Probabilidades y Estadística (C)

Clase Práctica 19

Matías Data

Departamento de Matemática, FCEyN, Universidad de Buenos Aires

10/11/2020

Introducción

- Hoy seguimos con estimación puntual, veremos sesgo, varianza, error cuadrático medio y consistencia de un estimador.

Definición

Dada una m.a. $X_1, \ldots, X_n \sim F(\theta)$ y dado un estimador $\widehat{\theta} = \delta(X_1, \ldots, X_n)$ de θ, se dice que es *insesgado* si $\mathbb{E}_\theta\left(\widehat{\theta}\right) = \theta$. El *sesgo* del estimador se define como $b\left(\widehat{\theta}\right) = \mathbb{E}_\theta\left(\widehat{\theta}\right) - \theta$.

Definición

Un estimador $\widehat{\theta}$ de θ se dice *asintóticamente insesgado* si

$$\mathbb{E}_\theta\left(\widehat{\theta}_n\right) \underset{n \to +\infty}{\longrightarrow} \theta$$

Es decir, cuando el tamaño de la muestra tiende a infinito el sesgo se hace cero.

Introducción

Definición

Dado un estimador $\widehat{\theta}$ de θ, se define la varianza del estimador por

$$\mathbb{V}_\theta\left(\widehat{\theta}\right) = Var\left(\widehat{\theta}\right)$$

Es decir, $\widehat{\theta}$ es una variable aleatoria, por lo tanto le calculamos su varianza (que depende de θ), esa es la varianza del estimador.

Definición

Un estimador $\widehat{\theta}$ de θ, se define su error cuadrático medio como:

$$ECM_\theta\left(\widehat{\theta}\right) = \mathbb{E}_\theta\left((\widehat{\theta} - \theta)^2\right)$$

Es decir, es la media (poblacional) de la distancia al cuadrado al parámetro θ.

Introducción

Proposición

Vale:

$$ECM_\theta\left(\widehat{\theta}\right) = b\left(\widehat{\theta}\right)^2 + \mathbb{V}_\theta\left(\widehat{\theta}\right)$$

Es decir, el error cuadrático medio se descompone como la suma del sesgo al cuadrado más la varianza.

Definición

Un estimador $\widehat{\theta}$ de θ se dice *consistente* si:

$$\widehat{\theta}_n \xrightarrow[n\to+\infty]{P} \theta$$

Es decir, dado $\varepsilon > 0$ tenemos que:

$$P\left(\left|\widehat{\theta}_n - \theta\right| > \varepsilon\right) \xrightarrow[n\to+\infty]{} 0$$

Introducción

Proposición

Dado un estimador $\widehat{\theta}$ de θ, si $ECM_\theta\left(\widehat{\theta}_n\right) \underset{n\to+\infty}{\longrightarrow} 0$ entonces el estimador es consistente.

Corolario

Dado un estimador $\widehat{\theta}$ de θ, si:

$$\mathbb{E}_\theta\left(\widehat{\theta}_n\right) \underset{n\to+\infty}{\longrightarrow} \theta$$

$$\mathbb{V}_\theta\left(\widehat{\theta}_n\right) \underset{n\to+\infty}{\longrightarrow} 0$$

entonces el estimador es consistente.

Introducción

Proposición

Sean $X_1, X_2, \ldots$ e $Y_1, Y_2, \ldots$ dos sucesiones de variables aleatorias tales que $X_n \xrightarrow{P} a$ e $Y_n \xrightarrow{P} b$. Sea $g : \mathbb{R}^2 \to \mathbb{R}$ continua en (a, b). Entonces $g(X_n, Y_n) \xrightarrow{P} g(a, b)$.

En particular:

- $X_n + Y_n \xrightarrow{P} a + b$.
- $X_n Y_n \xrightarrow{P} ab$.
- $X_n / Y_n \xrightarrow{P} a/b$ si $b \neq 0$.

Ejercicio 1

Ejercicio

Sean $X_1, \ldots, X_n$ una m.a. $N(0, \sigma^2)$.

