Inferencia Estadística
Propiedades de los EMV

Gabriel Martos Venturini
gmartos@utdt.edu

UTDT

UNIVERSIDAD TORCUATO DI TELLA

Hoja de ruta

- En este bloque analizaremos propiedades estadísticas de los EMV:
  - En muestras finitas (con n fijo).
  - Propiedades asintóticas ($n \to \infty$).
- Algunos resultados asintóticos respecto de los test de hipótesis y los intervalos de confianza (basados en estimadores MV) se discutirán oportunamente en el desarrollo de dichos temas.

Cumplen los principios de reducción

- Suficiencia (familias exponenciales).
- Verosimilitud.
- Invarianza.

Definition (Eficiencia)

Una estimador W_n insesgado para θ (en el contexto del modelo $f(x;\theta)$) es eficiente si su varianza alcanza la cota de Cramer–Rao (para todo n).

- Ratio de Eficiencia: $\text{RE}(W_n) = \frac{1/I_n(\theta)}{\text{Var}(W_n)} \leq 1$.
- W_n es Eficiente si $\text{RE}(W_n) = 1 \Rightarrow$ UMVUE.
- Ejemplos donde los EMV son los únicos eficientes (Lehmann–Scheffé):
  - $\{X_1, \ldots, X_n\} \overset{iid}{\sim} \text{N}(\mu, \sigma^2)$, luego $\widehat{\mu}_n = \overline{X}_n$.
  - $\text{Sesgo}^2(\widehat{\mu}_n) = 0$ y $\text{Var}(\widehat{\mu}_n) = \frac{\sigma}{n} = \text{CR}(\mu)$: $\text{RE}(\widehat{\mu}_n) = 1$ para todo n.
- En general, los EMV son *sesgados* y por tanto no eficientes :(
  - $\{X_1, \ldots, X_n\} \overset{iid}{\sim} \text{N}(\mu, \sigma^2)$, luego $\widehat{\sigma}_n^2 = (n-1)S_n^2/n$.
  - $\text{Sesgo}^2(\widehat{\sigma}_n^2) = \frac{\sigma^4}{n^2} \neq 0$ y $\text{Var}(\widehat{\sigma}_n^2) = \frac{2(n-1)\sigma^4}{n^2} > \frac{2(n-1)^2\sigma^4}{n^3} = \text{CR}(\sigma^2)$.
  - $\text{RE}(\widehat{\sigma}_n^2) = (n-1)/n < 1$.

Recapitulación

- En general los EMV son sesgados y por tanto no eficientes.
    - ¿Es la eficiencia una propiedad trascendental?

  Los estimadores UMVUE no siempre existen, si existen no suele ser trivial encontrarlos, y en general pueden existir estimadores sesgados (por tanto ineficientes) pero de menor riesgo cuadrático (compensen el sesgo con una gran reducción en la varianza).

- Las propiedades en muestras finitas parecen escasas, sin embargo desde el punto de vista asintótico, los EMV son sin duda muy buenos.

  Informalmente los estimadores MV se parecen bastante a los UMVUE (son "casi eficientes") cuando n es grande y su distribución se puede aproximar utilizando una distribución normal.

Agenda

1. Propiedades asintóticas
   - Consistencia
   - Normalidad Asintótica
   - Eficiencia Asintótica
2. Apéndice: Propiedades asintóticas de la función de verosimilitud

Agenda

1. Propiedades asintóticas
   - Consistencia
   - Normalidad Asintótica
   - Eficiencia Asintótica
2. Apéndice: Propiedades asintóticas de la función de verosimilitud

Consistencia

- Sea $\underline{X} = \{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x;\theta)$ y $W_n(\underline{X})$ un estimador de θ.

