Inferencia Estadística II

Dr. Olivier Nuñez Boyer

7 de Junio del 2000

Problema 1 (5 puntos). Se supone que la distancia (expresada en metros) entre dos coches en un atasco es una variable aleatoria X que sigue una distribución Gamma con parámetros k y 2 (Chi cuadrado con k grados de libertad). La función de densidad de esta distribución se define por

$$f(x) = \frac{2^{k/2}}{(k/2-1)!} \exp(-2x) \ x^{k/2-1},$$

para $x > 0$.
A partir de la observación de una muestra de n distancias $X_1, X_2, ..., X_n$ entre coches se quiere contrastar:

$$H_0 : k = 1 \text{ frente a } H_1 : k > 1.$$

1. Escribir el test UMP con nivel α para el contraste anterior.

2. En el caso en que $n = 1$, $\alpha = 5\%$ y si $X_1 = 2$ metros, cuál sera la decisión del test (nota: el cuantil 5% de un χ^2 con 1 grado de libertad es igual a $\chi^2_{1,5\%} = 3,84$).

Ahora, se modelizan las distancias con una Gamma de parámetros k y λ, ambos desconocidos. La función de densidad correspondiente se escribe

$$f(x) = \frac{\lambda^k}{(k-1)!} \exp(-\lambda x) \ x^{k-1},$$

para $x > 0$.

Se quiere hallar un test UMP en este nuevo marco (λ desconocido) para el contraste precedente, utilizando un test condicional.

1. Hallar el estadístico S suficiente y mínimo para λ y k bajo H_0.

2. Mostrar que la densidad de las observaciones condicionada por $S = s$ es
$$f(x_1, x_2, \ldots, x_n|s) = \frac{\exp\left[(k-1)\sum_{i=1}^{n}\ln(x_i)\right]}{C(s,k)}$$
si $\sum_{i=1}^{n} x_i = s$ y 0 en el caso contrario, y donde $C(s,k)$ es una constante que depende solamente de s y k. Justificar por qué esta función de densidad no depende de los párametros bajo H_0 y interpretar la constante $C(s,k)$ cuando $k = 1$.

3. Especificar en que medida el test UMP con nivel α que se obtiene mediante este procedimiento es diferente del test obtenido en 1.

Problema II (5 puntos). La probabilidad θ de tener un hijo de sexo femenino varía según las parejas. Por tanto, se considera que θ es una variable aleatoria cuya distribución *a priori* Q es una Beta $\beta(2,1)$. Consideramos el caso de una pareja que decide tener hijos hasta el nacimiento de la primera niña. El número de hijos que tendrá esta pareja es entonces una variable geométrica con parámetro θ desconocido (aquí aleatorio). Para evaluar el número esperado de hijos $1/\theta$ que tendrá esta pareja se predice θ. El error de una predicción d de θ se mide mediante la función de pérdidas $w(d,\theta) = (d-\theta)^2$.

1. Hallar la predicción d_Q que minimiza el riesgo bayesiano $\overline{w}_Q(d) = E_Q\left[(d-\theta)^2\right]$ y calcular el valor del riesgo de esta predicción.

2. Este error de predicción se puede mejorar si se hace la hipótesis de que la probabilidad de tener una niña siempre ha sido la misma en la familia de esta pareja. Tenemos los datos siguiente: sean $X_1, X_2, ..., X_n$ los números de hijos que tuvieron n parejas de esta familia hasta tener una niña. Esta muestra de datos proviene entonces de una distribución geométrica con parámetro θ.

   (a) Comprobar que la distribución de θ *a posteriori* de estos datos es también una Beta.

   (b) Deducir la decisión $d_Q(X)$ basada sobre los datos y la distribución a priori Q, que minimiza el riesgo bayesiano medio.

   (c) Mostrar que $\sum_{i=1}^n X_i \geq n$ y que la esperanza marginal $E\left(\sum_{i=1}^n X_i\right) = 2n$.

   (d) Utilizando el resultado anterior, comprobar que el riesgo bayesiano medio asociado a la decisión $d_Q(X)$ verifica

   $$W_Q\left(d_Q(X)\right) \leq \frac{(n+2)(n+1)}{(n+3)^2(n+4)},$$

   y justificar el comportamiento de este riesgo cuando n crece. Deducir de este resultado un valor de n a partir del cual $d_Q(X)$ será mejor que d_Q. Comentar.

Indicaciones:

(i) Si la variable Z sigue una distribución geométrica con parámetro θ, tenemos que $P(Z = z) = (1 - \theta)^{z-1}\theta$ para $z = 1, 2, 3, ...$

(ii) La función de densidad de una distribución Beta $\beta(a, b)$ con parámetro $a > 0$ y $b > 0$ es

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{C(a,b)}$$

para $0 \leq x \leq 1$ y 0 en caso contrario, y donde $C(a, b)$ es una constante que no depende de x. Su media es $\frac{a}{a+b}$ y su varianza es $\frac{ab}{(a+b)^2(a+b+1)}$.