Clase Auxiliar #8 Probabilidades y Estadística Prim '08 FC; HR ①

Teorema central del Límite; Estimadores.

Problema 1 Sea $(X_1, \ldots, X_n)$ una MAS de la distribución Potencia (α, θ), donde la densidad está dada por

$$f(x/\alpha,\theta) = \frac{\alpha+1}{\theta^{\alpha+1}} x^{\alpha} \mathbb{1}_{(0,\theta)}(x) \quad , \quad \begin{array}{l} \alpha \text{ conocido}, \quad \alpha > -1 \\ \theta \text{ desconocido}, \quad \theta > 0 \end{array}$$

a) Muestre que $\mathbb{E}(X^k) = \frac{\alpha+1}{\alpha+k+1}\theta^k$.

Sol

$$\mathbb{E}(X^k) = \int_{\mathbb{R}} x^k f_X(x/\alpha,\theta)\,dx = \int_{\mathbb{R}} x^k \frac{\alpha+1}{\theta^{\alpha+1}} x^{\alpha} \mathbb{1}_{(0,\theta)}(x)\,dx$$

$$= \frac{\alpha+1}{\theta^{\alpha+1}} \int_0^{\theta} x^{\alpha+k}\,dx = \frac{\alpha+1}{\theta^{\alpha+1}} \left. \frac{x^{\alpha+k+1}}{\alpha+k+1} \right|_0^{\theta}$$

$$= \frac{\alpha+1}{\theta^{\alpha+1}} \frac{\theta^{\alpha+k+1}}{\alpha+k+1} = \frac{\alpha+1}{\alpha+k+1}\theta^k.$$

b) Deduzca $\mathbb{E}(X)$, $\mathrm{Var}(X)$.

Sol

$$\mathbb{E}(X) = \frac{\alpha+1}{\alpha+2}\theta$$

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{\alpha+1}{\alpha+3}\theta^2 - \frac{(\alpha+1)^2}{(\alpha+2)^2}\theta^2$$

$$= \theta^2(\alpha+1)\left(\frac{1}{\alpha+3} - \frac{\alpha+1}{(\alpha+2)^2}\right)$$

$$= \theta^2(\alpha+1)\frac{(\alpha+2)^2 - (\alpha+1)(\alpha+3)}{(\alpha+3)(\alpha+2)^2} = \frac{\theta^2(\alpha+1)}{(\alpha+3)(\alpha+2)^2}\left(\cancel{\alpha^2}+\cancel{4\alpha}+4-\cancel{\alpha^2}-\cancel{4\alpha}-3\right)$$

$$= \frac{\theta^2(\alpha+1)}{(\alpha+3)(\alpha+2)^2}$$

c) Obtenga el estimador de los momentos $\hat{\theta}_n$ para θ. Calcule el sesgo y la varianza. ¿Es consistente?

Sol Por el método de los momentos: $m_k = \mu_k$.

$$(k=1) \qquad m_1 = \mu_1 \Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}(X^1) = \frac{\alpha+1}{\alpha+2}\theta$$

$$\Rightarrow \boxed{\hat{\theta}_n = \frac{\alpha+2}{\alpha+1}\,\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\alpha+2}{\alpha+1}\,\overline{X_n}}$$

$$\text{Sesgo}(\hat{\theta}_n) = \mathbb{E}(\hat{\theta}_n) - \theta = \mathbb{E}\left(\frac{\alpha+2}{\alpha+1}\,\overline{X}_n\right) - \theta$$

$$= \frac{\alpha+2}{\alpha+1}\,\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) - \theta$$

$$= \frac{\alpha+2}{\alpha+1}\,\frac{1}{n}\,n\left(\frac{\alpha+1}{\alpha+2}\theta\right) - \theta$$

$$= \theta - \theta = \emptyset.$$

El estimador es insesgado.

$$\mathbb{V}ar(\hat{\theta}_n) = \mathbb{V}ar\left(\frac{\alpha+2}{\alpha+1}\,\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{(\alpha+2)^2}{(\alpha+1)^2}\,\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}ar(X_i)$$

$$= \frac{(\alpha+2)^{\cancel{2}}}{(\alpha+1)^{\cancel{2}}}\,\frac{1}{n}\,\frac{\cancel{(\alpha+1)}\,\theta^2}{(\alpha+3)\cancel{(\alpha+2)}}$$

$$= \frac{1}{n}\cdot\theta^2\,\frac{\alpha+2}{(\alpha+1)(\alpha+3)}$$

¿Es consistente $\hat{\theta}_n$?

