Solución del Examen

Dr. Olivier Nuñez Boyer

Junio del 2002

Problema 1.

1. La distribución *a priori* Q de θ siendo una exponencial con parámetro $\lambda = 0.2$, deducimos que aquí la mejor predicción de θ (basada únicamente en Q) será $d_Q = E_Q(\theta) = \frac{1}{\lambda} = 5$. El error de esta predicción será

$$\overline{w}_Q\left(d_Q\right) = \mathrm{var}_Q\left(\theta\right) = \frac{1}{\lambda^2} = 25$$

   (a) Por la formula de Bayes, la densidad de la distribución *a posteriori* de θ se puede escribir:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{C}$$

   donde $x = (x_1, x_2, \ldots, x_n)$ es el conjunto de los datos y $C = \int f(x|\theta)f(\theta)d\theta$ es una constante que no depende de θ. Puesto que

$$\begin{aligned} f(x|\theta) &= \prod_{i=1}^{n} \frac{1}{x_i!}\theta^{x_i}e^{-\theta} \\ &= C_1\theta^{\sum_{i=1}^{n} x_i}e^{-n\theta} \end{aligned}$$

   C_1 siendo una constante que no depende de θ y $f(\theta) = \lambda e^{-\lambda\theta}$, deducimos que

$$f(\theta|x) = \frac{1}{C_2}\theta^{\sum_{i=1}^{n} x_i}e^{-(n+\lambda)\theta}$$

   que coincide con la densidad de la Gamma $G\left(\sum_{i=1}^{n} X_i + 1, n + \lambda\right).$

(b) Utilizando el resultado anterior, deducimos que la mejor predicción $d_Q(X)$ de θ basada en los datos y la distribución *a priori de* θ es

$$d_Q(X) = E\left(\theta|X\right) = \frac{\sum_{i=1}^{n} X_i + 1}{n+\lambda}$$

(c) Se puede escribir la anterior predicción de la siguiente manera

$$\begin{aligned} d_Q(X) &= \frac{n\overline{X}}{n+\lambda} + \frac{1}{n+\lambda} \\ &= \alpha_n \overline{X} + (1-\alpha_n)\frac{1}{\lambda} \end{aligned}$$

donde $\alpha_n = \frac{n}{n+\lambda} \in [0,1]$. Puesto que $\alpha_n \to 1$ cuando ($n \to \infty$ y/o λ es pequeño) y $\alpha_n \to 0$ cuando ($\lambda \to \infty$ y n es pequeño) . El peso dado a los datos crece con el tamaño muestral (o sea la precisión de $\overline{X}$) y el peso dado a la media *a priori* $d_Q = E(\theta) = 1/\lambda$ crece con la fiabilidad de esta predicción de θ ($\overline{w}_Q\left(d_Q\right) = \frac{1}{\lambda^2}$ decrece cuando $\lambda \to \infty$).

(d) Puesto que $E(\theta) = 5$ y $E(X_i|\theta) = \theta$ (condicionalmente a θ, X_i sigue una Poisson con media θ), deducimos que $E(X_i) = E(E(X_i|\theta)) = E(\theta) = 5$.
Sabemos que

$$\begin{aligned} W_Q(d_Q(X)) &= E\left[\operatorname{var}\left(\theta|X\right)\right] \\ &= E\left[\frac{\sum_{i=1}^{n} X_i + 1}{\left(n+\lambda\right)^2}\right] \\ &= \frac{\sum_{i=1}^{n} E\left[X_i\right] + 1}{\left(n+\lambda\right)^2} = \frac{5n+1}{\left(n+\lambda\right)^2} \end{aligned}$$

Por tanto $W_Q(d_Q(X)) \to 0$ cuando n crece hacia infinito. Se puede comprobar que $\frac{5n+1}{(n+0.2)^2}$ decrece cuando n crece, además para $n = 1$, tenemos que $W_Q(d_Q(X)) = 6/(1+0.2)^2 = 4.166 < w_Q(d_Q) = 25$. Deducimos que $W_Q(d_Q(X)) < w_Q(d_Q)$ cuando $n \geq 1$.
Si $n = 0$, $W_Q(d_Q(X)) = \frac{1}{\lambda^2} = w_Q(d_Q)$ como era de esperar, puesto que en este caso $d_Q(X)$ está basado sólo en la distribución *a priori*.

Problema 2.

1. Para $p \in]0,1[$, sea la función $h(p) = \ln\left(\frac{p}{1-p}\right) = \ln(p) - \ln(1-p)$. La derivada de h es $h'(p) = \frac{1}{p} + \frac{1}{1-p} > 0$ para cada $p \in]0,1[$. Por tanto h es creciente.

