Tema

4

Inferencia Bayesiana en el Modelo de Poisson

Índice

1. Distribuciones asociadas

1.1. La distribución Gamma

La familia de distribuciones Gamma son funciones de densidad no simétricas definidas para v.a. no negativas con mucha utilidad en diversos campos como en Fiabilidad Industrial, Medicina, Biología, etc.

Definición 1.1. *Se dice que una v.a. X sigue una distribución Gamma de parámetros $a > 0$ y $\lambda > 0$ (y se denota por $X \sim Ga(a, \lambda)$) si su función de densidad es de la forma*

$$f(x; a, \lambda) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \text{ para } x > 0, a > 0 \text{ y } \lambda > 0.$$

Nota 1. *El parámetro λ es una medida de escala y también se tiene, como caso particular, que la distribución Gamma con $a = 1$ es la distribución exponencial negativa de parámetro λ.*

Proposición 1.2. *Si $X \sim Ga(a, \lambda)$ entonces $Y = 2\lambda X \sim \chi^2_{2a}$.*

Proposición 1.3. *Si $X \sim Ga(a, \lambda)$ entonces*

$$E(X) = \frac{a}{\lambda}, \quad var(X) = \frac{a}{\lambda^2} \text{ y para } a > 1 \text{ se tiene } moda(X) = \frac{a-1}{\lambda}.$$

1.2. La distribución Gamma-Invertida

La distribución Gamma-Invertida es una función de densidad no simétrica definida para v.a. no negativas.

Definición 1.4. *Se dice que una v.a. X sigue una distribución Gamma-Invertida de parámetros $a > 0$ y $\lambda > 0$ (y se denota por $X \sim GaI(a, \lambda)$) si su función de densidad es de la forma*

$$f(x; a, \lambda) = \frac{\lambda^a}{\Gamma(a)} \frac{1}{x^{a+1}} e^{-\lambda/x} \text{ para } x > 0, a > 0 \text{ y } \lambda > 0.$$

Proposición 1.5. *Si $X \sim GaI(a, \lambda)$ entonces $Y = 1/X \sim Ga(a, \lambda)$.*

Proposición 1.6. *Si $X \sim GaI(a, \lambda)$ entonces para $a > 2$ se tiene que*

$$E(X) = \frac{\lambda}{a-1}, \quad var(X) = \frac{\lambda^2}{(a-1)^2(a-2)} \text{ y } moda(X) = \frac{\lambda}{a+1}.$$

2. La distribución de Poisson

La distribución de Poisson se ha estudiado ampliamente en Cálculo de Probabilidades de primer curso. La v.a. de Poisson sirve para modelar circunstancias especiales de recuento de un suceso cuando este puede ocurrir varias veces por unidades temporales. Por ejemplo, cuantas conexiones a determinada web se producen en una hora. También se vió como la distribución de Poisson resulta como límite de una binomial cuando $n \to \infty$, $p \approx 0$ y $np \to \lambda$. Recordemos la definición.

Definición 2.1. *Se dice que la variable aleatoria X sigue una distribución de Poisson de parámetro $\lambda > 0$ (y se denota por $X \sim Poi(\lambda)$) si su función de probabilidad es*

$$P(X = k; \lambda) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \text{ para } k = 0, 1, \ldots$$

Se demuestró que $E(X) = \lambda$ y que $var(X) = \lambda$. El parámetro λ se interpreta como la intensidad media de la v.a. Es decir, el número medio de veces que ocurre dicho suceso por unidad temporal.

3. La función de verosimilitud

Supongamos que $X|\theta \sim Poi(\theta)$. La función de verosimilitud para un conjunto de datos $\mathbf{x}$ procedente de una muestra aleatoria simple de tamaño n sería:

$$\begin{aligned} L(\theta|\mathbf{x}) &= \prod_{i=1}^{n} P(X = x_i|\theta) \\ &= \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{x_i}}{x_i\,!} \\ &= e^{-n\theta} \frac{\theta^{\sum x_i}}{\prod x_i\,!}. \end{aligned}$$

En realidad, para una visualización gráfica, únicamente interesa una función de θ proporcional:

$$L(\theta|\mathbf{x}) \propto e^{-n\theta} \theta^{\sum x_i}, \text{ para } \theta > 0.$$

4. Distribución a priori y a posteriori

En este caso no vamos a desarrollar el caso de la a priori no informativa completa ya que sería similar a la a priori de Jeffrey.

4.1. A priori no informativa de Fisher

más Sabemos que la función Información de Fisher viene dada por

$$I(\theta) = E\left[\left(\frac{\partial f(X|\theta)}{\partial \theta}\right)^2\right],$$

entonces es fácil obtener que

$$I(\theta) = 1/\theta.$$

Por tanto,

$$\pi(\theta) \propto \theta^{-1/2},$$

que obviamente se trata de una a priori impropia.

En definitiva, la densidad a posteriori sería:

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto L(\theta|\mathbf{x}) \cdot \pi(\theta) \\ &\propto e^{-n\theta}\theta^{\sum x_i} \cdot \theta^{-1/2} \\ &\propto e^{-n\theta}\theta^{\sum x_i - 1/2}.\end{aligned}$$

Es decir, $\theta|\mathbf{x} \sim Ga(\sum x_i + 1/2, n)$.

