

Inferencia Estadística II

Junio del 2002

Problema 1 (4 puntos). En un servicio de urgencias de pediatría se sabe que la frecuencia θ (expresada en casos por día) de niños atendidos por asma, varía según los meses. Mediante experiencias de años anteriores se determinó que la variable θ sigue una distribución (*a priori*) exponencial con parámetro $\lambda = 0.2$. Con el fin de atender mejor a estos niños, se quiere predecir θ en el mes corriente. La función de pérdidas de una predicción d será cuadrática: $w(d, \theta) = (d - \theta)^2$.

1. Hallar la mejor predicción de θ mediante su distribución *a priori* y calcular el riesgo (o error) de esta predicción.
2. Para mejorar la predicción de θ , se observa durante los n primeros días del mes el número de niños con asma atendidos en este servicio diariamente. Suponemos que condicionalmente a θ , la muestra de observaciones X_1, X_2, \dots, X_n proviene de una distribución de Poisson con media θ .
 - (a) Demostrar que la distribución *a posteriori* de θ es una Gamma $G(\sum_{i=1}^n X_i + 1, n + \lambda)$.
 - (b) Hallar la mejor predicción de θ basada en su distribución *a priori* y en los datos.
 - (c) Demostrar que la anterior predicción es una combinación lineal de la media muestral de los datos \bar{X} y de la media *a priori* de θ . Comentar esta expresión.
 - (d) Utilizando que $E(X_i) = E[E(X_i|\theta)]$, demostrar que $E(X_i) = 5$. Deducir el riesgo de la predicción hallada en (b). Demostrar que este riesgo tiende hacia 0 cuando n tiende hacia infinito y que es

inferior al riesgo de la predicción hallada en (1) cuando $n \geq 1$.
Comentar el caso $n = 0$.

Problema II (3 puntos). Se realizó una encuesta en una muestra de consumidores sobre su intención de compra de una vivienda. Los datos son n variables de Bernoulli X_1, X_2, \dots, X_n con parámetro $p \in [0, 1]$, siendo este parámetro la probabilidad de que un consumidor quiera comprarse una vivienda:

$$P(X = x) = p^x(1 - p)^{1-x} \quad x = 0, 1$$

$$= \begin{cases} p & \text{si } x = 1 \\ 1 - p & \text{si } x = 0 \end{cases} .$$

Se quiere contrastar la hipótesis $H_0 : p \leq \frac{1}{2}$ frente a $H_1 : p > \frac{1}{2}$.

1. Demostrar que la función $p \rightarrow \ln\left(\frac{p}{1-p}\right)$ es creciente.
2. Demostrar que la razón de verosimilitudes es creciente con respecto al estadístico $T = \sum_{i=1}^n X_i$
3. Deducir la expresión del test UMP para el contraste anterior. Utilizando la aproximación normal de la distribución de T , dar la decisión del test (nivel de significación $\alpha = 5\%$) si para $n = 100$, se observó $T = \sum_{i=1}^{100} X_i = 43$.

Problema III (3 puntos). Se comprobó que en la primavera de 2001, el nivel diario (en miligramos por metro cúbico) de dióxido de carbono (CO2) en Madrid seguía una distribución exponencial con media 3.65. Con el fin de verificar la actualidad de este resultado, se han realizado nuevas observaciones durante 12 días de la primavera de 2002 que aparecen en la tabla a continuación:

nivel de CO2	2.8	3.1	3.5	2.9	3.7	3.2	3.6	2.4	3.4	2.7	2.2	3.0
--------------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

1. Sea P la distribución del nivel de CO2 en la primavera de 2002 y P_0 la distribución exponencial con parámetro $\lambda = \frac{1}{3.65}$. Contrastar la hipótesis $H_0 : P = P_0$ frente a $H_1 : P \neq P_0$ mediante el test del χ^2 (nivel de significación $\alpha = 5\%$). Se utilizarán 4 categorías definidas por

los intervalos $C_1 = [0; 1.05]$, $C_2 =]1.05; 2.53]$, $C_3 =]2.53; 5.06]$, $C_4 =]5.06; \infty[$, verificando que las probabilidades $p_j^0 = P_0(C_j) = 1/4$ para cada $j = 1, \dots, 4$.

2. Dado que el nivel de CO2 es una variable continua y que la decisión del test anterior es muy sensible a la elección de las categorías, proponer un test más adecuado para contrastar H_0 . Dar la decisión del test para un nivel de significación $\alpha = 5\%$.

Indicaciones:

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

$$f(y) = \frac{y^{a-1} \exp(-by)}{C}$$

para $y > 0$ y 0 en caso contrario, donde $C = \int y^{a-1} \exp(-by) dy$ es una constante que no depende de y .

La media de esta distribución es $\frac{a}{b}$ y su varianza es $\frac{a}{b^2}$.

- La función de distribución (acumulada) F y la función de densidad f de una distribución exponencial con parámetro λ se escriben

$$F(y) = 1 - \exp(-\lambda y) \quad \text{y} \quad f(y) = \lambda \exp(-\lambda y)$$

para $y > 0$ y 0 en caso contrario. Su media es $1/\lambda$ y su varianza es $1/\lambda^2$

- Una variable Y sigue una distribución de Poisson con media θ si

$$P(Y = y) = \frac{\theta^y}{y!} \exp(-\theta)$$

para cada $y = 0, 1, 2, \dots$

- A continuación figuran los valores de los cuantiles 5% de la χ^2 con 11 grados de libertad, de la normal estandar y de la distribución del estadístico de Kolmogorov-Smirnov Δ_{12} :

$$\chi_{3,5\%}^2 = 7.81 ; z_{5\%} = 1.64 ; \Delta_{12,5\%} = 0.45$$