9 Estimación puntual II

9.1 Estimadores consistentes e insesgados

Según el teorema de Glivenko-Cantelli, sabemos que la muestra tiende a dar información completa sobre la distribución teórica cuando crece su tamaño.

Definición. Una sucesión de estimadores $T_1, T_2, \ldots$ asociada a los sucesivos tamaños muestrales n, se denomina consistente para estimar una función $g(\theta)$ si $T_n \xrightarrow{p} g(\theta)$ para $\theta \in \Theta$.

Ejemplo. La media muestral $\bar{X}_n$ es un estimador consistente de la media poblacional $\mu = g(\theta)$, ya que, por la ley débil de los grandes números, se tiene que $\bar{X}_n \xrightarrow{p} \mu$ para todo $\theta \in \Theta$. Esta consistencia es válida para cualquier distribución con μ finita.

Si $\lim_{n\to\infty} \mathbb{E}_\theta[T_n] = g(\theta)$ y $\lim_{n\to\infty} \operatorname{var}_\theta[T_n] = 0$ para todo $\theta \in \Theta$, entonces T_n es consistente para $g(\theta)$. En particular, como $\mathbb{E}[\bar{X}_n] = \mu$ y $\operatorname{var}[\bar{X}_n] = \sigma^2/n$, se cumple la condición anterior y se concluye que $\bar{X}_n$ es consistente para μ.

Ejemplo. Sea $X_1, \ldots, X_n \sim \operatorname{Exp}(\theta)$ una m.a.s. que modela la duración de n LEDs con tasa de fallo constante. El MLE de $\mu = \frac{1}{\theta}$ es $\overline{X}$, que es un estimador consistente. Se propone un estimador alternativo $T = nX_{(1)}$, que solo requiere esperar al primer fallo. Se tiene: $P(T \leq t) = 1 - e^{-\theta t} \Rightarrow T \sim \operatorname{Exp}(\theta)$, por lo que $E[T] = \frac{1}{\theta} = \mu$ y $\operatorname{var}[T] = \frac{1}{\theta^2}$. Aunque $\mathbb{E}[T] = \mu$, T no es consistente, ya que su distribución no depende de n y no se concentra alrededor de μ al crecer n.

Comentario. Los momentos y cuantiles muestrales se aproximan a los poblacionales al crecer el tamaño muestral, es decir, son estimadores consistentes.

Definición. El sesgo de T para estimar $g(\theta)$ es

$$b_T(g(\theta)) = \mathbb{E}_\theta[T] - g(\theta)$$

Un estimador es *insesgado* si $\mathbb{E}_\theta[T] = g(\theta)$.

Ejemplo. La media muestral $\bar{X}$ y la cuasivarianza S^2 son estimadores insesgados de μ y σ^2. La varianza muestral s^2 tiene sesgo: $b_{s^2}(\sigma^2) = \mathbb{E}[s^2] - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}$, que desaparece cuando $n \to \infty$, así que es asintóticamente insesgada.

Comentario. Si $\hat{\theta}$ es insesgado para θ, entonces $g(\hat{\theta})$ es insesgado para $g(\theta)$ solo si g es lineal. Por ejemplo, aunque S^2 es insesgado para σ^2, S no lo es para σ en general.

9.2 Error cuadrático medio como medida de la bondad de un estimador

El sesgo no siempre refleja bien la calidad de un estimador: un estimador puede ser insesgado y muy malo (como usar solo un dato), mientras que otro con ligero sesgo puede ser mucho más fiable al aprovechar mejor la información.

Sea T un estimador de θ y t un valor observado. La discrepancia entre θ y t se mide mediante una función de pérdida $L(\theta, t)$, que puede reflejar distintos tipos de errores o costes asociados, como el error relativo o el cuadrático.

Definición (Función de riesgo)**.** El riesgo del estimador T es la pérdida esperada

$$R_T(\theta) = \mathbb{E}_\theta[L(\theta, T)],$$

función que depende de θ.

Definición (Error cuadrático medio)**.** El *error cuadrático medio* de un estimador $\hat{\theta}$ es

$$\mathbb{E}_\theta[(\hat{\theta} - \theta)^2]$$

Podemos expresar el error cuadrático medio en términos de la varianza y el sesgo:

$$\begin{aligned}
\mathbb{E}_\theta[(\hat{\theta} - \theta)^2] &= \mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta(\hat{\theta}) + \mathbb{E}_\theta(\hat{\theta}) - \theta)^2] \\
&= \mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta(\hat{\theta}))^2] + [\mathbb{E}_\theta(\hat{\theta}) - \theta]^2 + 2\mathbb{E}_\theta[\mathbb{E}_\theta(\hat{\theta}) - \theta]\underbrace{\mathbb{E}_\theta[\hat{\theta} - \mathbb{E}_\theta(\hat{\theta})]}_{0} \\
&= \operatorname{var}(\hat{\theta}) + \operatorname{bias}^2(\hat{\theta})
\end{aligned}$$

Si nuestro objetivo es minimizar el error cuadrático medio, en ocasiones puede ser preferible utilizar un estimador sesgado con menor varianza.

