Distribución de Poisson

8 de octubre de 2018

1. Estimador Basado en Momentos

Dada una muestra aleatoria $X_1, ..., X_n \overset{iid}{\sim} \mathcal{P}(\lambda)$, por la Ley Fuerte de los Grandes Números sabemos que:

$$\overline{X} \xrightarrow[n\to\infty]{cs} \mathbb{E}_\lambda[X_1] = \lambda$$

Para obtener el estimador basado en el primer momento, igualamos:

$$\overline{X} = \hat{\lambda}$$

2. Estimador de Máxima Verosimilitud

2.1. Con verosimilitud

$$X \sim \mathrm{P}(\lambda) \Rightarrow f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$\boldsymbol{X} = (X_i)_{i=1}^N \,/\forall i, j : \mathrm{iid}\,(X_i; X_j)$

$$p_\lambda(\boldsymbol{x}) = \prod_{i=1}^N \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$p_\lambda(\boldsymbol{x}) = \lambda^{\sum_{i=1}^N x_i} e^{-\sum_{i=1}^N \lambda} \prod_{i=1}^N \frac{1}{x_i!}$$

$$p_\lambda(\boldsymbol{x}) = \lambda^{N\overline{X}} e^{-N\lambda} \cdot \left(\prod_{i=1}^N \frac{1}{x_i!}\right)$$

$\prod_{i=1}^N \frac{1}{x_i!}$ es una constante que no introduce dificultad alguna, así que por un rato la vamos a llamar arbitrariamente C.

$$\frac{\partial}{\partial\lambda} p_\lambda(\boldsymbol{x}) = C \cdot \left(N\overline{X} \cdot \lambda^{N\overline{X}-1} e^{-N\lambda} - \lambda^{N\overline{X}} \cdot N e^{-N\lambda}\right)$$

- $\hat{\lambda} = \lambda \in \mathbb{R} / \frac{\partial}{\partial\lambda} p_\lambda(\boldsymbol{x}) = 0$

$A \cdot B = 0 \Rightarrow A = 0 \vee B = 0$
$\forall x_i \in \mathbb{N} : \frac{1}{x_i!} \neq 0 \Rightarrow C \neq 0$

$$N\overline{X} \cdot \lambda^{N\overline{X}-1} e^{-N\lambda} - \lambda^{N\overline{X}} \cdot N e^{-N\lambda} = 0$$

$$\hat{\lambda}^{N\overline{X}-1} N \left(\overline{X} e^{N\hat{\lambda}} - \hat{\lambda} e^{N\hat{\lambda}} \right) = 0$$

$\hat{\lambda}^{N\overline{X}-1} N \neq 0$

$$\overline{X} e^{N\hat{\lambda}} - \lambda e^{N\hat{\lambda}} = 0$$

-
$$\overline{X} = \hat{\lambda}$$

2.2. Con Logverosimilitud

Primero obtenemos la verosimilitud del parámetro, que no es más que la probabilidad conjunta:

$$\mathcal{L}(\lambda) = p_\lambda(\boldsymbol{x}) = \lambda^{n\overline{X}} e^{-n\lambda} \cdot \left(\prod_{i=1}^{n} \frac{1}{x_i!} \right)$$

Queremos maximizar la verosimilitud. Aplicamos logaritmo para luego derivar más fácilmente:

$$\ln \mathcal{L}(\lambda) = \ln [p_\lambda(\boldsymbol{x})] = n\overline{X} \cdot \ln(\lambda) - n\lambda + \ln \left(\prod_{i=1}^{n} \frac{1}{x_i!} \right)$$

Derivamos con respecto a λ e igualamos a cero para buscar el valor del parámetro que maximiza esa función:

$$\frac{\partial}{\partial\lambda} \ln [p_\lambda(\boldsymbol{x})] = \frac{n\overline{X}}{\lambda} - n + 0 = \frac{n\overline{X}}{\lambda} - n = n \left(\frac{\overline{X}}{\lambda} - 1 \right)$$

- $\hat{\lambda} = \lambda \in \mathbb{R} / \frac{\partial}{\partial\lambda} \ln [p_\lambda(\boldsymbol{x})] = 0$

$$n \left(\frac{\overline{X}}{\hat{\lambda}} - 1 \right) = 0$$

$$\frac{\overline{X}}{\hat{\lambda}} - 1 = 0$$

$$\frac{\overline{X}}{\hat{\lambda}} = 1$$

- $$\hat{\lambda} = \overline{X}$$

Comprobamos que el estimador de máxima verosimilitud para λ coincide con el del método de los momentos basado en el primer momento.

