Auxiliar #4 MA34B-02 Prim'08 Prof Rodrigo ABT

Aux: Hector Olivero ; Victor Riquelme

Eficiencia ; Intervalos de confianza.

Previo: Información de Fisher.

Def: X v.a. con densidad $f(x/\theta)$, con $\theta \in \Theta$ parámetro.

$$I(\theta) = \mathbb{E}_X\left[\left(\frac{\partial \ln f(x/\theta)}{\partial \theta}\right)^2\right]$$

Props: a) $I(\theta) = \mathbb{E}_X\left[-\partial^2_\theta \ln f(x/\theta)\right]$

b) Si $\mathbb{E}\left[\partial_\theta \ln f(x/\theta)\right] = 0$, entonces $I(\theta) = \mathbb{V}ar\left[\partial_\theta \ln f(x/\theta)\right]$

c) Si $\vec{X} = (X_1, \ldots, X_n)$ m.a.s. de X con densidad $f(x/\theta)$. Entonces definimos

$$I_n(\theta) \equiv \mathbb{E}\left[-\partial^2_\theta \ln f(\vec{x}/\theta)\right]$$

$$\Rightarrow I_n(\theta) = n\, I(\theta).$$

Cota de Cramer-Rao:

Prop: Si $\hat{\theta}$ es un estimador insesgado de θ, entonces (si $(X_1, \ldots, X_n)$ m.a.s. de $X \sim f(x/\theta)$).

$$\mathbb{V}ar(\hat{\theta}) \geq \frac{1}{I_n(\theta)} = \frac{1}{n\, I(\theta)}$$

obs: Si $\hat{\theta}$ es tq alcanza la cota, entonces $\hat{\theta}$ es un estimador eficiente e insesgado.

Problema 1 Durante 200 intervalos de 15 min. se obtuvo k llegadas de personas en cada intervalo, a una cola de una caja de supermercado. La siguiente tabla indica el número de intervalos que registraron k personas.

k	0	1	2	3	4	5	6	7	>7
# intervalos	10	22	49	54	30	16	15	4	0

(a) Suponiendo que el # de personas que llega a la caja durante 15 minutos se distribuye Poisson(λ), encuentre el estimador de máxima verosimilitud para λ dada una MAS. $(X_i)_{i=1}^n$

Sol La fn de verosimilitud viene dada por

$$L(\vec{x},\theta) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i}$$

$$\Rightarrow l(\vec{x},\theta) = -n\lambda \cancel{\ln(e)} + \sum_{i=1}^n x_i \ln(\lambda) - \ln \prod_{i=1}^n x_i$$

$$\frac{\partial}{\partial\theta} l(\vec{x},\theta) = -n + \frac{1}{\lambda}\sum_{i=1}^n x_i = 0$$

$$\Rightarrow \boxed{\hat{\lambda} = \frac{1}{n}\sum_{i=1}^n x_i}$$

(b) Calcule la información de Fisher asociada a $(X_i)_{i=1}^n$

Sol Ya tenemos que $\partial_\theta l(\vec{x},\theta) = -n + \frac{1}{\lambda}\sum_{i=1}^n x_i$

$$\Rightarrow \partial_\theta^2 l(\vec{x},\theta) = -\frac{1}{\lambda^2}\sum_{i=1}^n x_i$$

$$\Rightarrow \mathbb{E}_X\left(-\partial_\theta^2 l(\vec{x},\theta)\right) = \mathbb{E}_X\left(\frac{1}{\lambda^2}\sum_{i=1}^n X_i\right) = \frac{1}{\lambda^2}\sum_{i=1}^n \mathbb{E}(X_i)$$

$$= \frac{1}{\lambda^2} n\lambda = \frac{n}{\lambda}$$

(c) Calcule la varianza del estimador $\hat{\lambda}$, despues de verificar que es insesgado.

Sol. ⟩ Insesgado: $\mathbb{E}(\hat{\lambda}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i)$

$= \frac{1}{n} n \lambda = \lambda$ //

⟩ $\mathbb{V}ar(\hat{\lambda}) = \mathbb{V}ar\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}ar(X_i) = \frac{1}{n^2}\cdot n\lambda = \frac{\lambda}{n} \xrightarrow{n\to\infty} 0$

Entonces, $\hat{\lambda}$ es consistente.

(d) ¿Alcanza $\hat{\lambda}$ la cota de Cramer-Rao? Concluya.