Ejemplo. Sea $X \sim \mathcal{B}(n, \theta)$ con n conocido. El estimador usual es $T_U = \frac{X}{n}$, insesgado, con varianza

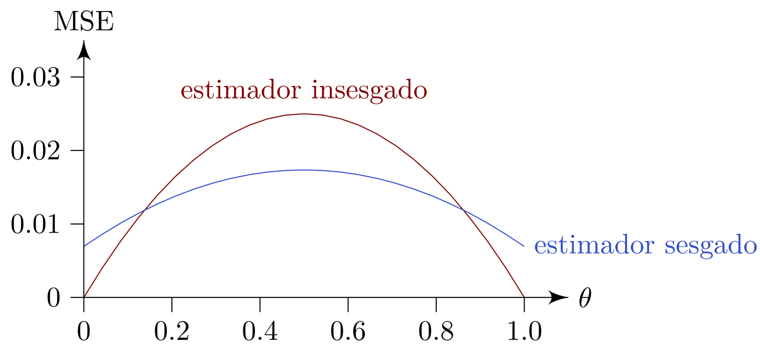
$$\operatorname{var}_\theta(T_U) = \frac{\operatorname{var}_\theta(X)}{n^2} = \frac{\theta(1-\theta)}{n}$$

Su error cuadrático medio es $\operatorname{mse}(T_U) = \frac{\theta(1-\theta)}{n}$. Consideremos el estimador alternativo $T_B = \frac{X+1}{n+2} = w\frac{X}{n} + (1-w)\frac{1}{2}$, con $w = \frac{n}{n+2}$. Su sesgo y varianza son

$$b(T_B) = \mathbb{E}_\theta[T_B] - \theta = (1-w)\left(\frac{1}{2} - \theta\right), \quad \operatorname{var}_\theta(T_B) = \frac{\operatorname{var}_\theta(X)}{(n+2)^2} = w^2\frac{\theta(1-\theta)}{n}$$

Por tanto, su MSE es $\operatorname{mse}(T_B) = w^2\frac{\theta(1-\theta)}{n} + (1-w)^2\left(\frac{1}{2} - \theta\right)^2$.

Podemos representar gráficamente el error cuadrático medio de cada estimador para distintos valores posibles de θ. A continuación, se muestra el caso en el que $n = 10$:



Este estimador sesgado tiene un menor error cuadrático medio, salvo cuando θ toma valores extremos.

Definición. Se dice que un estimador T_1 es preferible a otro T_2 para estimar $g(\theta)$ si $R_{T_1}(\theta) \leq R_{T_2}(\theta)$ para todo $\theta \in \Theta$ y $R_{T_1}(\theta) < R_{T_2}(\theta)$ para algún $R_{T_1}(\theta) < R_{T_2}(\theta)$.

Aunque sería ideal hallar un estimador preferible a todos, a menudo no existe.

Ejemplo. Sea una m.a.s. de tamaño n de una distribución $\mathcal{N}(\mu, \sigma)$. Se comparan $\overline{X}$ y la mediana muestral M como estimadores de μ. Se tiene: $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, $M \sim \mathcal{AN}\left(\mu, \frac{\sigma\sqrt{\pi}}{\sqrt{2n}}\right)$. Ambos son (asintóticamente) insesgados, pero el cociente de errores cuadráticos medios es

$$\frac{\text{MSE}(\overline{X})}{\text{MSE}(M)} \approx \sqrt{\frac{2}{\pi}} \approx 0{,}8$$

Por tanto, para n grande, $\overline{X}$ es preferible. Sin embargo, M es más robusto ante datos atípicos.