Definition (Consistencia)

Una secuencia de estimadores $\{W_n\}_{n\geq 1}$ se dice **consistente** para el parámetro θ si para todo $\varepsilon > 0$ (tan pequeño como quieras), se cumple:

$$\lim_{n\to\infty} P_\theta(|W_n - \theta| < \varepsilon) = 1, \text{ para todo } \theta \in \Theta.$$

- Equivalentemente W_n es consistente si: $\lim_{n\to\infty} P_\theta(|W_n - \theta| \geq \varepsilon) = 0$.
- La consistencia de W_n ocurre cuando su distribución de probabilidad tiende a concentrarse en torno de θ a medida que n tiende a infinito.
  - Ej: Si $X \sim N(\mu, \sigma_0^2)$ luego $\widehat{\mu}_n \overset{\text{EMV}}{=} \overline{X}_n \sim N(\mu, \sigma_0^2/n)$.
- Notar que por la ley de los grandes números (LGN) $\overline{X}_n$ es siempre consistente para $E(X)$. Sin embargo, en general, la consistencia depende tanto del estimador como del modelo estadístico.

(BackUp)

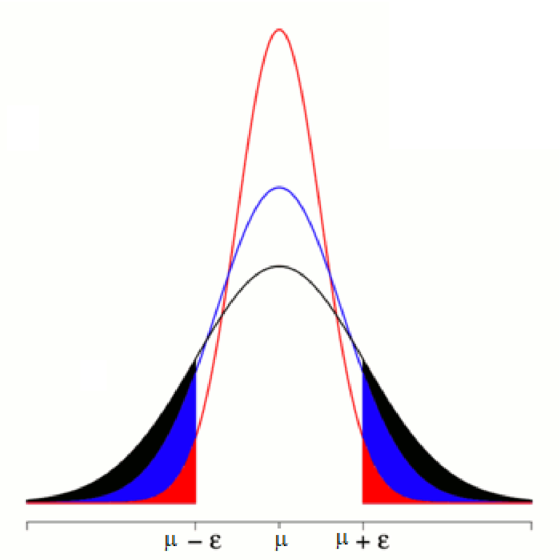

Figure: Si $X \sim N(\mu, \sigma_0^2)$, la distribución de $\widehat{\mu}_n$ se concentra en μ cuando $n \to \infty$.

Relación entre Consistencia y ECM

- Recordemos que el ECM de W_n se descomponía como:

$$\text{ECM}(W_n, \theta) = E[(W_n - \theta)^2] = \text{Var}_\theta(W_n) + \text{Sesgo}_\theta^2(W_n).$$

- Por Chebychev (CB pp–122, th 3.6.1):

$$P(|W_n - \theta| \geq \varepsilon) \leq \frac{E[(W_n - \theta)^2]}{\varepsilon^2} = \frac{\text{ECM}(W_n, \theta)}{\varepsilon^2}.$$

- Por lo que si[1] $\lim_{n\to\infty} \text{ECM}(W_n, \theta) = 0$, entonces W_n es consistente.
  - Si $\lim_{n\to\infty} \text{Sesgo}_\theta(W_n) = 0$ y $\lim_{n\to\infty} \text{Var}_\theta(W_n) = 0 \Rightarrow$ consistencia.

- Ejemplos: Estimadores MV cuando $X \sim N(\mu, \sigma^2)$.

[1] Esta es una condición suficiente para la consistencia.

Algunas reflexiones

- Un estimador puede ser sesgado y consistente:
  - Si W_n es insesgado y consistente para θ, luego $V_n = \frac{n}{n-1} W_n - \frac{1}{n}$ es un estimador sesgado y consistente para θ (¿porqué?).
- Un estimador puede ser insesgado e inconsistente:
  - $W_n = X_n$ para μ, con $\{X_1, \ldots, X_n\} \overset{iid}{\sim} N(\mu, \sigma^2)$ (¿porqué?).
- Notar que si W_n es consistente para θ y ψ es una función continua en Θ, luego $\psi(W_n)$ es consistente para $\psi(\theta)$ (th. del mapa continuo[2]).
  - Ejemplo: $\{X_1, \ldots, X_n\} \overset{iid}{\sim} \text{Bern}(\theta)$, luego $\widehat{\theta}_n = \overline{X}_n$ es consistente para θ (¿porqué?). Si me interesa el parámetro $\psi(\theta) \equiv \theta/(1-\theta)$, entonces $\psi(\widehat{\theta}_n) = \overline{X}_n/(1-\overline{X}_n)$ es un estimador consistente de $\psi(\theta)$.
  - Tener en cuenta que si W_n era insesgado para θ; en general $\psi(W_n)$ será sesgado para $\psi(\theta)$ (salvo el caso en que ψ es una función lineal).

