Clase 7

viernes, 31 de agosto de 2018 11:33

Consistencia

Definición 46: Una sucesión de estimadores $W_n = W_n(X_1, \dots, X_n)$ es una sucesión de estimadores consistentes del parámetro θ si para todo $\varepsilon > 0$ y para todo $\theta \in \Theta$

$$\lim_{n \to \infty} P_\theta(|W_n - \theta| < \varepsilon) = 1$$

es equivalente a decir $\lim_{n \to \infty} P_\theta(|W_n - \theta| \geq \varepsilon) = 0$

Ejemplo Sea $X_1, \dots, X_n$ iid con distribución $N(\theta, 1)$ y consideremos la sucesión

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$\bar{X} \sim N(\theta, 1/n)$ como ya sabemos luego

$$P_\theta(|\bar{X}_n - \theta| < \varepsilon) = \int_{\theta-\varepsilon}^{\theta+\varepsilon} \left(\frac{n}{2\pi}\right)^{1/2} e^{-n/2 (z-\theta)^2} dz =$$

cv $y = z - \theta$

$$= \int_{-\varepsilon}^{\varepsilon} \left(\frac{n}{2\pi}\right)^{1/2} e^{-n/2\, y^2} dy \underset{\substack{\downarrow \\ \text{cv } t = y\sqrt{n}}}{=} \int_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} \left(\frac{1}{2\pi}\right)^{1/2} e^{-(1/2)t^2} dt$$

$$= P(-\varepsilon\sqrt{n} < Z < \varepsilon\sqrt{n}) \xrightarrow[n \to \infty]{} 1 \qquad (Z \sim N(0,1))$$

luego $\bar{X}_n$ es un estimador consistente de θ

En general no es necesario hacer una demostración en base al cálculo de la probabilidad

Si usamos la desigualdad de Chebyshev tenemos

$$P_\theta(|W_n-\theta|\geq \varepsilon) \leq \frac{E_\theta[(W_n-\theta)^2]}{\varepsilon^2}$$

así que si se cumple para todo $\theta \in \Theta$ que

$$\lim_{n\to\infty} E_\theta[(W_n-\theta)^2] = 0$$

entonces la sucesión de estimadores es consistente.

Además ya vimos que:

$$E_\theta[(W_n-\theta)^2] = \text{Var}_\theta W_n + [\text{Sesgo}_\theta W_n]^2$$

A partir de esto obtenemos

Teorema 47: Si W_n una sucesión de estimadores de un parámetro θ que satisface:

i) $\lim_{n\to\infty} \text{Var}_\theta(W_n) = 0$

ii) $\lim_{n\to\infty} \text{Sesgo}_\theta(W_n) = 0$

para todo $\theta \in \Theta$ entonces W_n es una sucesión consistente de estimadores de θ

Ejemplo: Consideremos $X_1, \dots, X_n$ variables aleatorias iid con $E(X_i) = \theta$ y $Var(X_i) = \sigma^2 < \infty$ (no necesariamente normales) Tenemos

$$E_\theta(\bar{X}_n) = \theta \quad y \quad Var_\theta(\bar{X}_n) = \frac{\sigma^2}{n}$$

luego $\lim\limits_{n\to\infty} Var_\theta(\bar{X}_n) = \lim\limits_{n\to\infty} \frac{\sigma^2}{n} = 0$ y

el sesgo $(\bar{X}_n) = E_\theta(\bar{X}_n) - \theta = 0$ entonces

$\bar{X}_n$ es un estimador consistente para θ

Teorema 48: Sea W_n un estimador consistente de θ

Sean $\{a_n\}$ y $\{b_n\}$ sucesiones de constantes que cumplen

i) $\lim\limits_{n\to\infty} a_n = 1$

ii) $\lim\limits_{n\to\infty} b_n = 0$

Entonces la sucesión $U_n = a_n W_n + b_n$ es una sucesión consistente de estimadores de θ

Teorema 49: Sean $X_1, \dots, X_n$ iid $f(x/\theta)$ y sea $L(\theta/x) = \prod\limits_{i=1}^{n} f(x_i|\theta)$ la función de verosimilitud.