A) Hallar un estimador de σ^2 de la forma

$$\widehat{\sigma^2} = \alpha \left(\sum_{i=1}^{n} X_i^2 \right)$$

con α una constante de forma tal que minimize el ECM.

B) Deducir que es consistente.

C) Hallar el EMV de σ^2.

D) Cuál estimador prefiere?

Ejercicio 1

a) Calculemos el ECM para poder minimizarlo.

- Para esto primero calculemos la esperanza y varianza del estimador.

$$\mathbb{E}_{\sigma^2}\left(\widehat{\sigma^2}\right) = \mathbb{E}_{\sigma^2}\left(\alpha\left(\sum_{i=1}^{n} X_i^2\right)\right) = \alpha n \sigma^2$$

  ya que $\mathbb{E}_{\sigma^2}(X_i) = \sigma^2$.

- Ahora para la varianza recordamos que si $Z \sim N(0,1)$ entonces $Z^2 \sim \Gamma(1/2, 1/2)$, luego:

$$\mathbb{V}_{\sigma^2}\left(\widehat{\sigma^2}\right) = \mathbb{V}_{\sigma^2}\left(\alpha\left(\sum_{i=1}^{n} X_i^2\right)\right) = \alpha^2\sigma^4 n\, \mathbb{V}_{\sigma^2}\left(X_i^2/\sigma^2\right) =$$
$$= \alpha^2\sigma^4 2n$$

Ejercicio 1

- Luego:

$$ECM_{\sigma^2}\left(\widehat{\sigma^2}\right) = \mathbb{E}_{\sigma^2}\left(\widehat{\sigma^2} - \sigma^2\right)^2 + \mathbb{V}_{\sigma^2}\left(\widehat{\sigma^2}\right)$$
$$= \left(\sigma^2(\alpha n - 1)\right)^2 + \alpha^2\sigma^4 2n = \sigma^4\left((\alpha n - 1)^2 + \alpha^2 2n\right)$$

- Para minimizar esta expresión derivamos e igualamos a cero, luego:

$$0 = 2n(\alpha n - 1) + 4\alpha n$$

- Luego $\alpha = \frac{1}{n+2}$ realiza el mínimo.

Ejercicio 1

b) Calculemos su ECM para ver la consistencia:

$$ECM_{\sigma^2}\left(\widehat{\sigma^2}\right) = \sigma^4\left((\frac{n}{n+2}-1)^2 + \frac{1}{(n+2)^2}2n\right) =$$
$$= \sigma^4\left(\frac{4}{(n+2)^2} + \frac{2n}{(n+2)^2}\right) = \sigma^4\frac{2}{n+2}$$

- Luego $\widehat{\sigma^2}$ es consistente ya que el ECM converge a cero como $\sim 1/n$.
- Veamos otra forma de ver la consistencia, en lugar de calcular el ECM usando la LGN.
- Sabemos que:
$$\frac{\sum_{i=1}^n X_i^2}{n} \xrightarrow{P} \sigma^2$$
por la LGN.
- Luego:

$$\widehat{\sigma^2}_n = \frac{\sum_{i=1}^n X_i^2}{n+2} = \left(\frac{n}{n+2}\right)\left(\frac{\sum_{i=1}^n X_i^2}{n}\right) \xrightarrow{P} \sigma^2$$

Ejercicio 1

c) Calculemos el EMV de σ^2, dada una muestra $x_1, \ldots, x_n$:

$$\mathcal{L}(\sigma^2 \,|\, \underline{x}) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-x_i^2}{2\sigma^2}}$$

- Luego tomando logaritmos:

$$\ell(\sigma^2 \,|\, \underline{x}) = \sum_{i=1}^{n} -\log(\sigma) - \log(\sqrt{2\pi}) - \frac{x_i^2}{2\sigma^2}$$

- Derivando e igualando a cero tenemos que:

$$0 = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n} x_i^2}{\sigma^3}$$

- Luego $\widehat{\sigma^2_{MV}} = \frac{\sum_{i=1}^{n} X_i^2}{n}$.

Ejercicio 1

D) Para comparar los dos estimadores, calculemos el ECM del estimador de máxima verosimilitud.