[2] CB pp–233, th 5.5.4.

Theorem (Consistencia de EMV)

Sea $\{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x;\theta)$ *y* $\widehat{\theta}_n$ *el estimador MV de* θ. *Para* ψ *continua en* Θ *y bajo condiciones de regularidad generales (CB § 7.3.8) se tiene que:*

$$\lim_{n\to\infty} P_\theta(|\psi(\widehat{\theta}_n) - \psi(\theta)| < \varepsilon) = 1,$$

luego $\psi(\widehat{\theta}_n)$ *es un estimador consistente de* $\psi(\theta)$.

- Vale en particular para $\psi(\widehat{\theta}_n) = \widehat{\theta}_n$ respecto de $\psi(\theta) = \theta$.
- Condiciones de regularidad (informal): $L(\theta)$ (y por tanto $f(x;\theta)$) es una función suficientemente suave respecto de $\theta \in \Theta$.
- Corolario: Los EMV son *asintóticamente insesgados*.
- Los estimadores de momentos son (cond. gerales.) consistentes.

Agenda

Transformaciones estabilizadoras de la varianza

- Si $\widehat{\theta}_n$ es un estimador consistente de θ, el error de estimación:

$$E_n = \widehat{\theta}_n - \theta \to_P 0.$$

- Para poder estudiar la distribución del error de estimación cuando $n \to \infty$ necesitamos *amplificar los errores* multiplicandolos por una función de n que estabilice la varianza de E_n. Para $\alpha > 0$, hacemos:

$$n^{\alpha} E_n \to_F \text{Distribución conocida.}$$

- Al multiplicar por n^{α} tenemos una 'lupa' que nos permite estudiar el comportamiento asintótico del estimador (o del error de estimación).
- Para el estimador de máxima verosimilitud $\alpha = 1/2$.
    - Para simplificar, en lo que sigue vamos a considerar $\alpha = 1/2$.

Definition (Normalidad asintótica)

Un estimador $\widehat{\theta}_n$ de θ es asintóticamente normal si (asumimos $\alpha = 1/2$):

$$\sqrt{n}(\widehat{\theta}_n - \theta) \rightarrow_F N(0, v_\theta),$$

es decir, si para tamaños de muestra grande, la distribución de $\sqrt{n}(\widehat{\theta}_n - \theta)$ es aproximadamente $N(0, v_\theta)$. Llamaremos a v_θ y $\sqrt{v_\theta}$ a la varianza y el error (desvío) estandard asintóticos de $\widehat{\theta}_n$ como estimador de θ.

- Si $\widehat{\theta}_n$ es asintóticamente normal, entonces para n grande vale que:

$$E_n = \widehat{\theta}_n - \theta \sim_a N\left(0, \frac{v_\theta}{n}\right), \text{ o lo que los mismo } \widehat{\theta}_n \sim_a N\left(\theta, \frac{v_\theta}{n}\right)$$

- El término $\frac{v_\theta}{n}$ aproxima la varianza de $\widehat{\theta}_n$ cuando n es grande.
- Como θ es desconocido, para que estos resultados nos sirvan en la práctica haremos uso de un estimador consistente de v_θ (Slutzky).

- Slutzky: Si $W_n \rightarrow_F W$ y $V_n \rightarrow_P c$, entonces: $W_n V_n \rightarrow_F cW$.
- Si $\widehat{\theta}_n$ es asintóticamente normal y $\widehat{v}_\theta \rightarrow_P v_\theta$, luego:

Example

Consideremos una muestra aleatoria $\{X_1, \ldots, X_n\} \overset{iid}{\sim} X$ y llamemos $\mu = E(X)$ y $\sigma^2 = Var(X)$. Por el TCL se tiene que:

$$\sqrt{n}(\overline{X}_n - \mu) \to_F N(0, \sigma^2).$$

Luego $\overline{X}_n$ es asintóticamente normal con media cero y su varianza asintótica es $v_\mu = \sigma^2$ y por lo tanto para $n \gg 0$ la variabilidad del estimador se puede aproximar con σ^2/n.