Se cumple que i) las densidades $f(x/\theta)$ son identificables o sea si $\theta \neq \theta'$ entonces $f(x/\theta) \neq f(x/\theta')$

ii) $f(x/\theta)$ tienen el mismo soporte y $f(x/\theta)$ es diferenciable en Θ

iii) el espacio de parámetros Θ contiene un abierto ω en donde el verdadero valor del parámetro $\theta_0 \in \mathring{\omega}$

Sea $\hat{\theta}$ el estimador máximo verosímil de θ y sea $\tau(\theta)$ una función continua de θ.

Entonces para todo $\varepsilon > 0$ y para todo $\theta \in \Theta$

$$\lim_{n\to\infty} P_\theta\left(|\tau(\hat{\theta}) - \tau(\theta)| \geq \varepsilon\right) = 0$$

o sea $\tau(\hat{\theta})$ es un estimador consistente de $\tau(\theta)$

Dem: Ver el libro de Stuart, Ord y Arnold (1999)

Eficiencia

Esta propiedad concierne a la varianza asintótica del estimador. Dado un estimador T_n suele calcularse primeramente la $\operatorname{Var} T_n$ para una muestra finita y luego considerar una constante normalizadora k_n y evaluar $\lim_{n\to\infty} k_n \operatorname{Var} T_n$

En muchos casos $\operatorname{Var} T_n \to 0$ $(n \to \infty)$ por lo que necesitamos un factor k_n para fijar el límite.

Definición 50: Para un estimador T_n, si $\lim_{n\to\infty} k_n \operatorname{Var}(T_n) = \tau^2 < \infty$ donde $\{k_n\}$ una sucesión de constantes entonces τ^2 se llama varianza límite o límite de las varianzas.

Ejemplo: Sea $\bar{X}_n$ la media de n variables aleatorias normales iid con $E(X) = \mu$ y $Var(X) = \sigma^2$
Si tomamos $k_n = n$ tenemos que

$$\lim_n n \, Var(\bar{X}_n) = \lim_n n \frac{\sigma^2}{n} = \sigma^2$$

Definición 51: Consideremos un estimador T_n y supongamos que $k_n(T_n - \tau(\theta)) \to N(0, \sigma^2)$ en distribución. El parámetro σ^2 se llama varianza asintótica o varianza de la distribución límite de T_n.

Definición 52: Una sucesión de estimadores W_n es asintóticamente eficiente del parámetro $\tau(\theta)$ si

$$\sqrt{n}\,[W_n - \tau(\theta)] \to N(0, \upsilon(\theta)) \text{ en distribución}$$

y

$$\upsilon(\theta) = \frac{[\tau'(\theta)]^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta} \log f(x/\theta)\right)^2\right)}$$

o sea la varianza asintótica de W_n alcanza la cota de Cramér-Rao

Teorema 53: (Varianza asintótica del estimador máximo verosímil)
Sea $X_1, X_2, \ldots$ v.a iid $f(x/\theta)$. Sea $\hat{\theta}$ el estimador

máximo verosímil de Θ y sea $\tau(\Theta)$ una función continua de Θ. Si se cumplen las condiciones del teorema 49 y además se cumple

a) para todo $x \in \mathcal{X}$ la densidad $f(x/\theta)$ es tres veces diferenciable con respecto a Θ y la derivada tercera es continua en Θ.

b) $\int f(x/\theta)\,dx$ es tres veces diferenciable bajo el signo de la integral

c) Para todo $\Theta_0 \in \Theta$, existe un número positivo c y una función $M(x)$ (pueden depender ambos de Θ_0) tal que

$$\left[\frac{\partial^3}{\partial\theta^3} \log f(x/\theta)\right] \leq M(x) \quad \forall x \in \mathcal{X} \quad \theta_0 - c < \theta < \theta_0 + c$$

con $E_{\theta_0}\left[M(\mathbf{X})\right] < \infty$

entonces $\sqrt{n}\left[\tau(\hat{\theta}) - \tau(\theta)\right] \longrightarrow \mathcal{N}(0, v(\theta))$

donde $v(\theta)$ es la cota de Cramér-Rao.

