

# Inferencia Bayesiana (21 de Junio de 2007)

## Soluciones

**Problema 1.** El número de accidentes laborales sufridos por el sector de la construcción en una determinada región en las últimas 20 semanas han sido  $\{7, 4, 4, 3, 6, 3, 6, 3, 8, 5, 2, 3, 4, 5, 4, 2, 7, 7, 6, 6\}$ . Suponiendo que los datos pueden ser considerados como una muestra aleatoria de una distribución de Poisson,  $Pn(x | \lambda) = e^{-\lambda} \lambda^x / x!$ , determinar la probabilidad de que la semana próxima el sector sufra en esa región algún accidente laboral.

La función de información de Fisher correspondiente a una distribución de Poisson  $Pn(x | \lambda)$  es  $i(\lambda) = \lambda^{-1}$  y, consecuentemente, la función inicial de referencia es  $\pi(\lambda) = \lambda^{-1/2}$ . La función de verosimilitud es  $l(\lambda | \mathbf{x}) = \lambda^{n\bar{x}} \exp[-\lambda n]$ , de forma que  $(n, \bar{x})$  es suficiente y, utilizando el teorema de Bayes, la distribución final de referencia es  $\pi(\lambda | \bar{x}, n) = \text{Ga}(\lambda | n\bar{x} + \frac{1}{2}, n)$ . Utilizando la integral gamma,  $\int_0^\infty x^{\alpha-1} \exp(-\beta x) dx = \Gamma(\alpha) \beta^{-\alpha}$ , la distribución predictiva resulta ser

$$p(x | \mathbf{x}) = \int_0^\infty Pn(x | \lambda) \text{Ga}(\lambda | n\bar{x} + 1/2, n) dx = \frac{1}{x!} \frac{\Gamma(n\bar{x} + 1/2 + x)}{\Gamma(n\bar{x} + 1/2)} \frac{n^{n\bar{x}+1/2}}{(1+n)^{n\bar{x}+1/2+x}},$$

que para el caso  $x = 0$  se reduce a  $p(x = 0 | \mathbf{x}) = \left(\frac{n}{n+1}\right)^{n\bar{x}+1/2}$ . Con los datos del problema,  $n = 20$ ,  $\bar{x} = 4.75$ , resulta  $p(x = 0 | \mathbf{x}) \approx 0.9905$ .

**Problema 2.** Las rentas anuales de las familias de una determinada región del mundo (en miles de euros) tienen una distribución log-normal  $p(x | \mu) = (x\sqrt{2\pi})^{-1} e^{-(\mu - \log[x])^2/2}$ ,  $x > 0$ ,  $\mu \in \mathfrak{R}$ . Las rentas correspondientes a 5 familias escogidas al azar en esa región han resultado ser (en miles de euros)  $\{9.989, 5.563, 4.196, 13.610, 3.011\}$ . (i) Determinar la distribución final de referencia de  $\mu$ , y (ii) determinar el estimador puntual de  $\mu$  que corresponde a una pérdida cuadrática.

La distribución log-normal  $\text{Lno}(x | \mu) = (x\sqrt{2\pi})^{-1} e^{-(\mu - \log[x])^2/2}$  es una distribución acampanada, sobre  $\mathfrak{R}^+$ , con una moda única,  $\text{Mo}[x | \mu] = e^{\mu-1}$ , solución de la ecuación  $\partial \log[\text{Lno}(x | \mu)] / \partial x = 0$ .

(i) La función de información de Fisher correspondiente a  $\text{Lno}(x | \mu)$  es  $i(\mu) = -E_{x|\mu} \partial \log[\text{Lno}(x | \mu)] / \partial \mu^2 = 1$  y, consecuentemente, la función inicial de referencia es  $\pi(\mu) = 1$ . La función de verosimilitud es

$$l(\mu | \mathbf{x}) \propto \exp\left[-\frac{n}{2}(\mu - t)^2\right], \quad t = n^{-1} \sum_{j=1}^n \log[x_j],$$

de forma que  $(n, t)$  es suficiente y, utilizando el teorema de Bayes, la distribución final de referencia es

$$\pi(\mu | t, n) = \text{N}(\mu | t, 1/\sqrt{n}).$$

(ii) El estimador puntual correspondiente a una pérdida cuadrática es la media a posteriori, de forma que el estimador pedido es  $\mu^*(\mathbf{x}) = E[\mu | \mathbf{x}] = t$ . Con los datos del problema,  $\mu^*(\mathbf{x}) = t = 1.833$ .