Sol. La cota de CR dice que $\mathbb{V}ar(\hat{\lambda}) \geq \frac{1}{I_n(\lambda)} = \frac{\lambda}{n}$

En este caso, la varianza es igual a la cota. Entonces el estimador es eficiente, insesgado, consistente, de varianza mínima.

(e) Para la MAS dada al principio, de una estimación de λ.

Sol. En este caso, X va a ser la v.a. que indica cuanta gente llega durante 15 mins a la caja. Así, nuestra muestra es de tamaño $n=200$, de las cuales 10 veces se registraron 0 personas, 22 veces se registro 1 persona, y así sucesivamente.

Entonces, $\hat{\lambda} = \frac{1}{200}\left(0\cdot 10 + 1\cdot 22 + 2\cdot 49 + 3\cdot 54 + 4\cdot 30 + 5\cdot 16 + 6\cdot 15 + 7\cdot 4 + \underset{\text{rep. de } (7,\infty)}{8\cdot 0}\right)$

$= 3.$

Problema 2 Considere una MAS $T=(T_1,\dots,T_n)$ del tiempo de falla en 10 horas de n componentes electronicas. Suponga que el tiempo de falla de este tipo de componentes sigue una distribución Weibull con función de densidad dada por

$$f(t/\lambda) = \frac{1}{\lambda} t e^{-t^2/2\lambda}, \quad t>0 \ ; \ \lambda>0 \text{ fijo}$$

(a) Calcule el estimador de max. verosimilitud de λ.

$$L(\underline{t},\lambda) = \frac{1}{\lambda^n}\prod_{i=1}^{n} t_i \, e^{-\frac{1}{2\lambda}\sum_{i=1}^{n} t_i^2}$$

$$l(\underline{t},\lambda) = \ln L(\underline{t},\lambda) = -n\ln(\lambda) + \sum_{i=1}^{n}\ln t_i - \frac{1}{2\lambda}\sum_{i=1}^{n} t_i^2$$

$$\partial_\lambda l(\underline{t},\lambda) = -\frac{n}{\lambda} + \frac{1}{2\lambda^2}\sum_{i=1}^{n} t_i^2 = 0$$

$$\Rightarrow \cancel{\lambda}\sum_{i=1}^{n} t_i^2 = n\cdot 2\lambda^{\cancel{2}} \Rightarrow \boxed{\hat{\lambda} = \frac{1}{2n}\sum_{i=1}^{n} t_i^2}$$

(b) Muestre que $\hat{\lambda}$ es consistente

$$\mathbb{E}(\hat{\lambda}) = \frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}(t_i^2) = \frac{1}{2}\mathbb{E}(T_1^2)$$

Primero, por el teorema del cambio de variables, haciendo $Y=T^2 \Rightarrow T=\sqrt{Y} \rightarrow \frac{\partial t}{\partial y} = \frac{1}{2}\frac{1}{\sqrt{y}}$

$$f_Y(y) = f_T(t(y))\left|\frac{\partial t}{\partial y}(y)\right| = \frac{1}{\lambda}\cancel{\sqrt{y}}\, e^{-y/2\lambda}\cdot\frac{1}{2\cancel{\sqrt{y}}}$$

$$= \frac{1}{2\lambda}e^{-\frac{y}{2\lambda}}$$

$$\Rightarrow Y \sim Exp\left(\frac{1}{2\lambda}\right)$$

Así, $\mathbb{E}(T^2) = 2\lambda$

$$\Rightarrow \mathbb{E}(\hat{\lambda}) = \frac{1}{2}\cdot 2\lambda = \lambda \qquad \text{insesgado}$$

$$\mathbb{V}ar(\hat{\lambda}) = \mathbb{V}ar\left(\frac{\sum^{n} T_i^2}{2n}\right) = \frac{1}{4n^2}\, n\, \mathbb{V}ar(T^2) = \frac{1}{4n}\,(4\lambda^2)$$

$\mathbb{V}ar(Exp\,\lambda) = \frac{1}{\lambda^2}$

$$= \frac{\lambda^2}{n} \rightarrow 0$$

El estimador es consistente, y asintóticamente eficiente.