Rep: $\hat{\theta}_n$ es consistente si $\text{Sesgo}(\hat{\theta}_n) \xrightarrow{n\to\infty} \emptyset$

$\mathbb{V}ar(\hat{\theta}_n) \xrightarrow{n\to\infty} 0$

En este caso, $\hat{\theta}_n$ es consistente.

1) Obtenga el estimador de máxima verosimilitud $\hat{\theta}_V$ de θ.

Sol La función de verosimilitud del vector $(X_1 - X_n)$ es

$$L(\theta) = f_n(x_1, -, x_n | \theta) \overset{\text{indep}}{=} \prod_{i=1}^{n} f_{X_i}(x_i|\theta) = \prod_{i=1}^{n} \frac{\alpha+1}{\theta^{\alpha+1}} x_i^{\alpha} \mathbb{1}_{(0,\theta)}(x_i)$$

$$= \frac{(\alpha+1)^n}{\theta^{n(\alpha+1)}} \prod_{i=1}^{n} x_i \; \mathbb{1}_{(0,\theta)^n}(\vec{x}) \qquad (*)$$

Notemos que lo anterior no es derivable como función de θ. Entonces vemos que la verosimilitud se maximiza tomando θ lo más chico posible (pues $\theta \mapsto \frac{1}{\theta^k}$ es decreciente), de modo que $(*)$ no se haga nulo por la indicatriz.

Entonces, tomamos el menor θ tal que $\theta \geq x_i$, $\forall i = 1, \dots, n$,

entonces, $\hat{\theta}_V = \max\{X_1, \dots, X_n\}$.

2) Calcule las funciones de distribución y densidad de $\hat{\theta}_V$. Verifique que $\hat{\theta}_V$ tiene distribución de Potencia. Calcule sesgo, varianza y verifique consistencia.

Encuentre c tq $c\hat{\theta}_V$ es insesgado.

Sol Calculamos la distribución de $\hat{\theta}_V$.

$$F_{\hat{\theta}_V}(t) = \mathbb{P}(\hat{\theta}_V \leq t) = \mathbb{P}(\max\{X_1, -, X_n\} \leq t)$$

$$= \mathbb{P}(X_1 \leq t, \dots, X_n \leq t) \overset{(X_i)\text{ iid}}{=} \mathbb{P}(X_1 \leq t)^n$$

$$= \left(F_{X_1}(t)\right)^n$$

con $F_{X_1}(t) = \int_{-\infty}^{t} f_X(x)\,dx = \int_{x=0}^{t} \frac{\alpha+1}{\theta^{\alpha+1}} x^{\alpha}\,dx = \frac{\cancel{\alpha+1}}{\theta^{\alpha+1}} \frac{t^{\alpha+1}}{\cancel{\alpha+1}}$, $t \in (0,\theta)$

$$= \frac{t^{\alpha+1}}{\theta^{\alpha+1}} \mathbb{1}_{(0,\theta)}(t) + \mathbb{1}_{[\theta,\infty)}(t)$$

Entonces, $$F_{\hat{\theta}_v}(t) = \frac{t^{n(\alpha+1)}}{\theta^{n(\alpha+1)}} \mathbb{1}_{(0,\theta)}(t) + \mathbb{1}_{[\theta,\infty)}(t)$$

y $$f_{\hat{\theta}_v}(t) = \frac{n(\alpha+1)}{\theta^{n(\alpha+1)}}\, t^{n(\alpha+1)-1}\, \mathbb{1}_{(0,\theta)}(t)$$

$$= \frac{(n(\alpha+1)-1)+1}{\theta^{(n(\alpha+1)-1)+1}}\, t^{n(\alpha+1)-1}\, \mathbb{1}_{(0,\theta)}(t)$$

tomando $\tilde{\alpha} = n(\alpha+1)-1$, se tiene que $\hat{\theta}_v \sim$ Potencia $(\tilde{\alpha}, \theta)$.