- Sabemos que es insesgado, luego:

$$ECM_{\sigma^2}(\widehat{\sigma^2_{MV}}) = \mathbb{V}_{\sigma^2}\left(\frac{\sum_{i=1}^n X_i^2}{n}\right) = \frac{\sigma^4}{n}\mathbb{V}_{\sigma^2}\left(\frac{X_1}{\sigma^2}\right) = \sigma^4\frac{2}{n}$$

- Como vemos el ECM es más grande, por lo tanto por el Principio de estimación de menor error cuadrático medio nos quedamos con el estimador del item A).

Ejercicio 2

Ejercicio

Sean $X_1, \ldots, X_n$ una m.a. cuya distribución está dada por:

$$f(x, \theta) = \frac{\theta}{x^2} \mathbb{1}_{[\theta, +\infty)}(x)$$

para $\theta > 0$.

a) Hallar $\widehat{\theta}_{MV}$, el estimador de máxima verosimilitud de θ.

b) Probar que es asintóticamente insesgado.

c) Calcular su varianza y su ECM.

d) Deducir que es consistente.

Ejercicio 2

a) Calculamos:

$$\mathcal{L}(\theta \mid \underline{x}) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \frac{\theta}{x_i^2} \mathbb{1}_{[\theta,+\infty)}(x_i) =$$
$$= C\theta^n \mathbb{1}_{(0,\min x_i]}(\theta)$$

ya que $\prod_{i=1}^{n} \mathbb{1}_{[\theta,+\infty)}(x_i) = 1$ si y solo si $\theta \leq x_i$ para todo i, si y solo si $\theta \leq \min x_i$.

- Luego el máximo se alcanza cuando $\widehat{\theta} = \min x_i$, ya que es una función creciente en ese intervalo.
- Por lo tanto $\widehat{\theta}_{MV} = \min X_i$.

Ejercicio 2

B) Para calcular la esperanza del estimador, primero calculamos su distribución.

$$P(\widehat{\theta}_n \geq x) = P(\min X_i \geq x) = P(X_1 \geq x)^n = \left(\frac{\theta}{x}\right)^n$$

- Luego $F_{\widehat{\theta}_n}(x) = 1 - \left(\frac{\theta}{x}\right)^n$ y por lo tanto:

$$f_{\widehat{\theta}_n} = n\frac{\theta^n}{x^{n+1}}$$

- Ahora calculemos $\mathbb{E}_\theta\left(\widehat{\theta}_n\right)$.

$$\mathbb{E}_\theta\left(\widehat{\theta}_n\right) = \int_\theta^{+\infty} x\frac{n\theta^n}{x^{n+1}}\,dx = n\theta^n\left(\frac{-x^{-(n-1)}}{n-1}\right)\Bigg|_\theta^{+\infty} =$$
$$= \frac{n\theta}{n-1}$$

- Por lo tanto, $\widehat{\theta}_{MV}$ es asintóticamente insesgado.

Ejercicio 2

c) Para calcular la varianza primero calculamos el segundo momento:

$$\mathbb{E}_\theta\left(\widehat{\theta}_n^2\right) = \int_\theta^{+\infty} x^2 \frac{n\theta^n}{x^{n+1}}\, dx = n\theta^n \left(\frac{-x^{-(n-2)}}{n-2}\right)\Bigg|_\theta^{+\infty} =$$

$$= \frac{n\theta^2}{n-2}$$

- Luego:

$$\mathbb{V}_\theta\left(\widehat{\theta}_n\right) = \frac{n\theta^2}{n-2} - \left(\frac{n\theta}{n-1}\right)^2 = \frac{n\theta^2}{(n-1)^2(n-2)}$$

- Por lo tanto:

$$ECM_\theta\left(\widehat{\theta}_n\right) = \mathbb{E}_\theta\left(\widehat{\theta}_n - \theta\right)^2 + \mathbb{V}_\theta\left(\widehat{\theta}_n\right) = \frac{2\theta^2}{(n-1)(n-2)}$$

d) Como el ECM tiende a cero (como $\sim 1/n^2$) el estimador es consistente.