g. el muestreador de Gibbs[2]).

3. Distribución geométrica

La distribución geométrica, $\text{Geo}(x;\theta)$ modela el número de pólizas siniestradas, x, antes de observar 1 póliza no siniestrada, donde la probabilidad de siniestro es $0 < \theta < 1$. La distribución está dada por:

$$p_X(x) = Pr(X = x) \propto \theta^x (1 - \theta), \tag{3}$$

para $x \in \mathscr{X}$. Esta distribución es un caso particular de la distribución binomial negativa, con $r = 1$. El análisis y estimaciecuentista y bayesiano de esta distribución es notoriamente más simple que en el caso de la distribución binomial negativa. Sin embargo, su implementación y uso en la práctica es menos robusto que el de la distribución binominal negativa.

Los supuestos que operan en este caso son similares a los de la distribución binomial negativa.

4. Distribución Poisson

La distribución Poisson para la frecuencia de siniestros, $\text{Po}(x;\lambda)$, siendo $\lambda > 0$ la tasa de siniestros, es quizás la distribución más empleada en la práctica. La distribución de probabilidad está dada

[2] Robert, C. y Casella, G. (1997) *Introduction to MCMC Methods*. Springer.

por:

$$p_X(x) = Pr(X = x) \propto \frac{\lambda^x}{x!}, \tag{4}$$

para $x \in \mathscr{X}$.

Esta distribución tiene las siguientes propiedades:

$$\begin{aligned}
\mathbb{E}(X) &= \lambda. \\
\mathbb{V}(X) &= \lambda. \\
M_X(t) &= \exp\left\{\lambda(e^t - 1)\right\}.
\end{aligned}$$

El estimador de máxima verosimilitud de λ es $\hat{\lambda} = 1/n \sum_i \mathbb{I}(\text{siniestro}_i)$. Bajo el enfoque bayesiano de inferencia, la distribución inicial conjugada para λ es $\text{Ga}(\lambda|\alpha_0, \beta_0)$. Así, la distribución final para λ es $\text{Ga}(\lambda|\alpha_1, \beta_1)$, donde $\alpha_1 = \alpha_0 + \sum_i \mathbb{I}(\text{siniestro}_i)$ y $\beta_1 = \beta_0 + n$.