Ejemplo. Consideramos una muestra aleatoria $X_1, X_2, \ldots, X_n \sim U(0, \theta)$ con $\theta > 0$ desconocido. Para construir un estimador insesgado de θ, usamos $X_{(n)} = \max\{X_1, \ldots, X_n\}$. Primero, obtenemos su función de densidad:

$$F_{X_{(n)}}(z) = P(X_{(n)} \leq z) = P(X_1 \leq z, \ldots, X_n \leq z) = [F_X(z)]^n = \left(\frac{z}{\theta}\right)^n$$

para $0 < z < \theta$. Derivando, obtenemos la función de densidad:

$$f_{X_{(n)}}(z) = \frac{d}{dz}\left(\frac{z}{\theta}\right)^n = \frac{nz^{n-1}}{\theta^n}, \quad 0 < z < \theta$$

Ahora, calculamos la esperanza del máximo

$$\mathbb{E}[X_{(n)}] = \int_0^\theta z \cdot \frac{nz^{n-1}}{\theta^n}\, dz = \frac{n}{\theta^n}\int_0^\theta z^n\, dz = \frac{n}{\theta^n} \cdot \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta$$

Buscamos una constante c tal que $T = cX_{(n)}$ sea insesgado, es decir, $\mathbb{E}[T] = \theta$. Como $\mathbb{E}[X_{(n)}] = \frac{n}{n+1}\theta$, se deduce que $c = \frac{n+1}{n}$, y así $T = \frac{n+1}{n}X_{(n)}$ es un estimador insesgado de θ.

Finalmente, este estimador es consistente, ya que $X_{(n)} \xrightarrow{p} \theta$ al crecer n, y el factor $\frac{n+1}{n} \to 1$, por lo tanto $T \xrightarrow{p} \theta$.

Ejemplo. Buscamos c_0 tal que $T = c_0 X_{(n)}$ minimice el MSE dentro de los estimadores $cX_{(n)}$. Tenemos que $\mathbb{E}[X_{(n)}] = \frac{n}{n+1}\theta$ y $\mathbb{E}[X_{(n)}^2] = \frac{n}{n+2}\theta^2$. Entonces, para $T = cX_{(n)}$:

$$\text{sesgo} = \left(c\frac{n}{n+1} - 1\right)\theta, \quad \text{var}(T) = c^2\left(\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right)\theta^2$$

$$\text{MSE}(T) = \left(\frac{n}{n+2}c^2 - 2\frac{n}{n+1}c + 1\right)\theta^2 = h(c)\theta^2$$

El mínimo de $h(c)$ se alcanza en $c_0 = \frac{n+2}{n+1}$. Por tanto, el mejor estimador de la forma $cX_{(n)}$ (en términos de MSE) es: $T = \frac{n+2}{n+1}X_{(n)}$.

Ejemplo. Sea $X_1, \dots, X_n \sim \text{U}(0, \theta)$, con $\theta > 0$, y sea $\mu = \frac{\theta}{2}$. Se comparan dos estimadores de μ: $T_1 = \frac{c}{2}X_{(n)}$ con $c = \frac{n+2}{n+1}$, y $T_2 = \bar{X}$. $\text{MSE}(T_1) = \frac{1}{4}\text{MSE}(cX_{(n)}) = \frac{1}{4}h(c)\theta^2 \quad$ con $h(c) = \frac{n}{n+2}c^2 - 2\frac{n}{n+1}c + 1$. Para $c = \frac{n+2}{n+1}$, se obtiene

$$\text{MSE}(T_1) = \frac{-(n(n+2) - (n+1)^2)}{4(n+1)^2}\theta^2 = \frac{1}{4(n+1)^2}\theta^2$$

Para T_2,

$$\text{MSE}(\bar{X}) = \text{var}(\bar{X}) = \frac{\theta^2}{12n}$$

Por tanto, $\frac{1}{4(n+1)^2} < \frac{1}{12n} \quad \forall n \Rightarrow \text{MSE}(T_1) < \text{MSE}(T_2)$. $T_1 = \frac{n+2}{2(n+1)}X_{(n)}$ es preferible a $\bar{X}$ para estimar $\mu = \frac{\theta}{2}$. Además, su MSE converge a 0 más rápido que el de $\bar{X}$.

9.3 Estimador centrado de uniformemente mínima varianza

Definición (Estimador centrado de uniformemente mínima varianza)**.** Un estimador centrado T de $g(\theta)$ con varianza finita es ECUMV si para todo estimador centrado T' con varianza finita se cumple $\text{var}_\theta(T) \leq \text{var}_\theta(T')$ para todo $\theta \in \Theta$.

Comentario. Si existe, el ECUMV es único y preferible pues su MSE coincide con la varianza.

Teorema (de Rao-Blackwell)**.** Si S es suficiente para $\{F_\theta\}$ y T_1 es estimador insesgado centrado de $g(\theta)$, entonces $T_2 = \mathbb{E}[T_1 \mid S]$ cumple:

(a) T_2 es independiente de θ.