- Cómo $S_n^2 \to_p \sigma^2$, luego también vale: $\sqrt{n}(\overline{X}_n - \mu)/S_n \to_F N(0, 1)$.

$$\overline{X}_n \sim_a N(\mu, S_n^2/n), \text{ si } n \gg 0.$$

- Este tipo de construcciones vale también para los EMV. La varianza asintótica de los EMV viene dada por la cota de Cramér–Rao.

(refresh CR / información de Fisher)

- Para $\underline{X} = \{X_1, \dots, X_n\} \overset{iid}{\sim} f(x;\theta)$ y $W_n(\underline{X}) \in \mathcal{C}_\theta$ (H1, H2).

$$V(W_n) \geq \frac{1}{n \underbrace{E\left(\left[\frac{\partial}{\partial\theta} \log f(X;\theta)\right]^2\right)}_{i(\theta)}} = \frac{1}{E\left(\left[\frac{\partial}{\partial\theta}\ell(\theta|\underline{X})\right]^2\right)} = \frac{1}{I_n(\theta)}.$$

- Información de Fisher: $I_n(\theta) = ni(\theta)$.

- Cuando se trata de modelos de la familia exponencial:

$$I_n(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\ell(\theta|\underline{X})\right) \overset{iid}{=} -n \underbrace{E\left(\frac{\partial^2}{\partial\theta^2} \log f(X;\theta)\right)}_{i(\theta)}.$$

- Para los EMV se tiene que: $v_\theta = 1/i(\theta)$.

Theorem (Distribución asintótica de los EMV)

Bajo condiciones de regularidad generales y asumiendo que $0 < i(\theta) < \infty$, los estimadores MV tienen asintóticamente una distribución normal:

$$\sqrt{n}(\widehat{\theta}_n - \theta) \to_F N(0, 1/i(\theta)).$$

- Para $n \gg 0$, el error (desvío) standard del EMV se aproxima como:

$$\mathrm{se}_\theta \overset{n \gg 0}{=} \sqrt{\frac{1}{ni(\theta)}} \overset{\text{c.m.t.}}{=} \sqrt{\frac{1}{ni(\widehat{\theta}_n)}} = \widehat{\mathrm{se}}_\theta.$$

- Si $i(\theta)$ es una función continua, también vale que (Slutsky):

$$\sqrt{ni(\widehat{\theta}_n)}(\widehat{\theta}_n - \theta) \to_F N\big(0, 1\big).$$

- Por lo tanto, para $n \gg 0$ vale la aproximación:

$$\boxed{\widehat{\theta}_n \sim_a N\left(\theta, \frac{1}{ni(\widehat{\theta}_n)}\right).}$$

Ejemplo: $X \sim \text{Bern}(\theta)$

¿Utilidad de este resultado?

- Nos permite construir intervalos de confianza y/o diseñar tests para el parámetro del modelo a partir de la distribución aproximada del EMV (validos cuando $n \gg 0$). Como oportunamente veremos se cumple:

$$\text{IC}_{1-\alpha}(\theta) = \left\{ \widehat{\theta}_n \pm z_{\alpha/2} \frac{1}{\sqrt{n i(\widehat{\theta}_n)}} \right\} = \left\{ \widehat{\theta}_n \pm z_{\alpha/2} \widehat{\text{se}}_\theta \right\}.$$

- Ejemplo: Intervalos de confianza con el modelo $\text{Bern}(\theta)$.

$$\text{IC}_{1-\alpha}(\theta) = \left\{ \widehat{\theta}_n \pm z_{\alpha/2} \underbrace{\sqrt{\frac{\widehat{\theta}_n(1-\widehat{\theta}_n)}{n}}}_{\widehat{\text{se}}_\theta} \right\}.$$

- ¿Ejemplo en `R`? (volveremos sobre este tema en la próxima clase).

