Capítulo 7

Estimación puntual

7.1. Introducción

Definición 7.1.1 Un *estimador puntual* es cualquier función $W(X_1, \cdots, X_n)$ de la muestra. Es decir, cualquier estadística es una estimador puntual.

Se debe tener clara la diferencia entre *estimador* y *estimación*. Un *estimador* es una función de una muestra, mientras que una *estimación* es el valor obtenido al aplicar un *estimador* a los datos de una muestra. Es decir, un estimador es una función de las variables aleatorias $X_1, \cdots, X_n$ mientras que una estimación es una función de los valores muestrales $x_1, \cdots, x_n$.

7.2. Métodos para encontrar estimadores

En muchos casos habrá un candidato evidente o natural para ser el estimador puntual de un parámetro particular y a menudo la intuición puede inducirnos a obtener buenos estimadores. Por ejemplo, la media muestral es un candidato natural para estimar la media poblacional.

7.2.1. Métodos de momentos

Se trata quizás de uno de los métodos más antiguos para hallar estimadores. Tiene la ventaja de ser fácil de usar y casi siempre permite obtener algún tipo de estimador que en algunos casos puede ser mejorado.

Sea $X_1, \cdots, X_n$ una muestra de una población con función de probabilidad o densidad $f(x|\theta_1, \cdots, \theta_k)$. Los estadísticos por el método de momentos se encuentran igualando los k primeros momentos muestrales a sus correspondientes k momentos poblacionales y resolviendo simultáneamente las ecuaciones resultantes. Es decir, se definen:

$$\begin{aligned}
m_1 &= \frac{1}{n}\sum_{i=1}^{n} X_i^1, \qquad \mu_1 = \mathrm{E}[X^1] \\
m_2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2, \qquad \mu_2 = \mathrm{E}[X^2] \\
&\vdots \\
m_k &= \frac{1}{n}\sum_{i=1}^{n} X_i^k, \qquad \mu_k = \mathrm{E}[X^k]
\end{aligned}$$

El momento poblacional μ_j es, por lo general, una función de $\theta_1, \cdots, \theta_k$ digamos $\mu_j(\theta_1, \cdots, \theta_k)$. El estimador por el método de momentos $(\tilde{\theta}_1, \cdots \tilde{\theta}_k)$ de $(\theta_1, \cdots, \theta_k)$ se obtiene resolviendo el siguiente sistema de ecuaciones en términos de $(m_1, \cdots, m_k)$:

$$\begin{aligned}
m_1 &= \mu_1(\theta_1, \cdots, \theta_k) \\
m_2 &= \mu_2(\theta_1, \cdots, \theta_k) \\
&\vdots \\
m_k &= \mu_k(\theta_1, \cdots, \theta_k)
\end{aligned}$$

Ejemplo 7.2.1 Sea $X_1, \cdots, X_n$ una muestra aleatoria obtenida desde la distribución $\mathcal{N}(\mu, \sigma^2)$. Hallar los estimadores de μ y σ^2 usando el método de momentos.

Ejemplo 7.2.2 Sea $X_1, \cdots, X_n$ una muestra aleatoria obtenida desde la distribución $\mathcal{BI}(k, p)$ con ambos parámetros desconocidos. Hallar sus estimadores por el método de momentos.

7.2.2. Estimadores de máxima verosimilitud

Es uno de los métodos más populares ya que permite encontrar estimadores con propiedades importantes para el proceso de inferencia. Consiste en usar la información disponible en la muestra para elegir el valor del parámetro para el cual es más probable haber observado los resultados muestrales.

Definición 7.2.1 Sea $f(\mathbf{x}|\theta)$ que denota la función de probabilidad o densidad conjunta de la muestra $\mathbf{X} = (X_1, \cdots, X_n)$. Entonces, dado que $\mathbf{X} = \mathbf{x}$ es observado, la función de θ definida por:

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$$

es llamada *función de verosimilitud.* Si $X_1, \cdots, X_n$ es una muestra independiente e identicamente distribuida de una población con función de probabilidad o densidad $f(\mathbf{x}|\theta_1, \cdots, \theta_k)$, *la función de verosimilitud* se define por:

$$L(\theta|\mathbf{x}) = L(\theta_1, \cdots, \theta_k|x_1, \cdots, x_n) = \prod_{i=1}^{n} f(x_i|\theta_1, \cdots, \theta_k) \tag{7.2.1}$$

Definición 7.2.2 Sea $\hat{\theta}(\mathbf{x})$ el valor del parámetro en el que la función de verosimilutud $L(\theta|\mathbf{x})$ toma su máximo valor como función de θ, entonces el *estimador de máxima verosimilitud* del parámetro θ basado en la muestra $\mathbf{X}$ es $\hat{\theta}(\mathbf{X})$.

