Inferencia Estadística II

Septiembre del 2001

Problema 1 (3 puntos).

1. Se realizaron pruebas del nivel de colesterol en un grupo de padres e hijos. Los datos recogidos (clasificados en tres niveles de colesterol: Bajo, Medio y Alto) aparecen en la siguiente tabla:

		Padres		
		Bajo	Medio	Alto
Hijos	Bajo	2023	175	90
	Medio	125	355	152
	Alto	76	121	215

   Se quiere contrastar la hipótesis de independencia entre el nivel de colesterol de los padres y de los hijos. Calcular el estadístico χ^2 para contrastar esta hipótesis y deducir la decisión del test para un nivel de 5%.

2. Utilizando el test basado en ele estadístico χ^2, contrastar la hipótesis de simetría de la tabla con un nivel $\alpha = 5\%$.

Problema II (4 puntos). Se quiere comprobar estadísticamente si en media las mujeres españolas y francesas tardan lo mismo en tener su primer hijo. Se observaron dos muestras de edades a las cuales mujeres españolas y francesas tuvieron su primer hijo: $X_1, X_2, \ldots, X_n$ (mujeres españolas) e $Y_1, Y_2, \ldots, Y_m$ (mujeres francesas). Suponemos que las dos muestras son independientes y que la primera muestra de edad viene de una distribución exponencial $Exp\,(\lambda_E)$ y la segunda de una exponencial $Exp\,(\lambda_F)$. Queremos contrastar $H_0 : \lambda_E = \lambda_F$ frente a su alternativa $H_1 : \lambda_E \neq \lambda_F$.

1. Calcular los estimadores máximo verosímiles de λ_E y λ_F bajo H_0 y bajo cualquier hipótesis.

2. Calcular el test de razón de verosimilitudes para $\alpha = 5\%$.

3. Deducir la decisión del test en el caso en el cual se observa $\sum_{i=1}^{100} X_i = 2125$, $\sum_{j=1}^{120} Y_j = 2730$ con $n = 100$, $m = 120$.

Problema III (3 puntos). En un modelo estadístico complejo, se quiere evaluar la información sobre un párametro desconocido. El valor cuantitativo de esta información, que denotamos θ, varía en función de la muestra de datos observados. A partir de varias muestras del mismo tamaño, se comprobó que el valor θ de esta información sigue una distribución Q exponencial con media 10. Para medir el error de una predicción d del valor de la información, se considera la función de pérdida cuadrática $w(d,\theta) = (d-\theta)^2$

1. Calcular la mejor predicción d_Q basada en la distribución *a priori* Q y calcular el riesgo bayesiano $\overline{w}_Q\,(d_Q)$.

2. Para mejorar esta predicción se observan nuevas muestras del mismo tamaño. Mediante cada muestra, obtenemos una estimación del parámetro desconocido. Denotamos $X = (X_1, X_2, \ldots, X_n)$ el conjunto de las n estimaciones centradas que obtuvimos a partir de n muestras de datos. Se sabe que condicionalmente al conocimiento del valor de la infomación θ, estas estimaciones corresponden a n realizaciones independientes de una distribución normal centrada y con varianza $1/\theta$.

   (a) Demostrar que la distribución de θ a posteriori es una distribución Gamma
   $G\left(\sum_{i=1}^{n} X_i^2/\,2 + 0.1, n/2+1\right)$

   (b) Deducir la mejor predicción $d_Q(X)$ basada en X y en la distribución a priori Q, y dar una expresión del error de esta predicción.

Indicaciones:

1. (i) La función de densidad de una distribución gamma $G(a,b)$ se puede escribir como

$$f(x) = \frac{x^{a-1}\exp(-bx)}{C}$$

para $x > 0$ y 0 en caso contrario, donce C es una constante que no depende de x.
La media de esta distribución es $\frac{a}{b}$ y su varianza es $\frac{a}{b^2}$.

(ii) Una distribución exponencial $Exp(\lambda)$ con parámetro λ es una Gamma $G(1, \lambda)$ y su función densidad se puede escribir

$$f(x) = \lambda \exp(-\lambda x)$$

para $x > 0$ y 0 en caso contrario.

(iii) La función de densidad de una distribución normal con media m y varianza v se puede escribir como

$$f(x) = \frac{1}{(2\pi)^{\frac{1}{2}}\, v^{\frac{1}{2}}} \exp\left(-\frac{1}{2v}(x-m)^2\right),$$

para $x \in \mathbb{R}$.

(iv) A continuación, figuran los valores del cuantil 5% de la distribución χ^2 para distintos valores del número de grados de libertad:

$$\begin{aligned} \chi^2_{1,5\%} &= 3.84\ ;\ \chi^2_{2,5\%} = 5.99\ ;\ \chi^2_{3,5\%} = 7.81 \\ \chi^2_{4,5\%} &= 9.48\ ;\ \chi^2_{5,5\%} = 11.07\ ;\ \chi^2_{6,5\%} = 12.59 \end{aligned}$$