2. La verosimilitud de los datos es aquí

$$\begin{aligned} L_p(x) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^{n} X_i}(1-p)^{n-\sum_{i=1}^{n} X_i} \end{aligned}$$

Sea $p' > p$, el logaritmo de la razón de verosimilitudes será

$$\begin{aligned} \ln\left[\frac{L_{p'}(x)}{L_p(x)}\right] &= \ln(p')\sum_{i=1}^{n} X_i + \ln(1-p')\left(n - \sum_{i=1}^{n} X_i\right) \\ &\quad - \ln(p)\sum_{i=1}^{n} X_i - \ln(1-p)\left(n - \sum_{i=1}^{n} X_i\right) \\ &= [h(p') - h(p)]\sum_{i=1}^{n} X_i + C, \end{aligned}$$

C siendo una constante que no depende de los datos y $h(p) = \ln(p) - \ln(1-p)$. Puesto que h es creciente y $p' > p$, el coeficiente $(h(p') - h(p))$ es positivo. Por tanto, el logaritmo de la razón de verosimilitud es una función (aquí lineal) creciente de $T = \sum_{i=1}^{n} X_i$.

3. Puesto que T es una variable discreta, el test UMP tiene la expresión

$$\phi = \begin{cases} 1 & \text{si } T > u \\ \gamma & \text{si } T = u \\ 0 & \text{si } T < u \end{cases}.$$

Sin embargo, si se decide aproximar la distribución de T (una binomial $B(n,p)$) mediante una normal, estamos considerando que T es continua y el test será entonces: "rechazamos H_0 si $T > u$". La función de potencia siendo creciente, el nivel del test será $\alpha = P(T > u)$ cuando $p = \frac{1}{2}$. Para $n = 100$ y $p = \frac{1}{2}$, la media $E(T) = 50$ y $\mathrm{var}(T) = 25$. Por tanto,

$$\begin{aligned} 5\% &= P(T > u) \\ &= P\left(\frac{T-50}{5} > \frac{u-50}{5}\right) \simeq P\left(Z > \frac{u-50}{5}\right) \end{aligned}$$

donde Z sigue una normal estándar. Puesto que $z_{5\%} = 1.64$, deducimos que el umbral u debe verificar $(u-50)/5 = 1.64$, o sea, $u = 58.2$. Hemos observado $T = 43$, por tanto la decisión del test será aceptar H_0.

Problema 3.

1. Sea N_j el número de datos observado en la catgoría C_j, $j = 1, \ldots, 4$. Tenemos que $N_1 = 0$, $N_2 = 2$, $N_3 = 10$, y $N_4 = 0$. Bajo H_0, el número de datos esperado en cada categoría es el mismo e igual a $np_j^0 = 12 \times \frac{1}{4} = 3$. Por tanto el estadístico K_{12} del test del χ^2 será:

$$\begin{aligned} K_{12} &= \sum_{j=1}^{4} \frac{\left(N_j - np_j^0\right)^2}{np_j^0} \\ &= \frac{1}{3}3^2 + \frac{1}{3}1^2 + \frac{1}{3}7^2 + \frac{1}{3}3^2 \\ &= \frac{1}{3}68 = 22.667 \end{aligned}$$

Sabemos que la distribución (asintótica) de K_{12} es una χ^2 con $k-1$ grados de libertad, siendo k el número de categoría. Puesto que $\chi^2_{3,5\%} = 7.81$ rechazamos H_0 con un nivel (asintótico $\alpha = 5\%$).

1. Aquí, el test idóneo está basado en el estadístico de Kolmogorov-Smirnov Δ_n. El calculo de

$$\Delta_n = \max\left(\max_{1\le i\le n}\left[\frac{i}{n} - F_0(X_{(i)})\right], \max_{1\le i\le n}\left[F_0(X_{(i)}) - \frac{i-1}{n}\right]\right)$$

donde F_0 es la función de distribución de la exponencial con media 3.65, se deduce de la tabla siguiente ($n = 12$) :

$x_{(i)}$	$F_n\left(x_{(i)}\right) = \frac{i}{n}$	$F_0\left(x_{(i)}\right)$	$\frac{i}{n} - F_0(X_{(i)})$	$F_0(X_{(i)}) - \frac{i-1}{n}$
2.20	0.0833	0.4527	-0.3693	0.4527
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Sabemos que $\Delta_{12,5\%} = 0.45$ y sólo con la primera fila de esta tabla, tenemos que $\Delta_{12} \geq 0.4527$, por tanto $\Delta_{12} > \Delta_{12,5\%}$. Por consiguiente, rechazamos H_0.