Inferencia Estadística II

Septiembre de 2005

Problema I (3 puntos) El test adecuado para contrastar la independencia entre el nivel de renta de un matrimonio y el número de hijos con los datos de los que disponemos está basado en el estadístico de Pearson:

$$\begin{aligned} K_n^* &= \sum_{i=1}^{3}\sum_{j=1}^{4} \frac{\left(N_{ij} - \frac{N_{i.}N_{.j}}{n}\right)^2}{\frac{N_{i.}N_{.j}}{n}} \\ &= \left[\frac{\left(25 - \frac{51\cdot 45}{180}\right)^2}{\frac{51\cdot 45}{180}} + \ldots + \frac{\left(28 - \frac{43\cdot 68}{180}\right)^2}{\frac{43\cdot 68}{180}}\right] \\ &> 20 \end{aligned}$$

donde $n = 180$.
Sabemos que la distribución asintótica de este estadístico es una χ^2 con $(3-1)(4-1) = 6$ grados de libertad. El cuantil 5% de esta distribución siendo igual a $\chi^2_{6,5\%} = 12.59$, rechazamos la hipótesis de independencia.

Problema II (3 puntos).

1. El test de Neyman-Pearson (NP) está basado en el ratio de verosimilitudes:

$$\frac{f_1(X)}{f_0(X)} = \frac{e^{-3X+1.15}}{e^{5X-3.38}} = Ce^{-8X}$$

Claramente este ratio crece cuando X decrece, por lo tanto el test de NP se puede expresar como

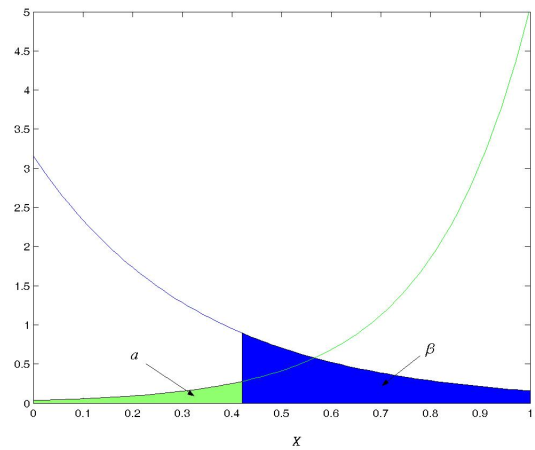
Rechazar H_0 si $X < u$

El nivel del test $\alpha = 5\%$, entonces el umbral u debe verificar

$$\begin{aligned} 5\% &= \mathbf{P}\left(X < u | H_0\right) \\ &= \int_0^u f_0\left(x\right) dx = \int_0^u e^{5x-3.38} dx = \frac{e^{-3.38}}{5}\left(e^{5u} - 1\right) = .05 \end{aligned}$$

De la anterior ecuación se deduce que $u_\alpha \simeq 0.42$.

2. En el gráfico a continuación aparece los dos riesgos α y β asociados al test de NP:



3. La potencia del test se define como la probabilidad de rechazar H_0 cuando es falsa, o sea

$$\begin{aligned} 1-\beta &= \mathbf{P}\left(X < u_\alpha | H_1\right) \\ &= \int_0^{u_\alpha} e^{-3x+1.15} dx \\ &= \frac{e^{1.15}}{3}\left[e^{-3x}\right]_{0.42}^{0} = \frac{e^{1.15}}{3}\left(1-e^{-3\times 0.42}\right) \simeq 75\% \end{aligned}$$

Por lo tanto, con este test podremos a partir de la observación de X detectar que Greta es zurda en el 75% de los casos.

Problema III (4 puntos)..

1. La función de perdida siendo cuadrática, la mejor predicción de θ será aquí:

$$d_Q = \mathbf{E}_Q\left(\theta\right) = 15$$

puesto que la distribución *a priori* Q es exponencial con media $\frac{1}{\lambda} = 15$. El riesgo bayesiano de esta predicción es $\overline{w}_Q\left(d_Q\right) = \mathbf{var}_Q\left(\theta\right) = \frac{1}{\lambda^2} = 225$.

2. .

   (a) La distribución de θ condicionada por X es una Gamma $G(X+1,\lambda+1)$ con $\lambda = 0.07$ (ver demostración en la solución del examen de Junio de 2002).

   (b) Por lo tanto, la mejor predicción de θ conociendo X será:

$$d_Q\left(X\right) = \mathbf{E}_Q\left(\theta|X\right) = \frac{X+1}{\lambda+1} = \frac{X+1}{1.07}$$

   (c) El riesgo de esta predicción es

$$\begin{aligned} W_Q\left(d_Q\left(X\right)\right) &= \mathbf{E}\left[\mathbf{var}_Q\left(\theta|X\right)\right] \\ &= \mathbf{E}\left[\frac{X+1}{\left(\lambda+1\right)^2}\right] \\ &= \frac{\mathbf{E}\left[X\right]+\mathbf{1}}{\left(\lambda+1\right)^2} = \frac{\lambda^{-1}+1}{\left(\lambda+1\right)^2} = \frac{1}{\lambda\left(\lambda+1\right)} \end{aligned}$$

   porque $\mathbf{E}\left[X\right] = \mathbf{E}\left[\mathbf{E}\left(X|\theta\right)\right] = \mathbf{E}\left[\theta\right] = \lambda^{-1}$.
   Puesto que $\lambda\left(\lambda+1\right) > \lambda^2$ concluimos que $W_Q\left(d_Q\left(X\right)\right) < \overline{w}_Q\left(d_Q\right)$, o sea que $d_Q(X)$ es más preciso que d_Q.