3. Propiedades del estimador $\hat{\lambda} = \overline{X}$

3.1. El estadístico T es suficiente

Dada la función de la muestra $T = r(\underline{\mathrm{X}}) = \sum_{i=1}^{n} X_i$, vamos a ver que se trata de un estadístico suficiente para λ. Podemos factorizar la función de probabilidad conjunta del siguiente modo:

$$p(\underline{x}, \lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} = e^{-n\lambda} e^{\ln(\lambda) \sum_{i=1}^{n} x_i} \frac{1}{\prod_{i=1}^{n} x_i!}$$

donde $A(\lambda) = e^{-n\lambda}$, $r_1(\underline{x}) = \sum_{i=1}^{n} x_i$, $c_1(\lambda) = \ln(\lambda)$, $h(\underline{x}) = \frac{1}{\prod_{i=1}^{n} x_i!}$. Comprobamos que la distribución Poisson es una familia exponencial a un parámetro. Por ende, dada una muestra aleatoria $X_1, X_2, ..., X_n \overset{iid}{\sim} \mathcal{P}(\lambda)$, el estadístico $T = \sum_{i=1}^{n} X_i$es suficiente para el parámetro λ.

3.2. El estadístico T es completo

Poisson es una familia exponencial a un parámetro. Sea:

$$\Lambda = \{\ln(\lambda) : \lambda \in \mathbb{R}^+\}$$

Observamos que si $\lambda \in (0, 1)$, entonces $\ln(\lambda) \in (-\infty, 0)$, mientras que si $\lambda \in (1, +\infty)$, entonces $\ln(\lambda) \in (0, +\infty)$, de modo que:

$$\Lambda = \mathbb{R}$$

Como Λ contiene una bola en $\mathbb{R}$, el estadístico $T = \sum_{i=1}^{n} X_i$ es completo.

3.3. El estimador $\overline{X}$ es insesgado

Sea $\delta_n(T) = \frac{T}{n} = \overline{X}$ un estimador de λ basado en el estadístico suficiente y completo $T = \sum_{i=1}^{n} X_i$. Veamos que la esperanza del estimador equivale al parámetro a estimar:

$$\mathbb{E}_\lambda[\overline{X}] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_\lambda[X_i] = \frac{1}{n} \sum_{i=1}^{n} \lambda = \frac{n}{n} \lambda = \lambda$$

$$\mathbb{E}_\lambda[\overline{X}] = \lambda$$

Comprombamos que el estimador $\overline{X}$ es insesgado.

3.4. El estimador $\overline{X}$ es IMVU

Vimos que el estadístico T es suficiente y completo y que el estimador $\overline{X}$ basado en T es insesgado. Por ende, el estimador $\overline{X}$ es IMVU.

3.5. El estimador $\overline{X}$ es fuerte y débilmente consistente

Dado que la muestra construida con eventos que siguen una distribución Poisson cumple las hipótesis de tener elementos iid que siguen una distribución tal que la esperanza de su módulo es finita, se sigue por la Ley Fuerte de los Grandes Números que:

$$\overline{X} \xrightarrow[n\to\infty]{cs} \mathbb{E}[X_1] = \lambda$$

De la afirmación anterior se sigue inmediatamente la consistencia fuerte del estimador $\overline{X}$. Como la convergencia casi segura implica convergencia en probabilidad, el estimador $\overline{X}$ también es débilmente consistente.

3.6. Error Cuadrático Medio de $\overline{X}$

Como las variables X_i son iid, tenemos que la varianza del estimador es:

$$\mathrm{Var}_\lambda\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} X_i = \frac{1}{n^2} n\cdot\lambda = \frac{\lambda}{n}$$

$$\mathrm{Var}_\lambda\left(\overline{X}\right) = \frac{\lambda}{n}$$

Como se trata de un estimador insesgado, su ECM coincide con su varianza:

$$\mathrm{ECM}\left(\overline{X}\right) = \frac{\lambda}{n} + 0^2 = \frac{\lambda}{n} = \mathrm{Var}_\lambda(\overline{X})$$

Adicionalmente, observamos que el error tiende a cero:

$$\lim_{n\to\infty} \mathrm{ECM}\left(\overline{X}\right) = \lim_{n\to\infty} \frac{\lambda}{n} = 0$$