Por lo tanto, $$\mathbb{E}(\hat{\theta}_v) = \frac{\tilde{\alpha}+1}{\tilde{\alpha}+2}\,\theta = \frac{n(\alpha+1)}{n(\alpha+1)+1}\,\theta$$

$$\mathbb{V}ar(\hat{\theta}_v) = \frac{\theta^2(\tilde{\alpha}+1)}{(\tilde{\alpha}+3)(\tilde{\alpha}+2)^2} = \frac{\theta^2\, n(\alpha+1)}{(n(\alpha+1)+2)(n(\alpha+1)+1)^2} \xrightarrow{n\to\infty} 0$$

y $$\text{Sesgo}(\hat{\theta}_v) = \mathbb{E}(\hat{\theta}_v) - \theta = \left(\frac{n(\alpha+1)}{n(\alpha+1)+1} - 1\right)\theta = \frac{n(\alpha+1) - n(\alpha+1) - 1}{n(\alpha+1)+1}\,\theta$$

$$= \frac{-\theta}{n(\alpha+1)+1} \xrightarrow{n\to\infty} 0$$

se concluye que $\hat{\theta}_v$ es consistente.

Ahora, tomando $c = \frac{n(\alpha+1)+1}{n(\alpha+1)}$, tenemos que $(c\,\hat{\theta}_v)$ es insesgado.

$$\mathbb{E}(c\,\hat{\theta}_v) = c\,\mathbb{E}(\hat{\theta}_v) = \frac{n(\alpha+1)+1}{n(\alpha+1)} \cdot \frac{n(\alpha+1)}{n(\alpha+1)+1}\,\theta = \theta$$

Problema 2 Se dice que un estimador $\hat{\theta}$ de un parámetro θ es consistente si

$\lim_{n\to\infty} p\ \hat{\theta} = \theta$ (límite en probabilidad).

a) Demuestre que $ECM(\hat{\theta}) = \mathbb{V}ar(\hat{\theta}) + Sesgo(\hat{\theta})^2$

b) Pruebe que $\hat{\theta}$ es consistente si $\mathbb{V}ar(\hat{\theta}) \xrightarrow{n\to\infty} 0$ y $Sesgo(\hat{\theta}) \xrightarrow{n\to\infty} 0$.

sol a)

$$ECM(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta}-\theta)^2\right] = \mathbb{E}\left[(\hat{\theta}-\mathbb{E}(\hat{\theta})+\mathbb{E}(\hat{\theta})-\theta)^2\right]$$

$$= \mathbb{E}\Big((\hat{\theta}-\mathbb{E}(\hat{\theta}))^2 + (\mathbb{E}(\hat{\theta})-\theta)^2 + 2(\hat{\theta}-\mathbb{E}(\hat{\theta}))(\mathbb{E}(\hat{\theta})-\theta)\Big)$$

$$= \mathbb{E}\left[(\hat{\theta}-\mathbb{E}(\hat{\theta}))^2\right] + (\mathbb{E}(\hat{\theta})-\theta)^2 + 2\,\underbrace{\mathbb{E}(\hat{\theta}-\mathbb{E}(\hat{\theta}))}_{0}\,(\mathbb{E}(\hat{\theta})-\theta)$$

$$= \mathbb{V}ar(\hat{\theta}) + Sesgo(\hat{\theta})^2$$

b) Recordemos la definición de límite en probabilidad.

$$\lim_{n\to\infty} p\ \hat{\theta} = \theta \iff (\forall \varepsilon > 0)\ \mathbb{P}(|\hat{\theta}-\theta| > \varepsilon) \xrightarrow{n\to\infty} 0.$$

Sea $\varepsilon > 0$. Calculemos $\mathbb{P}(|\hat{\theta}-\theta| > \varepsilon)$ (o una cota).

Por la desigualdad de Chebyshev,

$$\mathbb{P}(|\hat{\theta}-\theta| > \varepsilon) \leq \frac{\mathbb{E}(|\hat{\theta}-\theta|^2)}{\varepsilon^2} = \frac{ECM(\hat{\theta})}{\varepsilon^2}$$

$$= \frac{1}{\varepsilon}\left[\mathbb{V}ar(\hat{\theta}) + Sesgo(\hat{\theta})^2\right]$$

Si $\mathbb{V}ar(\hat{\theta}) \xrightarrow{n\to\infty} 0$ y $Sesgo(\hat{\theta}) \xrightarrow{n\to\infty} 0$, entonces al tomar $n \nearrow \infty$, el lado derecho tiende a 0. Por lo tanto, $\lim_{n\to\infty} p\ \hat{\theta} = \theta$, y entonces el estimador es consistente.