O sea $\tau(\hat{\theta})$ es un estimador consistente y asintóticamente eficiente de $\tau(\theta)$

Dem: Veremos solo el caso de Θ (omitimos el caso $\tau(\theta)$)

Notamos por $l(\theta|x) = \sum \log f(x_i|\theta)$

$$l'(\theta|x) = l'(\theta_0|x) + (\theta - \theta_0)\, l''(\theta_0|x) + \dots$$

omitimos los términos de mayor orden

Si $\hat{\theta}$ es el estimador máximo verosímil entonces $\ell'(\hat{\theta}|x)=0$ entonces

$$\hat{\theta}-\theta_0 = -\frac{\ell'(\theta_0|x)}{\ell''(\theta_0|x)}$$

de donde

$$\sqrt{n}(\hat{\theta}-\theta_0) = \sqrt{n}\,\frac{-\ell'(\theta_0|x)}{\ell''(\theta_0|x)} = \frac{1/\sqrt{n}\,\ell'(\theta_0|x)}{-1/n\,\ell''(\theta_0|x)}$$

Si notamos $I(\theta_0) = E[\ell'(\theta_0|X)]^2 = 1/\nu(\theta)$ la información de una observación

Por el teorema central del límite

$$\frac{1}{\sqrt{n}}\ell'(\theta_0|X) = \sqrt{n}\left[\frac{1}{n}\sum_i \frac{\frac{d}{d\theta}f(x_i|\theta)}{f(x_i|\theta)}\right] = \sqrt{n}\left[\frac{1}{n}\sum_i w_i\right]$$

$$w_i = \frac{\frac{d}{d\theta}f(x_i/\theta)}{f(x_i|\theta)} \quad y \quad E(w_i)=0 \,, \; Var(w_i) = I(\theta_0)$$

luego por el teorema central del límite

$$\frac{1}{\sqrt{n}}\ell'(\theta_0|X) \to N(0, I(\theta_0))$$

Además $-\frac{1}{n}\ell''(\theta_0|x) = \frac{1}{n}\sum w_i^2 - \frac{1}{n}\sum_i \frac{\frac{d^2}{d\theta^2}f(x_i|\theta)}{\quad}$

Además $-\frac{1}{n} l''(\theta_0|x) = \frac{1}{n}\sum_i W_i^2 - \frac{1}{n}\sum_i \frac{\frac{d^2}{d\theta^2} f(x_i|\theta)}{f^2(x_i|\theta)}$

$$E\left(\frac{1}{n}\sum_i W_i^2\right) = E(W_i^2) = I(\theta_0)$$

y $E\left(\frac{1}{n}\sum_i \frac{\frac{d^2}{d\theta^2} f(x_i|\theta)}{f^2(x_i|\theta)}\right) = 0$

entonces aplicando la ley débil de los grandes números

$$-\frac{1}{n} l''(\theta_0|x) \longrightarrow I(\theta_0) \text{ en probabilidad}$$

o sea

$$\sqrt{n}(\hat{\theta}-\theta_0) = \frac{\overbrace{1/\sqrt{n}\, l'(\theta_0|x)}^{\to N(0, I(\theta_0))}}{\underbrace{1/n\, l''(\theta_0|x)}_{\xrightarrow{P} I(\theta_0)}} \longrightarrow N\left(0, 1/I(\theta_0)\right)$$

Observación: Decir que un estimador es eficiente y consistente es redundante pues la normalidad asintótica implica consistencia. Supongamos que

$$\sqrt{n}\,\frac{W_n-\mu}{\sigma} \longrightarrow Z \sim N(0,1) \text{ en distribución}$$

Aplicando Slutsky

$$W_n-\mu = \left(\frac{\sigma}{\sqrt{n}}\right)\left(\sqrt{n}\,\frac{W_n-\mu}{\sigma}\right) \longrightarrow \lim_{n\to\infty}\left(\frac{\sigma}{\sqrt{n}}\right) Z = 0$$

o sea $W_n-\mu \longrightarrow 0$ en distribución.

