Estadística I

Soluciones a los problemas del examen 10 de enero de 2013

---

1) (a) El estimador de máxima verosimilitud se obtiene maximizando en θ la función de verosimilitud

$$L_n(X_1, \dots, X_n; \theta) = \frac{\theta^n (-\log \theta)^{\sum X_i}}{X_1! \dots X_n!}.$$

Para maximizar en θ tomamos logaritmos y calculamos la derivada

$$\frac{\partial}{\partial \theta} \log L_n(X_1, \dots, X_n; \theta) = \frac{n}{\theta} + \frac{1}{\theta \log \theta} \sum_{i=1}^{n} X_i$$

La única solución de la ecuación $\frac{\partial}{\partial \theta} \log L_n(X_1, \dots, X_n; \theta) = 0$ es

$$\hat{\theta}_n = e^{-\bar{X}}.$$

Este valor corresponde a un máximo porque la derivada segunda en $\hat{\theta}_n$ es negativa.

Aplicando el TCL sabemos que $\sqrt{n}(\bar{X} - \mathbb{E}_\theta(X)) \xrightarrow{d} N(0, \sigma(\theta))$, siendo $\sigma(\theta) = V_\theta(X)^{1/2}$. Como $\hat{\theta}_n = g(\bar{X})$ con $g(u) = e^{-u}$ y esta función es derivable con derivada continua, podemos aplicar el método delta para obtener

$$\sqrt{n}(g(\bar{X}) - g(\mu)) \xrightarrow{d} N(0, |g'(\mu)| \sigma(\theta)),$$

denotando $\mu = \mathbb{E}_\theta(X)$. En definitiva, hemos obtenido

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \theta \sqrt{-\log \theta}),$$

(observemos que $-\log \theta > 0$ porque $\theta \in (0, 1)$).

(b)

$$I(\theta) = \mathbb{E}_\theta \left(-\frac{\partial^2}{\partial \theta^2} \log f(X; \theta) \right)$$

Tenemos

$$\frac{\partial}{\partial \theta} \log f(X; \theta) = \frac{1}{\theta} + \frac{1}{\theta \log \theta} X$$

$$\frac{\partial^2}{\partial\theta^2}\log f(X;\theta) = \frac{-1}{\theta^2} - \frac{X}{\theta^2\log\theta} - \frac{X}{\theta^2\log^2\theta}$$

$$I(\theta) = \mathbb{E}_\theta\left(-\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right) = \frac{1}{\theta^2} + \frac{\mathbb{E}_\theta(X)}{\theta^2\log\theta} + \frac{\mathbb{E}_\theta(X)}{\theta^2\log^2\theta} = -\frac{1}{\theta^2\log\theta}$$

La cantidad $I(\theta)$ es importante por varios motivos: bajo ciertas condiciones se verifica $V_\theta(T_n) \geq 1/(nI(\theta))$ (cota de Fréchet-Cramer-Rao) para estimadores insesgados T_n de θ. También (bajo condiciones de regularidad) el estimador de máxima verosimilitud $\hat{\theta}_n$ verifica

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, 1/\sqrt{I(\theta)}),$$

de manera que $I(\theta)^{-1}$ es la varianza de la distribución asintótica. En efecto, obsérvese que en este caso $I(\theta)^{-1}$ coincide con la varianza de la distribución límite obtenida en el apartado anterior.

(c) Si denotamos $g(u) = e^{-u}$,

$$(\hat{\theta}_n - \theta)^2 = (g(\bar{X}) - g(-\log\theta))^2 \leq (\bar{X} - (-\log\theta))^2. \qquad (*)$$

Para obtener esta desigualdad hemos usado el Teorema del Valor Medio, junto con el hecho de que $|g'(u)| = |e^{-u}| \leq 1$ para $u \geq 0$.

Tomando esperanzas,

$$\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] \leq \mathbb{E}_\theta[(\bar{X} - (-\log\theta))^2] = V_\theta(\bar{X}) = \frac{V_\theta(X)}{n} = \frac{-\log\theta}{n}.$$

Por tanto, dado cualquier $\epsilon > 0$,

$$\mathbb{P}_\theta\{n^{1/3}|\hat{\theta}_n - \theta| > \epsilon\} = \mathbb{P}_\theta\{n^{2/3}(\hat{\theta}_n - \theta)^2 > \epsilon^2\} \leq \text{ (usando la desigualdad de Markov)}$$
$$\leq \frac{n^{2/3}\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2]}{\epsilon^2} \overset{\text{(usando (*))}}{\leq} \frac{n^{2/3}(-\log\theta)}{n\epsilon^2} = \frac{-\log\theta}{n^{1/3}\epsilon^2} \to 0,$$

lo que demuestra $n^{1/3}(\hat{\theta}_n - \theta) \xrightarrow{P} 0$.

2) Supongamos que la v.a. X = "contenido de un paquete elegido al azar" tiene distribución $N(\mu, \sigma)$. Queremos contrastar

$$H_0 : \mu \geq 800 \text{ frente a } H_1 : \mu < 800.$$

Tenemos una muestra de $n = 20$ observaciones independientes de la v.a. X para la cual $\bar{x} = 793$, $s = 15$. La región crítica del test usual de nivel α para este problema es

$$\bar{x} - 800 < t_{n-1;1-\alpha}\frac{s}{\sqrt{n}}.$$

En este caso, $\bar{x} - 800 = -7$, $t_{n-1;1-\alpha} = t_{19,0.95} = -t_{19;0.05} = -1.729$, $s/\sqrt{n} = 15/\sqrt{20} = 3.354102$.