Si la función de verosimilitud es diferenciable en θ_i, los posibles candidatos para estimadores de máxima verosimilitud son los valores de $(\theta_1, \cdots, \theta_k)$ que resuelven:

$$\frac{\partial}{\partial \theta_i} L(\theta|\mathbf{x}) = 0, \ \ i = 1, \cdots, k \tag{7.2.2}$$

Ejemplo 7.2.3 Sea $X_1, \cdots, X_n$ una muestra aleatoria obtenida desde la distribución $\mathcal{N}(\theta, \sigma^2)$ con $\sigma^2 = 1$. Hallar el estimador de máxima verosimilitud para θ.

En muchos casos es fácil trabajar con el logaritmo natural de la función de verosimilitud denotado por $l(\theta|\mathbf{x})$. Lo anterior es posible debido a que la función *logverosimilitud* es estrictamente decreciente sobre $(0, \infty)$.

Ejemplo 7.2.4 Sea $X_1, \cdots, X_n$ una muestra aleatoria obtenida desde la distribución $\mathcal{B}(p)$. Hallar el estimador de máxima verosimilitud para p.

Teorema 7.2.1 (Propiedad de invariancia) Si $\hat{\theta}$ es el estimador de máxima verosimilitud de θ, entonces para toda función $\tau(\theta)$, su estimador de máxima verosimilitud es $\tau(\hat{\theta})$.

Usando el teorema anterior, se puede establecer que en el problema 7.2.3 el estimador de máxima verosimilitud de θ^2 es $\overline{X}^2$. Además, el estimador de máxima verosimilitud de $\sqrt{p(1-p)}$ es $\sqrt{\overline{X}(1-\overline{X})}$ en el problema 7.2.4.

La propiedad de invariancia también puede aplicarse al caso multivariado. Suponga que el estimador de máxima verosimilitud de $(\theta_1, \cdots, \theta_k)$ es $(\hat{\theta}_1, \cdots, \hat{\theta}_k)$. Si $\tau(\theta_1, \cdots, \theta_k)$ es una función de los parámetros entonces su estimador de máxima verosimilitud es $\tau(\hat{\theta}_1, \cdots, \hat{\theta}_k)$.

Si $\theta = (\theta_1, \cdots, \theta_k)$ es multidimensional se requieren de más condiciones relacionadas con la segunda derivada para hallar el estimador de máxima verosimilitud.

Ejemplo 7.2.5 Sea $X_1, \cdots, X_n$ una muestra aleatoria obtenida desde la distribución $\mathcal{N}(\theta, \sigma^2)$ con ambos parámetros desconocidos. Entonces:

$$L(\theta, \sigma^2|\mathbf{x}) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\theta)^2}$$

$$l(\theta, \sigma^2|\mathbf{x}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\theta)^2$$

las derivadas parciales con respecto a θ y σ^2 son:

$$\frac{\partial}{\partial\theta} l = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\theta) = 0 \quad \text{y} \quad \frac{\partial}{\partial\sigma^2} l = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (x_i-\theta)^2 = 0$$

entonces $\hat{\theta} = \overline{x}$ y $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (x_i - \bar{x})^2$ son los candidatos a estimadores de máxima verosimilitud, pero ¿se trata de un máximo global? Para eso al menos una derivada parcial de segundo orden deber ser negativa:

$$\left.\frac{\partial^2}{\partial\theta^2} l\right|_{\theta=\overline{x}} = \frac{1}{\sigma^2}\sum_{i=1}^n (-1) = -\frac{n}{\sigma^2} < 0$$

y además el Jacobiano debe ser positivo:

$$= \left| \begin{array}{cc} \frac{\partial^2}{\partial\theta^2} l & \frac{\partial^2}{\partial\theta\partial\sigma^2} l \\ \frac{\partial^2}{\partial\theta\partial\sigma^2} l & \frac{\partial^2}{\partial(\sigma^2)^2} l \end{array} \right|_{\theta = \hat{\theta}, \sigma^2 = \hat{\sigma}^2}$$