La convergencia en distribución a un punto implica la convergencia en probabilidad a ese punto luego W_n es un estimador consistente de μ

Método Delta

Sea $T_1, \dots, T_k$ variables aleatorias con medias $\theta_1, \dots, \theta_k$ y definamos $T = (T_1, \dots, T_k)$ y $\theta = (\theta_1, \dots, \theta_k)$

Sea $g(T)$ un estimador y que es diferenciable

Definamos $g_i'(\theta) = \frac{\partial}{\partial t_i} g(t) \Big|_{t_1=\theta_1, \dots, t_k=\theta_k}$

entonces $g(t) = g(\theta) + \sum_{i=1}^{k} g_i'(\theta)(t_i - \theta_i) + \text{resto}$

Si desechamos el resto

$$g(t) \approx g(\theta) + \sum_{i=1}^{k} g_i'(\theta)(t_i - \theta_i)$$

luego

$$E_\theta(g(T)) \approx g(\theta) + \sum_{i=1}^{k} g_i'(\theta) \underbrace{E_\theta(T_i - \theta_i)}_{0 \text{ pues } E(T_i)=\theta_i} = g(\theta)$$

Podemos aproximar la varianza de $g(T)$ por

$$\text{Var}_\theta(g(T)) \approx E_\theta([g(T) - g(\theta)]^2) \approx$$

$$\approx E_\theta\left(\left(\sum_{i=1}^{k} g_i'(\theta)(T_i - \theta_i)\right)^2\right) =$$

$$= \sum_{i=1}^{k} [g_i'(\theta)]^2 \operatorname{Var}_\theta(T_i) + 2 \sum_{i>j} g_i'(\theta) g_j'(\theta) \operatorname{Cov}_\theta(T_i, T_j)$$

Si X es una variable aleatoria con $E_\mu(X) = \mu \neq 0$

así

$$g(X) = g(\mu) + g'(\mu)(X-\mu)$$

Si usamos $g(X)$ como un estimador de $g(\mu)$ tenemos

$$E_\mu(g(X)) = g(\mu)$$

$$\operatorname{Var}_\mu(g(X)) \simeq [g'(\mu)]^2 \operatorname{Var}_\mu(X)$$

Como ejemplo específico si tomamos $g(\mu) = 1/\mu$, podemos estimar $1/\mu$ con $1/X$ y podemos ver que

$$E_\mu(1/X) \simeq 1/\mu$$

$$\operatorname{Var}_\mu(1/X) \simeq \left(\frac{1}{\mu}\right)^4 \operatorname{Var}_\mu(X)$$

Usando el desarrollo de Taylor obtenemos una generalización útil del teorema central del límite conocido como el método delta

Teorema 54: (Método Delta) Sea Y_n una sucesión de variables aleatorias que satisfacen $\sqrt{n}(Y_n - \theta) \to N(0, \sigma^2)$ en distribución. Para una función g y un valor específico de θ, supongamos que $g'(\theta)$ existe y no es 0

Entonces

$$\sqrt{n}[g(Y_n)-g(\theta)] \to N(0,\sigma^2[g'(\theta)]^2) \text{ en distribución}$$

Dem: Mediante el desarrollo de Taylor de $g(Y_n)$ alrededor de $Y_n=\theta$ tenemos

$$g(Y_n)=g(\theta)+g'(\theta)(Y_n-\theta)+\text{resto}$$

donde el resto $\to 0$ cuando $Y_n \to \theta$.

Dado que $Y_n \to \theta$ en probabilidad se concluye que el resto $\to 0$ en probabilidad.

Aplicando Slutsky

$$\sqrt{n}[g(Y_n)-g(\theta)]=g'(\theta)\sqrt{n}(Y_n-\theta)$$

y entonces como $\sqrt{n}(Y_n-\theta)\to N(0,\sigma^2)$ podemos poner

$$g'(\theta)\sqrt{n}(Y_n-\theta)=g'(\theta)\sigma\underbrace{\left(\frac{\sqrt{n}(Y_n-\theta)}{\sigma}\right)}_{\to Z\sim N(0,1)}\to N(0,\sigma^2[g'(\theta)]^2)$$

Ejemplo Consideremos $X_1,\dots,X_n$ con $E(X_i)=\mu$ y sea $\bar{X}$. Podemos querer estimar $1/\mu$ entonces con $g(\mu)=1/\mu$ y tenemos $g(\bar{X})=1/\bar{X}$.