(b) T_2 es insesgado para $g(\theta)$.

(c) $\text{var}_\theta(T_2) \leq \text{var}_\theta(T_1)$, con igualdad solo si $T_2 = T_1$ casi seguro.

Demostración. (a) Como la distribución de la muestra condicionada a S no depende de θ, tampoco $\mathbb{E}_\theta[T_1 \mid S]$, luego T_2 es un estimador independiente de θ.

(b) Se tiene $\mathbb{E}_\theta[T_2] = \mathbb{E}_\theta[\mathbb{E}[T_1 \mid S]] = \mathbb{E}_\theta[T_1] = g(\theta)$, por lo que T_2 es insesgado.

(c) Por la descomposición de varianza, $\text{var}_\theta[T_1] = \text{var}_\theta(\mathbb{E}[T_1 \mid S]) + \mathbb{E}_\theta[\text{var}(T_1 \mid S)] \geq \text{var}_\theta[T_2]$, con igualdad si y sólo si $\text{var}(T_1 \mid S) = 0$ casi seguro, es decir, $T_1 = T_2$ casi seguro.

□

Anteriormente se han considerado sólo estimadores insesgados, donde el ECM coincide con la varianza. Rao-Blackwell mejora un estimador insesgado T_1 obteniendo $T_2 = \mathbb{E}[T_1 \mid S]$, función del estadístico suficiente S, con varianza menor o igual a la de T_1. Así, el ECUMV, si existe, debe ser función de S. Si T_1 no es función de S, entonces $T_2 \neq T_1$ es preferible, por lo que T_1 no es ECUMV. Además, T_2 es función de S y satisface $\mathbb{E}[T_2 \mid S] = \mathbb{E}[T_2 \cdot 1 \mid S] = T_2 \cdot \mathbb{E}[1 \mid S] = T_2$.

Frecuentemente, existe una única función h del estadístico suficiente minimal S que es insesgada para $g(\theta)$. En este caso, por Rao-Blackwell, $h(S)$ debe ser el ECUMV, y además: $h(S) = \mathbb{E}[T_1 \mid S]$ para cualquier estimador insesgado T_1.

Teorema (Lehmann-Scheffé)**.** Si S es un estadístico suficiente y completo, y existe T de θ insesgado, entonces existe un único ECUMV para θ, y viene dado por $\mathbb{E}[T \mid S]$.

En todos los ejemplos, el estadístico suficiente s_m es completo, por lo que el ECUMV es la única función centrada de s_m. A veces se obtiene directamente, otras calculando $\mathbb{E}[T \mid S]$.

Ejemplo. Para las siguientes familias y funciones g, el ECUMV es la única función centrada del estadístico suficiente s_m:

(a) $\mathcal{B}(1, \theta)$, $g(\theta) = \theta$: $SM = \bar{X}$ es suficiente e insesgado para $\mu = \theta$, luego ECUMV.

(b) $\mathcal{P}(\theta)$, $g(\theta) = \theta$: $SM = \bar{X}$ es suficiente e insesgado para $\mu = \theta$, luego ECUMV.

(c) $\mathcal{N}(\mu, \sigma)$, $g(\mu, \sigma) = \mu$: $SM = (\bar{X}, S^2)$, y $\bar{X}$ es insesgado para μ y función de SM, luego ECUMV.

(d) $\mathcal{N}(\mu, \sigma)$, $g(\mu, \sigma) = \sigma^2$: $SM = (\bar{X}, S^2)$, y S^2 es insesgado para σ^2 y función de SM, luego ECUMV.

Ejemplo. Sea $X_1, \dots, X_n$ una m.a.s. de $\mathcal{B}(1, \theta)$. Queremos estimar $g(\theta) = \theta(1 - \theta) = \sigma^2$. Como $SM = \sum X_i$ es suficiente y completo, el ECUMV es la única función centrada de SM. Se tiene que $SM = \bar{X}$. No parece sencillo encontrar de modo directo una función centrada del estadístico suficiente $\bar{X}$.