Distribución asintóticas de funciones del EMV

- Si $\psi'(\theta) \neq 0$ luego (**método delta** + normalidad asintótica):

$$\sqrt{n}(\psi(\widehat{\theta}_n) - \psi(\theta)) \rightarrow_F N\Big(0, (\psi'(\theta))^2[i(\theta)]^{-1}\Big).$$

- Para $n \gg 0$, el error standard de $\widehat{\psi}_n \equiv \psi(\widehat{\theta}_n)$ se aproxima como:

$$\text{se}_\psi \overset{n \gg 0}{=} \sqrt{\frac{(\psi'(\theta))^2}{ni(\theta)}} \overset{\text{c.m.t.}}{=} \sqrt{\frac{(\psi'(\widehat{\theta}_n))^2}{ni(\widehat{\theta}_n)}} = \widehat{\text{se}}_\psi.$$

- Asumiendo continuidad en $i(\theta)$ (y usando Slutsky):

$$\sqrt{\frac{ni(\widehat{\theta}_n)}{(\psi'(\widehat{\theta}_n))^2}}(\psi(\widehat{\theta}_n) - \psi(\theta)) \rightarrow_F N(0, 1).$$

- Por lo tanto, para $n \gg 0$ se tiene que:

$$\widehat{\psi}_n \sim_a N\left(\psi; \frac{(\psi'(\widehat{\theta}_n))^2}{ni(\widehat{\theta}_n)}\right).$$

Extensiones (modelos multiparámetro)

- Los resultados anteriores también valen cuando $\dim(\Theta) = d > 1$:

$$\boxed{\widehat{\boldsymbol{\theta}}_n \overset{n \gg 0}{\sim}_a N_d(\boldsymbol{\theta}, \mathbf{I}_n(\boldsymbol{\theta})^{-1}).}$$

- Por lo tanto (Slutsky), con n grande: $\widehat{\boldsymbol{\theta}}_n \overset{n \gg 0}{\sim}_a N_d(\boldsymbol{\theta}, \mathbf{I}_n(\widehat{\boldsymbol{\theta}}_n)^{-1})$.
- Otras aproximaciones asintóticas también son válidas:
  - $2\left(\ell_n(\widehat{\boldsymbol{\theta}}_n|\underline{X}) - \ell_n(\boldsymbol{\theta}|\underline{X})\right) \rightarrow_F \chi^2_d$, donde $d = \dim(\Theta)$.
  - Estas aproximaciones son habituales al plantear test (de ratios de verosimilitud) para los parámetros del modelo (discutido más adelante).
  - Ejemplo elipse de confianza para el modelo normal.

Agenda

1. Propiedades asintóticas
   - Consistencia
   - Normalidad Asintótica
   - Eficiencia Asintótica

2. Apéndice: Propiedades asintóticas de la función de verosimilitud

- En las transparencias anteriores vimos que los EMV son consistentes y asintóticamente normales. También se cumple que su varianza, a medida que n crece, se aproxima a la cota de CR.
- Si $\sqrt{n}(W_n - \theta) \to_F N(0, \sigma^2_W(\theta))$ (W_n es asintóticamente normal[3]); la eficiencia asintótica de W_n:

$$\mathrm{EA}(W_n) = \frac{1/i(\theta)}{\sigma^2_W(\theta)} \leq 1.$$

- Los EMV son asintóticamente eficientes porque $\mathrm{EA}(\widehat{\theta}_n) = 1$.

Informalmente para $n \gg 0$ los estimadores MV son casi insesgados, tienen una varianza muy parecida a la de los estimadores UMVUE y su distribución se puede aproximar con una normal :) :) :)

[3]Notar que $\sigma^2_W(\theta)$ es constante respecto de n, pero puede depender del valor de θ.

Eficiencia asintótica relativa

Definition (Eficiencia asintótica relativa)

Sean W_n y V_n dos estimadores consistentes de θ tales que:

$$\sqrt{n}(W_n - \theta) \to_F N(0, \sigma^2_W(\theta)), \text{ y } \sqrt{n}(V_n - \theta) \to_F N(0, \sigma^2_V(\theta)),$$

definimos la eficiencia asintótica relativa $\text{EAR}(W_n, V_n) = \sigma^2_W(\theta)/\sigma^2_V(\theta)$.