A partir del teorema tenemos que

$$\sqrt{n}\left[\frac{1}{\overline{X}}-\frac{1}{\mu}\right]\longrightarrow N\left(0,\left(\frac{1}{\mu}\right)^4 Var_\mu X_1\right)$$

Si no conocemos $Var_\mu X_1$ usaremos una aproximación digamos S^2 o sea tendríamos $\left(\frac{1}{\mu}\right)^4 S^2$.

Puede suceder que no conocemos μ y lo aproximamos por $\overline{X}$ y la varianza por $\left(\frac{1}{\overline{X}}\right)^4 S^2$

Ahora consideremos

$$\frac{\sqrt{n}\left(\frac{1}{\overline{X}}-\frac{1}{\mu}\right)}{\left(\frac{1}{\overline{X}}\right)^2 S}=\underbrace{\frac{\left(\frac{1}{\mu}\right)^4 Var_\mu X_1}{\left(\frac{1}{\overline{X}}\right)^2 S}}_{\xrightarrow{P} 1}\underbrace{\frac{\sqrt{n}\left(\frac{1}{\overline{X}}-\frac{1}{\mu}\right)}{\left(\frac{1}{\mu}\right)^4 Var_\mu(X_1)}}_{\to N(0,1)}$$

y $\overline{X}\xrightarrow{P}\mu$ y $S\xrightarrow{P} Var_\mu(X_1)$ entonces

$$\frac{\left(\frac{1}{\mu}\right)^4 Var_\mu X_1}{\left(\frac{1}{\overline{X}}\right)^4 S}\xrightarrow{P} 1$$

luego $\dfrac{\sqrt{n}\left(\frac{1}{\overline{X}}-\frac{1}{\mu}\right)}{\left(\frac{1}{\overline{X}}\right)^4 S}\longrightarrow N(0,1)$ en distribución

Veamos que sucede si $g'(\theta)=0$.

Veamos que sucede si $g'(\theta)=0$.

Tomamos

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + \frac{g''(\theta)(Y_n-\theta)^2}{2} + \text{Resto}$$

Si $g'(\theta)=0$ entonces

$$g(Y_n) - g(\theta) = \frac{g''(\theta)}{2}(Y_n - \theta)^2 + \text{resto}$$

recordando que el cuadrado de una normal $(0,1)$ es una χ^2_1 tenemos

Teorema 55: (Método delta de segundo orden)
Sea Y_n una sucesión de v.a. que satisface $\sqrt{n}(Y_n - \theta) \to N(0, \sigma^2)$ en distribución.
Para una función g y θ específico supongamos que $g'(\theta)=0$ y $g''(\theta)$ existe y no es 0 entonces

$$n[g(Y_n) - g(\theta)] \to \sigma^2 \frac{g''(\theta)}{2} \chi^2_1 \text{ en distribución}$$

Método Delta Multivariado

Teorema 56: Sean $X_1, \ldots, X_n$ una muestra aleatoria con $E(X_{ij}) = \mu_i$ y $\text{Cov}(X_{ik}, X_{jk}) = \sigma_{ij}$.
Para una función g con derivadas parciales de primer orden continuas, un valor específico de

$\mu = (\mu_1, \ldots, \mu_p)$ para el cual

$$Z^2 = \sum \sum \sigma_{ij} \frac{\partial g(u)}{\partial u_i} \frac{\partial g(u)}{\partial u_j} > 0$$

$$\sqrt{n}\left[g(\bar{X}_1, \ldots, \bar{X}_p) - g(\mu_1, \ldots, \mu_p)\right] \to N(0, Z^2)$$

en distribución

con $\bar{X}_i = \sum_{k=1}^{n} X_{ik}$