Por tanto,

$$t_{19;0.05}\frac{s}{\sqrt{n}} = -1.729 \cdot 3.354102 = -5.799242$$

Como $-7 < -5.799241$, se concluye que se ha encontrado suficiente evidencia estadística, al nivel 0.05, para aceptar H_1.

Si consideramos el nivel $\alpha = 0.01$, se tiene $t_{19;0.01}\frac{s}{\sqrt{n}} = -2.539 \cdot 3.354102 = -8.516065$.

Por tanto, al nivel 0.01 NO se ha encontrado evidencia estadística suficiente a favor de H_1.

Se concluye que el p-valor debe de ser mayor que 0.01 ya que el p-valor es el ínfimo de los valores del nivel de significación para los cuales se rechaza la hipótesis nula.

3) La función de potencia es la probabilidad de rechazar la hipótesis nula:

$$\beta_\alpha(\mu) = \mathbb{P}_\mu(R) = \mathbb{P}_\mu\left\{\bar{X} > z_\alpha\frac{\sigma}{\sqrt{n}}\right\} = \mathbb{P}_\mu\left\{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \left(z_\alpha\frac{\sigma}{\sqrt{n}} - \mu\right)\Big/\frac{\sigma}{\sqrt{n}}\right\} = \mathbb{P}\left\{Z > z_\alpha - 10\,\frac{\mu}{\sigma}\right\}$$
$$= 1 - \Phi\left(z_\alpha - 10\,\frac{\mu}{\sigma}\right)$$

donde Z es una v.a. $N(0,1)$ y Φ su función de distribución. Aquí hemos usado que, como $X \sim N(\mu,\sigma)$, tenemos que $\bar{X} \sim N\left(\mu, \frac{\sigma}{10}\right)$. Se concluye que $\beta_\alpha(\mu)$ es estrictamente creciente porque Φ lo es (ya que la densidad normal es estrictamente positiva en todo $\mathbb{R}$).

Si $\alpha = 0.05$, $\mu = 1$ y $\sigma = 2$, obtenemos

$$\beta_{0.05}(1) = \mathbb{P}\left\{Z > z_{0.05} - 10\,\frac{1}{2}\right\} = \mathbb{P}\{Z > 1.645 - 5\} \simeq 1.$$

4) La densidad $\pi(\theta|x_1,\ldots,x_n)$ de la distribución a posteriori de θ es proporcional a

$$f(x_1,\ldots,x_n|\theta)\,\pi(\theta),$$

donde

$$f(x_1,\ldots,x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \mathbb{P}_\theta\{X = x_i\} = \prod_{i=1}^n (1-\theta)^{x_i}\theta = \theta^n(1-\theta)^{\sum_{i=1}^n x_i}$$

es la función de verosimilitud de la muestra. Como

$$f(x_1, \ldots, x_n|\theta)\, \pi(\theta) = \theta^n (1-\theta)^{\sum_{i=1}^n x_i} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \mathbb{I}_{[0,1]}(\theta)$$

y $\Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta)$ es simplemente una constante de proporcionalidad (no depende de θ), tenemos que $\pi(\theta|x_1, \ldots, x_n)$ es proporcional a

$$\theta^{n+\alpha-1}\, (1-\theta)^{(\sum_{i=1}^n x_i)+\beta-1} \mathbb{I}_{[0,1]}(\theta),$$

que corresponde a una beta de parámetros $n+\alpha$ y $(\sum_{i=1}^n x_i) + \beta$. El estimador Bayes de θ es la esperanza de esta distribución a posteriori:

$$T_n(x_1, \ldots, x_n) = \frac{n+\alpha}{n+\alpha+\sum_{i=1}^n x_i + \beta}.$$

Para probar la consistencia c.s. de T_n, reescribimos el estimador así

$$T_n(x_1, \ldots, x_n) = \frac{1+\frac{\alpha}{n}}{1+\bar{x}+\frac{\alpha+\beta}{n}}.$$

Recordemos que $T_n \xrightarrow[n\to\infty]{\text{c.s.}} \theta$ si y sólo si $1 = \mathbb{P}_\theta\{w : T_n(X_1(w), \ldots, X_n(w)) \xrightarrow[n\to\infty]{} \theta\}$, es decir, si la probabilidad del conjunto en el que se da la convergencia de T_n a θ es uno. Por la ley fuerte de los grandes números sabemos que $\bar{X} \xrightarrow[n\to\infty]{\text{c.s.}} \mathbb{E}_\theta(X) = \dfrac{1+\theta}{\theta}$. Además el denominador de T_n siempre será mayor o igual que 1. Por tanto, para todo w salvo en un conjunto de probabilidad 0, se cumple que

$$T_n(X_1(w), \ldots, X_n(w)) = \frac{1+\frac{\alpha}{n}}{1+\bar{X}(w)+\frac{\alpha+\beta}{n}} \xrightarrow[n\to\infty]{} \frac{1}{1+\frac{1+\theta}{\theta}} = \theta,$$

es decir, $T_n \xrightarrow[n\to\infty]{\text{c.s.}} \theta$.

5)

```
xx<-read.table('datos.txt')
x<-xx$V1
y<-xx$V2
boxplot(x,y)
lm(y ~ x)
```