- EAR como criterio para elegir entre estimadores consistentes.
- Ejemplo: Estimando $\psi = P(X = 0)$ cuando $X \sim \text{Poiss}(\lambda)$.
  - Estrategia 1: Utilizar invarianza $\widehat{\psi} = \psi(\widehat{\lambda})$.
  - Estrategia 2: Bernullizar el problema $\widehat{p} = \widehat{P}(Y = 1)$ (?).
  - Como $\sigma^2_{\widehat{\psi}} \leq \sigma^2_{\widehat{p}}$, preferimos el primer estimador.

Conclusiones

- Los EMV cumplen los 3 principios de inferencia.
- En muestras finitas, en general son sesgados y por tanto no eficientes.
- Si embargo son consistentes y asintóticamente eficientes.
  - Con $n \gg 0$, el sesgo de los EMV es pequeño y su varianza se parece a la de los estimadores eficientes (es muy parecido al UMVUE).
  - Aproximamos su varianza (error estandard) con la inversa de la información de Fisher evaluada en la estimación máximo verosímil.
- Podemos aproximar su distribución asintótica (y la de cualquier función continua del EMV) utilizando una distribución normal.
  - Esto nos permite construir intervalos y testear hipótesis (hacer inferencia) para θ a partir de estos estimadores.

Ejercicio integrador

Sea $\{X_1, \ldots, X_n\} \overset{iid}{\sim} \text{Exp}(\theta)$, se pide:

1. Obtenga el EMV de θ verificando la CSO.
2. ¿Es el EMV una función del estadístico suficiente?
3. ¿Es el EMV el UMVUE?
4. Compute el riesgo cuadrático del estimador.
5. ¿Es el estimador consistente?
6. Determine la distribución asintótica del EMV.
7. De una muestra de tamaño $n = 10$ se tiene $\sum_{i=1}^{10} X_i = 20$; calcule la estimación máximo verosímil de θ, aproxime el error estándar del estimador (¿para qué sirve esta cantidad?).

Agenda

1. Propiedades asintóticas
2. Apéndice: Propiedades asintóticas de la función de verosimilitud

- Bajo las siguientes condiciones de regularidad:

  1. El soporte del modelo, $\{x \mid f(x;\theta) > 0\}$,es el mismo para todo $\theta \in \Theta$.
  2. Identificabilidad: $\theta \neq \theta' \Rightarrow f(x;\theta) \neq f(x;\theta')$ para todo θ, $\theta' \in \Theta$.
  3. θ^* es un punto *interior* de Θ, siendo Θ compacto.

  ... que se cumplen en general en la familia exponencial.

- Llamemos θ^* al verdadero parámetro en la población:

$$\lim_{n \to \infty} P\Big(L_n(\theta^* | \underline{X}) > L_n(\theta | \underline{X})\Big) = 1, \text{ para todo } \theta \neq \theta^*.$$

- Cuando $n \gg 0$, con una probabilidad alta ocurre que el máximo de $L_n(\theta)$ ocurre en el verdadero valor de θ en la población.

- Cuando $n \to \infty$ identificamos al "modelo" <u>correcto</u> con certeza.

$X \sim N(\mu = 2, \sigma_0^2 = 1)$

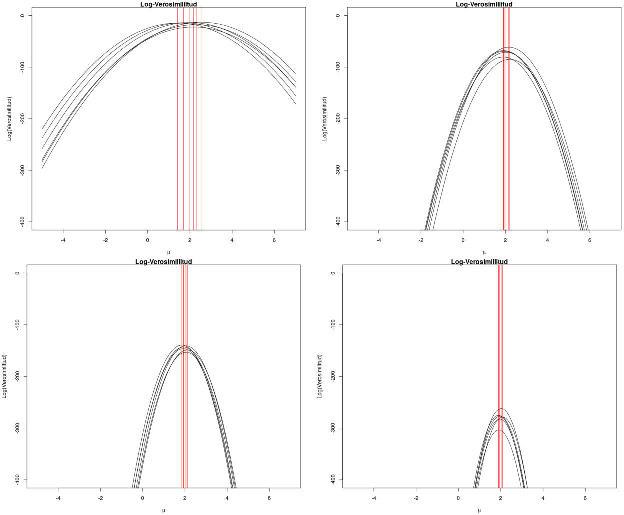

Figure: Algunas realizaciones de $\ell_n(\mu)$ para $n = 10, 50, 100, 200$.