3.7. La varianza de $\overline{X}$ alcanza la cota de Rao-Cramér

Vimos que la varianza del estimador $\overline{X}$ es: $\mathrm{Var}_\lambda[\overline{X}] = \frac{\lambda}{n}$. Por el teorema de Rao-Cramér sabemos que la varianza de un estimador tiene la siguiente cota inferior:

$$\mathrm{Var}_\lambda[\delta_\mathrm{n}(\underline{\mathrm{X}})] = \mathrm{Var}_\lambda[\overline{\mathrm{X}}] \geq \frac{[q'(\lambda)]^2}{nI_1(\lambda)}$$

En nuestro caso, tenemos que $q(\lambda) = \lambda$, $q'(\lambda) = 1$, $[q'(\lambda)]^2 = 1$. Por otro lado, una manera de obtener el número de información de Fisher es la siguiente:

$$I_1(\lambda) = -E_\lambda \left[\frac{\partial^2}{\partial \lambda^2} \ln p(X, \lambda) \right]$$

En nuestro caso, tenemos:

$$p(x, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$\ln p(x, \lambda) = x \ln(\lambda) - \ln(x!) - \lambda$$

$$\frac{\partial}{\partial \lambda} \ln p(x, \lambda) = \frac{x}{\lambda} - 1$$

$$\frac{\partial^2}{\partial \lambda^2} \ln p(x, \lambda) = -\frac{x}{\lambda^2}$$

$$I_1(\lambda) = -\mathbb{E}_\lambda \left[-\frac{X}{\lambda^2} \right] = \frac{1}{\lambda^2} \mathbb{E}_\lambda[X] = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

La cota de Rao-Cramér queda entonces como:

$$\frac{[q'(\lambda)]^2}{n I_1(\lambda)} = \frac{1}{n \frac{1}{\lambda}} = \frac{\lambda}{n}$$

lo que equivale a la varianza del estimador $\overline{X}$, de modo que el estimador alcanza su cota de Rao-Cramér (y por ende es IMVU).

3.8. El estimador $\overline{X}$ es asintóticamente normal y eficiente

Para que un estimador δ_n sea asintóticamente normal y eficiente (ANE), debe cumplir que:

$$\sqrt{n}(\delta_n - q(\theta)) \xrightarrow[n \to \infty]{\mathcal{D}} \mathcal{N} \left(0, \frac{[q'(\lambda)]^2}{I_1(\lambda)} \right)$$

Para nuestro estimador, vimos que $[q'(\lambda)]^2 = 1$ y que $I_1(\lambda) = \frac{1}{\lambda}$, de modo que es ANE si cumple que:

$$\sqrt{n}(\overline{X} - \lambda) \xrightarrow[n \to \infty]{\mathcal{D}} \mathcal{N}(0, \lambda)$$

Por el Teorema del Límite Central, sabemos que:

$$\sqrt{n}(\overline{X} - \mathbb{E}[X]) \xrightarrow[n \to \infty]{\mathcal{D}} \mathcal{N}(0, \mathrm{Var}[X])$$

Con una v. a. distribuida como Poisson, tenemos que $\mathbb{E}[X] = \lambda$ y $\mathrm{Var}[X] = \lambda$. Así pues, el TLC afirma:

$$\sqrt{n}(\overline{X} - \lambda) \xrightarrow[n \to \infty]{\mathcal{D}} \mathcal{N}(0, \lambda)$$

Esta afirmación coincide exactamente con lo que queríamos mostrar. En consecuencia, el estimador $\overline{X}$ es ANE.

4. Distribución a Posteriori si Consideramos que a priori es Exponencial

- No parece idóneo tomar la beta como distribución de los parámetros porque $\lambda \in \mathbb{R}^+$. Probamos con la distribución exponencial.