Luego, se tiene que:

$$
\begin{aligned}
&= \left. \begin{vmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\theta)^2 \\ -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\theta)^2 & \frac{n}{\sigma^4}-\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\theta)^2 \end{vmatrix} \right|_{\theta=\hat{\theta},\sigma^2=\hat{\sigma}^2} \\
&= \left. \frac{n^2}{2\sigma^6} \right|_{\theta=\hat{\theta},\sigma^2=\hat{\sigma}^2} \\
&= \frac{n^2}{2\hat{\sigma}^6} > 0
\end{aligned}
$$

Finalmente, $\hat{\theta} = \overline{X}$ y $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ son los estimadores de máxima verosimilitud.

7.3. Métodos de evaluación de estimadores

Para cada parámetro de interés pueden existir varios estimadores diferentes. En general, se escoge al que posea mejores propiedades que los restantes. En esta sección se presentan algunos criterios básicos para evaluar estimadores.

7.3.1. Error cuadrático medio

Definición 7.3.1 El *error cuadrático medio* del estimador W de un parámetro θ se define por $\mathrm{E}_\theta[(W-\theta)^2]$. Se puede probar que:

$$\mathrm{E}_\theta[(W-\theta)^2] = \mathrm{Var}_\theta(W) + \mathrm{Sesgo}_\theta^2[W] \tag{7.3.1}$$

Definición 7.3.2 El *sesgo* del estimador puntual W del parámetro θ, es la diferencia entre su valor esperado y θ. Es decir, $\mathrm{Sesgo}_\theta[W] = \mathrm{E}_\theta[W] - \theta$. Un estimador cuyo sesgo es cero es llamado *insesgado.* Para estimadores insesgados se tiene $\mathrm{E}_\theta[(W-\theta)^2] = \mathrm{Var}_\theta(W)$.

El error cuadrático medio incorpora un componente que mide la variabilidad del estimador (precisión) y otro componente que mide el sesgo. Un estimador con buenas propiedades es aquel que mantiene tanto la varianza como el sesgo en un valor pequeño o realiza un intercambio efectivo entre ellos.

Ejemplo 7.3.1 Sea $X_1, \cdots, X_n$ una muestra aleatoria obtenida desde la distribución $\mathcal{N}(\mu, \sigma^2)$. Las estadísticas $\overline{X}$ y S^2 son ambos estimadores insesgados ya que:

$$\mathrm{E}[\overline{X}] = \mu \quad \text{y} \quad \mathrm{E}[S^2] = \sigma^2$$

Lo anterior es cierto aún sin el supuesto de normalidad. El error cuadrático medio de estos estadísticos es:

$$\mathrm{E}[(\overline{X} - \mu)^2] = \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

$$\mathrm{E}[(S^2 - \sigma^2)^2] = \mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

un estimador alternativo para σ^2 es el estimador de máxima verosimilitud $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{(n-1)}{n}s^2$, entonces:

$$\mathrm{E}[\hat{\sigma}^2] = \mathrm{E}\left[\frac{n-1}{n}S^2\right] = \frac{n-1}{n}\sigma^2$$

es decir, $\hat{\sigma}^2$ es un estadístico sesgado de σ^2. La variancia de $\hat{\sigma}^2$ puede calcularse como:

$$\mathrm{Var}(\hat{\sigma}^2) = \mathrm{Var}\left(\frac{n-1}{n}S^2\right) = \left(\frac{n-1}{n}\right)^2 \mathrm{Var}(S^2) = \frac{2(n-1)}{n^2}\sigma^4$$

Luego:

$$\begin{aligned}
\mathrm{ECM}[\hat{\sigma}^2] &= \mathrm{E}[(\hat{\sigma}^2 - \sigma^2)^2] \\
&= \mathrm{Var}(\hat{\sigma}^2) + \mathrm{Sesgo}^2[\hat{\sigma}^2] \\
&= \frac{2(n-1)}{n^2}\sigma^4 + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 \\
&= \left(\frac{2n-1}{n^2}\right)\sigma^4
\end{aligned}$$

Finalmente:

$$\mathrm{ECM}[\hat{\sigma}^2] = \left(\frac{2n-1}{n^2}\right)\sigma^4 < \left(\frac{2}{n-1}\right)\sigma^4 = \mathrm{ECM}[S^2]$$

7.3.2. Mejores estimadores insesgados

Una comparación de estimadores basada en el error cuadrático medio no permite establecer claramente cual es mejor. El objetivo de esta sección es obtener un método para encontrar el mejor estimador insesgado.