(c) Verifique si $\hat{\lambda}$ es eficiente para λ.

sol Calculemos la información de Fisher

$$\partial_\lambda l(\underline{t},\lambda) = \frac{-n}{\lambda} + \frac{1}{2\lambda^2}\sum_{i=1}^{n} t_i^2$$

$$\partial^2_\lambda l(\underline{t},\lambda) = \frac{n}{\lambda^2} - \frac{1}{\lambda^3}\sum_{i=1}^{n} t_i^2$$

$$-\mathbb{E}_T\left(\partial^2_\lambda l(\underline{t},\lambda)\right) = \frac{-n}{\lambda^2} + \frac{1}{\lambda^3}\, n\, 2\lambda = -\frac{n}{\lambda^2} + \frac{2n}{\lambda^2} = \frac{n}{\lambda^2}$$

Así, $CR(\lambda) = \frac{\lambda^2}{n}$ es la cota de Cramer Rao.

y se tiene que $\mathbb{V}ar(\hat{\lambda}) = CR(\lambda)$. El estimador es eficiente.

(d) Con los datos obtenidos $n=10$, $\sum t_i = 8{,}2$, $\sum t_i^2 = 10$, calcule la estimación de máxima verosimilitud de la probabilidad que un componente de este tipo dure más de 100 horas.

sol
$$\mathbb{P}(T \geq 100) = \int_{100}^{\infty} \frac{1}{\lambda}\, t\, e^{-\frac{t^2}{2\lambda}}\, dt = -e^{-t^2/2\lambda}\Big|_{100}^{\infty} = e^{-\frac{100^2}{2\lambda}}$$

Ahora, se reemplaza λ por $\hat{\lambda}$ (propiedad de los EMV: $g(\hat{\theta}) = \widehat{g(\theta)}_{MV}$

$$\widehat{\mathbb{P}(T \geq 100)}_{MV} = e^{-\frac{100^2 \cdot 2n}{\sum_{i=1}^{n} t_i^2}} = e^{-\frac{100^2 \cdot 20}{10}}$$

Problema 3 Una máquina llena paquetes de cafe de 500 gr. La maquina hace el llenado según una distribución $N(\mu, \sigma^2)$, $\sigma^2 = 400\ gr^2$. Se sospecha que la máquina está fuera de control. Se selecciona una m.a.s. de 16 paquetes, observandose $\overline{X}_n = 492$ gr.

(a) De un intervalo de confianza para la media μ, de nivel de confianza 95%.

Sol Damos el intervalo de confianza de largo mínimo.

Como $X_i \sim N(\mu, \sigma^2)$, $\overline{X}_n \sim N(\mu, \sigma^2/n)$. Entonces, $\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

Buscamos un intervalo (b,a) tal que $\mathbb{P}(\mu \in (b,a)) = 95\% = 0{,}95$.

$$\mathbb{P}(b \leq \mu \leq a) = \mathbb{P}(\overline{X}_n - a \leq \overline{X}_n - \mu \leq \overline{X}_n - b)$$

$$= \mathbb{P}\Big(\underbrace{\frac{\overline{X}_n - a}{\sigma/\sqrt{n}}}_{\alpha} \leq \underbrace{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}_{Z} \leq \underbrace{\frac{\overline{X}_n - b}{\sigma/\sqrt{n}}}_{\beta}\Big)$$

Se buscan números α, β tq $\mathbb{P}(\alpha \leq Z \leq \beta) = 0{,}95$, con $Z \sim N(0,1)$. Tomamos intervalo simetrico, por ser de largo mínimo. (así, $\alpha = -\beta$)

$$\Rightarrow \begin{cases} \beta = 1{,}96 & \text{es tal que } \mathbb{P}(Z > \beta) = 0{,}025 \\ \alpha = -1{,}96 & \text{es tal que } \mathbb{P}(Z < \alpha) = 0{,}025 \end{cases}$$

Así, $\overline{X}_n - b = \beta \dfrac{\sigma}{\sqrt{n}} \Rightarrow b = \overline{X}_n - \beta \dfrac{\sigma}{\sqrt{n}} = 492 - 1{,}96 \cdot \dfrac{20}{\sqrt{16}} = 482{,}2$

$\overline{X}_n - a = -\beta \dfrac{\sigma}{\sqrt{n}} \Rightarrow a = \overline{X}_n + \beta \dfrac{\sigma}{\sqrt{n}} = 492 + 1{,}96 \cdot \dfrac{20}{\sqrt{16}} = 501{,}8$

(b) ¿Que puede concluir?

Que la máquina con un 95% de confianza funciona dentro de lo normal ($\mu = 500$)