Problema 3 Una línea aérea quiere planificar ciertos vuelos de 200 pasajeros. Para mantener los costos en un rango competitivo, la línea está obligada a sobrevender los vuelos. La experiencia muestra que 5% de los pasajeros no llegan o llegan después de cerrado el vuelo. ¿Cuál es el máximo de pasajeros por vuelo que puede vender la compañía, para que la prob. de dejar pasajeros abajo sea menor que 0,1?

sol Sea n = # de pasajeros vendidos, y X_i una v.a. que indica si el pasajero i llega o no al vuelo:

$$X_i = \begin{cases} 1 & \text{si el pasajero llega, con prob } p = 0{,}95 \\ 0 & \text{else, con prob } 1-p = 0{,}05. \end{cases}$$

Entonces, $\sum_{i=1}^{n} X_i$ es el # de pasajeros que llegan al vuelo, y nos piden

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i > 200\right) < 0{,}1.$$

Por el teo. central del límite, $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0,1)$, con $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

donde $\mu = \mathbb{E}(X_1)$, $\sigma^2 = \mathrm{Var}(X_1)$.

En este caso, $\mu = p = 0{,}95$, $\sigma^2 = p(1-p) = 0{,}95 \cdot 0{,}05 = 0{,}0475$.

Entonces, $$\mathbb{P}\left(\sum_{i=1}^{n} X_i > 200\right) = \mathbb{P}\left(\frac{\frac{1}{n}\sum X_i - \mu}{\sigma/\sqrt{n}} > \frac{\frac{200}{n} - \mu}{\sigma/\sqrt{n}}\right)$$

$$\cong \mathbb{P}\left(Z > \frac{200 - n\mu}{\sqrt{n}\,\sigma}\right) \leq 0{,}01 \quad , Z \sim N(0,1)$$

Mirando en una tabla de distribuciones de $N(0,1)$, vemos que (en el caso límite)

$$\mathbb{P}(Z > z) = 0{,}1 \Rightarrow z = 1{,}28$$

Entonces, $z = \dfrac{200 - n\mu}{\sqrt{n}\sigma} \Rightarrow \sqrt{n}\sigma z = 200 - n\mu$

Tomando $N = \sqrt{n}$, $\quad N\sigma z = 200 - N^2\mu$

$$\mu N^2 + \sigma z N - 200 = 0$$

$$\Rightarrow N = \frac{-\sigma z \pm \sqrt{\sigma^2 z^2 + 4 \cdot 200 \cdot \mu}}{2\mu}$$ tomamos el positivo.

$$N = \sqrt{n} = 14,36$$

$$\Rightarrow n = 206,26.$$

Entonces, hay que vender 206 tickets a lo más. //

Obs: Usando la desigualdad de Chebyshev unilateral, imponiendo la desigualdad:
(con $Y = \sum_{i=1}^{n} X_i$, $\nu = \mathbb{E}(Y) = np$, $\sigma^2 = np(1-p)$, $c = 200$; $\alpha = 0,1$)

$$\mathbb{P}(Y > c) = \mathbb{P}(Y - \nu > c - \nu) \leq \frac{\sigma^2}{\sigma^2 + (c-\nu)^2} = \alpha$$

$$\Rightarrow \frac{np(1-p)}{np(1-p) + (c-np)^2} = \alpha$$

$$np(1-p) = \alpha np(1-p) + (c-np)^2\alpha$$

$$0 = -(1-\alpha)p(1-p)n + \alpha c^2 - 2\alpha cpn + \alpha p^2 n^2$$

$$\underbrace{(\alpha p^2)}_{r} n^2 - \underbrace{(2\alpha cp + (1-\alpha)p(1-p))}_{s} n + \underbrace{\alpha c^2}_{t} = 0$$

$$\Rightarrow n = \frac{s \pm \sqrt{s^2 - 4rt}}{2r} = \begin{cases} 200,7 \\ 220,8 \end{cases}$$ $\to$ deja más pasajeros abajo

$\Rightarrow n = 201$ (es peor que la aprox del teo. central del límite).