Sea $T_1 = \mathbb{I}_{\{X_1=1,\, X_2=0\}}$, que es centrado para $g(\theta)$, puesto que $\mathbb{E}_\theta[T_1] = 1 \cdot \mathbb{P}(T_1 = 1) + 0 \cdot \mathbb{P}(T_1 = 0) = \mathbb{P}(X_1 = 1, X_2 = 0) = \mathbb{P}(X_1 = 1) \cdot \mathbb{P}(X_2 = 0) = \theta(1 - \theta) = g(\theta)$. Es ahora más cómodo considerar el estadístico suficiente $SM = \sum_{i=1}^{n} X_i$, que tiene distribución $\mathcal{B}(n,\theta)$. Queremos calcular $\mathbb{E}[T_1 \mid SM = s] = \mathbb{P}(T_1 = 1 \mid SM = s)$ y

$$\begin{aligned}
\mathbb{P}(T_1 = 1 \mid SM = s) &= \mathbb{P}(X_1 = 1, X_2 = 0 \mid \textstyle\sum_{i=1}^{n} X_i = s) \\
&= \frac{\mathbb{P}(X_1 = 1, X_2 = 0, \sum_{i=3}^{n} X_i = s-1)}{\mathbb{P}(\sum_{i=1}^{n} X_i = s)} \\
&= \frac{\theta \cdot (1-\theta) \cdot \mathbb{P}(\mathcal{B}(n-2,\theta) = s-1)}{\mathbb{P}(\mathcal{B}(n,\theta) = s)} \\
&= \frac{\theta(1-\theta)\binom{n-2}{s-1}\theta^{s-1}(1-\theta)^{n-s-1}}{\binom{n}{s}\theta^{s}(1-\theta)^{n-s}} \\
&= \frac{\binom{n-2}{s-1}}{\binom{n}{s}}.
\end{aligned}$$

Utilizando la identidad combinatoria: $\frac{\binom{n-2}{s-1}}{\binom{n}{s}} = \frac{s(n-s)}{n(n-1)}$, concluimos que el ECUMV para $g(\theta) = \theta(1-\theta) = \sigma^2$ es: $\mathbb{E}[T_1 \mid SM] = \frac{SM(n-SM)}{n(n-1)} = \frac{\bar{X}(1-\bar{X})}{n-1}$.

Ejemplo. Sea $X_1, \ldots, X_n$ una m.a.s. de $P(\theta)$. Queremos estimar $g(\theta) = P_\theta(X > 0) = 1 - e^{-\theta}$. El ECUMV es único y es la única función centrada del estadístico suficiente $S_M = \sum_{i=1}^{n} X_i$. Sea $T_1 = \mathbb{I}_{\{X_1 > 0\}}$, entonces: $\mathbb{E}[T_1] = P(X_1 > 0) = 1 - e^{-\theta} = g(\theta)$. Por tanto, aplicando el teorema,

$$\begin{aligned}
\mathbb{E}[T_1 \mid S_M = s] &= P(X_1 > 0 \mid S_M = s) = 1 - P(X_1 = 0 \mid S_M = s) \\
&= 1 - \frac{P(X_1 = 0) \cdot P(\sum_{i=2}^{n} X_i = s)}{P(S_M = s)} \\
&= 1 - \frac{e^{-\theta} \cdot \frac{[(n-1)\theta]^s}{s!} e^{-(n-1)\theta}}{\frac{(n\theta)^s}{s!} e^{-n\theta}} = 1 - \left(\frac{n-1}{n}\right)^s
\end{aligned}$$

Luego, el ECUMV es $\mathbb{E}[T_1 \mid S_M] = 1 - \left(\frac{n-1}{n}\right)^{S_M} = 1 - \left(1 - \frac{1}{n}\right)^{X}$. El EMV de $g(\theta)$ es $1 - e^{-\hat{\theta}} = 1 - e^{-X}$. y ambos son próximos para n grande, ya que $\left(1 - \frac{1}{n}\right)^n \to e^{-1}$. Además, si T_1, T_2 son ECUMV para $g_1(\theta)$ y $g_2(\theta)$, entonces $T = aT_1 + bT_2$ es ECUMV para $ag_1(\theta) + bg_2(\theta)$.

Ejemplo. Consideremos una m.a.s. de una distribución teórica $B(1,\theta)$. Obtenemos el ECUMV para $g(\theta) = \theta^2$. Se tiene que $g(\theta) = g_1(\theta) - g_2(\theta)$, con $g_1(\theta) = \theta$ y $g_2(\theta) = \theta(1-\theta)$. En ejemplos anteriores obtuvimos el ECUMV para $g_1(\theta)$, que es $T_1 = \bar{X}$, y el ECUMV para $g_2(\theta)$, que es $T_2 = \frac{n}{n-1}\bar{X}(1-\bar{X})$. Entonces, el ECUMV para $g(\theta) = g_1(\theta) - g_2(\theta) = \theta^2$ es $T = T_1 - T_2 = \bar{X} - \frac{n}{n-1}\bar{X}(1-\bar{X}) = \frac{n}{n-1}\bar{X}^2 - \bar{X}$.