$$p\left(\lambda|\boldsymbol{T}=t\right)=\frac{\frac{e^{-\lambda}\lambda^{\sum_{i=1}^{n}x_i}}{\sum_{i=1}^{n}x_i!}\cdot\theta\cdot e^{-\theta\lambda}}{\int_0^\infty\frac{e^{-\lambda}\lambda^{\sum_{i=1}^{n}x_i}}{\sum_{i=1}^{n}x_i!}\cdot\theta\cdot e^{-\theta\lambda}d\lambda}$$

$$p\left(\lambda|\boldsymbol{T}=t\right)=\frac{e^{-\lambda}\lambda^{\sum_{i=1}^{n}x_i}\cdot e^{-\theta\lambda}}{\int_0^\infty\lambda^{\sum_{i=1}^{n}x_i}\cdot e^{-\lambda(\theta+1)}d\lambda}$$

o una gamma
$Y\sim\Gamma\left(\alpha;\lambda\right)\Rightarrow f_Y\left(y\right)=\frac{\lambda^\alpha}{\Gamma(\alpha)}e^{-\lambda y}y^{\alpha-1}$
$\Lambda\sim\Gamma\left(\sum_{i=1}^{n}x_i+1;\theta+1\right)\Rightarrow f_\lambda\left(\lambda\right)=\frac{(\theta+1)^{\sum_{i=1}^{n}x_i+1}}{\Gamma(x+1)}e^{-(\theta+1)\lambda}\lambda^{\sum_{i=1}^{n}x_i}$

$$p\left(\lambda|\boldsymbol{T}=t\right)=\frac{e^{-\lambda}\lambda^{x}\cdot e^{-\theta\lambda}}{\frac{\Gamma\left(\sum_{i=1}^{n}x_i+1\right)}{(\theta+1)^{\sum_{i=1}^{n}x_i+1}}\cdot\underbrace{\int_0^\infty\frac{(\theta+1)^{\sum_{i=1}^{n}x_i+1}}{\Gamma\left(\sum_{i=1}^{n}x_i+1\right)}\lambda^{\sum_{i=1}^{n}x_i}\cdot e^{-\lambda(\theta+1)}d\lambda}_{1}}$$

$$p\left(\lambda|\boldsymbol{T}=t\right)=\frac{(\theta+1)^{\sum_{i=1}^{n}x_i+1}\cdot e^{-\lambda}\lambda^{\sum_{i=1}^{n}x_i}\cdot e^{-\theta\lambda}}{\Gamma\left(\sum_{i=1}^{n}x_i+1\right)}$$

$p\left(\lambda|\boldsymbol{T}=t\right)=\frac{(\theta+1)^{\sum_{i=1}^{n}x_i+1}\lambda^{\sum_{i=1}^{n}x_i}\cdot e^{-(\theta+1)\lambda}}{\Gamma\left(\sum_{i=1}^{n}x_i+1\right)}$
$p\left(\lambda|\boldsymbol{T}=t\right)=\frac{(\theta+1)^{\sum_{i=1}^{n}x_i+1}}{\Gamma\left(\sum_{i=1}^{n}x_i+1\right)}\lambda^{\sum_{i=1}^{n}x_i}\cdot e^{-(\theta+1)\lambda}$
Es decir que la distribución es una gamma $\Lambda\sim\Gamma\left(\sum_{i=1}^{n}x_i+1;\theta+1\right)$.

4.1. Si Buscamos Maximizar la probabilidad:

$$p\left(\lambda|\boldsymbol{T}=t\right)=\frac{(\theta+1)^{\sum_{i=1}^{n}x_i+1}\,\lambda^{\sum_{i=1}^{n}x_i}\cdot e^{-(\theta+1)\lambda}}{\Gamma\left(\sum_{i=1}^{n}x_i+1\right)}$$

$$p\left(\lambda|\boldsymbol{T}=t\right)=\frac{1}{\Gamma\left(\sum_{i=1}^{n}x_i+1\right)}\cdot\frac{(\theta+1)^{\sum_{i=1}^{n}x_i+1}\,\lambda^{\sum_{i=1}^{n}x_i}}{e^{(\theta+1)\lambda}}$$

$$p\left(\lambda|\boldsymbol{T}=t\right)=\frac{(\theta+1)^{\sum_{i=1}^{n}x_i+1}}{\Gamma\left(\sum_{i=1}^{n}x_i+1\right)}\cdot\frac{\lambda^{\sum_{i=1}^{n}x_i}}{e^{(\theta+1)\lambda}}$$

$$p\left(\lambda|\boldsymbol{T}=t\right)=\frac{(\theta+1)^{\sum_{i=1}^{n}x_i+1}}{\Gamma\left(\sum_{i=1}^{n}x_i+1\right)}\cdot\frac{e^{\sum_{i=1}^{n}x_i\ln\lambda}}{e^{(\theta+1)\lambda}}$$

- $$\max\left(\sum_{i=1}^{n} x_i \ln\lambda - (\theta+1)\,\lambda\right)$$