4.2. A priori gamma

Supongamos el caso conjugado, es decir $\theta \sim Ga(\alpha, \beta)$ donde $\alpha > 0$ y $\beta > 0$ son valores conocidos.

Por tanto,

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto L(\theta|\mathbf{x}) \cdot \pi(\theta) \\ &\propto e^{-n\theta}\theta^{\sum x_i} \cdot e^{-\theta\beta}\theta^{\alpha-1} \\ &\propto e^{-(n+\beta)\theta}\theta^{\sum x_i + \alpha - 1}.\end{aligned}$$

Es decir, $\theta|\mathbf{x} \sim Ga(\sum x_i + \alpha, n + \beta)$.

4.3. Elección de una a priori

En este caso no es fácil la elección de una a priori cuando se tiene sospecha sobre el valor de θ. La mejor opción sería elegir una distribución gamma con moda igual a θ_0 (valor que se cree más probable) a partir de

$$\frac{\alpha - 1}{\beta} = \theta_0,$$

conjuntamente con otra ecuación que relacione los parámetros α y β y la información que se disponga.

5. Estimadores de Bayes e Intervalos de Credibilidad

5.1. Estimadores de Bayes

Como hemos demostrado anteriormente, la distribución a posteriori es $\theta|\mathbf{x} \sim Ga(a(\mathbf{x}), b(\mathbf{x}))$, luego si se usa la funicón de pérdida cuadrática el estimador de Bayes será la media de la gamma:

$$\hat{\theta}_B = \frac{a(\mathbf{x})}{b(\mathbf{x})},$$

donde $a()$ y $b()$ son funciones de los datos y del tamaño muestral que dependerá su expresión de la distribución a priori que se seleccione.

Para el caso de la función de pérdida valor absoluto el estimador de Bayes es la mediana de una distribución gamma y para su cálculo tenemos que recurrir al software **R**.

5.2. Intervalos de Credibilidad

Debido a que la distribución a posteriori es una distribución Gamma, se usará el método de HPDI o el método de las colas igualmente ponderadas para proponer un intervalo de credibilidad asegurándonos que la moda de la distribución a posteriori pertenezca a dicho intervalo. En cualquier caso, también tendremos que recurrir al **R** para su cálculo.

6. Factor de Bayes

Supongamos que deseamos realizar el siguiente contraste de hipótesis:

$$H_0 : \theta \leq \lambda_0 \text{ frente a } H_1 : \theta > \lambda_0.$$

Partimos de un conjunto de datos $\mathbf{x}$ procedente de la realización n pruebas independientes del modelo de Poisson. Luego en este caso se tendría que

$$FB_{01} = \frac{P(\theta \leq \lambda_0|\mathbf{x})}{P(\theta > \lambda_0|\mathbf{x})} \cdot \frac{P(\theta > \lambda_0)}{P(\theta \leq \lambda_0)}.$$

El primer factor corresponde a las distribuciones a posteriori mientras que el segundo a las distribuciones a priori. También es evidente que necesitaremos **R** para llevar a cabo este cálculo. Posteriormente decidiremos la evidencia de las hipótesis en función del valor del FB_{01} según la tabla de Jeffrey.

7. Distribución Predictiva

Supongamos como hasta ahora que $\theta|\mathbf{x} \sim Ga(a,b)$, donde por simplicidad en la notación se ha eliminado la dependencia de los datos muestrales $\mathbf{x}$. El objetivo es calcular $P(Y=k|\mathbf{x})$. Por tanto, se tiene que

$$\begin{aligned}
P(Y=k|\mathbf{x}) &= \int_0^{\infty} P(Y=k|\mathbf{x},\theta)\pi(\theta|\mathbf{x})d\theta \\
&= \int_0^{\infty} e^{-\theta}\frac{\theta^k}{k!}\,\frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}d\theta \\
&= \frac{b^a}{\Gamma(a)}\frac{1}{k!}\int_0^{\infty}\theta^{k+a-1}e^{-\theta(b+1)}d\theta \\
&= \frac{b^a}{\Gamma(a)}\frac{1}{k!}\frac{\Gamma(a+k)}{(b+1)^{a+k}} \\
&\text{en el caso que } a\in\mathbb{N} \text{ se tiene que} \\
&= \frac{(a+k-1)!}{(a-1)!k!}\left(\frac{b}{b+1}\right)^a\left(\frac{1}{b+1}\right)^k \quad \text{para } k=0,1,\ldots
\end{aligned}$$

Es decir, $Y|\mathbf{x}$ sigue una distribución binomial negativa con parámetros $(a,\frac{b}{b+1})$.

Ejercicios de autoaprendizaje. Supongamos que $X|\lambda \sim Poi(\lambda)$ e $Y|\lambda \sim Poi(\lambda)$ y ambas son independientes. Si se toman muestras aleatorias de tamaño n y m respectivamente, se pide:

1. Calcular la función de verosimilitud de λ y representarla.
2. Asumiendo la a priori de Fisher, calcular la a posteriori.
3. Calcular la distribución predictiva para una nueva observación de ambas.