Definición 7.3.3 Un estimador W^* es el *mejor estimador insesgado* de $\tau(\theta)$ si satisface que $\mathrm{E}_\theta[W^*] = \tau(\theta)$, para todo θ y para cualquier otro estimador W con $\mathrm{E}_\theta[W] = \tau(\theta)$ se tiene $\mathrm{Var}_\theta\,(W^*) \leq \mathrm{Var}_\theta\,(W)$ para todo θ. W^* también es llamado *estimador insesgado uniforme de mínima variancia* de $\tau(\theta)$.

Ejemplo 7.3.2 Sea $X_1, \cdots, X_n$ una muestra aleatoria obtenida desde la distribución $\mathcal{P}(\lambda)$. Recordar que para la función de probabilidad de Poisson, la media y la varianza son iguales a λ. Luego aplicando el teorema 5.2.2 se tiene:

$$\mathrm{E}_\lambda[\overline{X}] = \lambda$$
$$\mathrm{E}_\lambda[S^2] = \lambda$$

es decir que ambos son estimadores insesgados de λ. Para determinar cuál es mejor se comparan las varianzas, aplicando nuevamente el teorema 5.2.2, obteniendose que $\mathrm{Var}_\lambda(\overline{X}) \leq \mathrm{Var}_\lambda(S^2)$. Aún cuando que $\overline{X}$ es mejor que S^2, considere la siguiente clase de estimadores:

$$W_a(\overline{X}, S^2) = a\overline{X} + (1-a)S^2$$

para toda constante a, $\mathrm{E}_\lambda[W_a(\overline{X}, S^2)] = \lambda$, es decir se tienen infinitos estimadores insesgados de λ. La pregunta es, aún siendo $\overline{X}$ mejor estimador que S^2, ¿$\overline{X}$ es mejor que $W_a(\overline{X}, S^2)$ para todo a?

Teorema 7.3.1 (Cramér-Rao) Sea $X_1, \cdots X_n$ una muestra con función de probabilidad o densidad $f(\mathbf{x}|\theta)$ y sea $W(\mathbf{X}) = W(X_1, \cdots, X_n)$ algún estimador donde $\mathrm{E}_\theta[W(\mathbf{X})]$ es una función diferenciable de θ. Suponga que la función de densidad conjunta $f(\mathbf{x}|\theta) = f(x_1, \cdots, x_n|\theta)$ satisface:

$$\frac{d}{d\theta}\int\cdots\int h(\mathbf{x})f(\mathbf{x}|\theta)dx_1\cdots dx_n = \int\cdots\int h(\mathbf{x})\frac{\partial}{\partial\theta}f(\mathbf{x}|\theta)dx_1\cdots dx_n \tag{7.3.2}$$

para cualquier función $h(\mathbf{x})$ con $\mathrm{E}_\theta[|h(\mathbf{X})|] < \infty$. Entonces:

$$\mathrm{Var}_\theta\left(W(\mathbf{X})\right) \geq \frac{\left(\frac{d}{d\theta}\mathrm{E}_\theta[W(\mathbf{X})]\right)^2}{\mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right)^2\right]} \tag{7.3.3}$$

Corolario 7.3.1 (Cramér-Rao caso independiente e identicamente distribuido) Sean $X_1, \cdots, X_n$ independientes e identicamente distribuidos con función de probabilidad o densidad $f(x|\theta)$ y sea $W(\mathbf{X}) = W(X_1, \cdots, X_n)$ cualquier estadística tal que $\mathrm{E}_\theta[W(\mathbf{X})]$ es una función diferenciable de θ. Si la función de densidad conjunta $f(\mathbf{x}|\theta) = \prod f(x_i|\theta)$ satisface 7.3.2 :

$$\mathrm{Var}_\theta\left(W(\mathbf{X})\right) \geq \frac{\left(\frac{d}{d\theta}\mathrm{E}_\theta[W(\mathbf{X})]\right)^2}{n\mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right]}$$

La cantidad $\mathrm{E}_\theta[\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(X_i|\theta))^2]$ es llamada *número de información* o *información de Fisher* de la muestra.

El nombre proviene del hecho que el número de información es la inversa de la variancia del mejor estimador de θ. Cuando el número de información es grande entonces se tiene una variancia pequeña para el mejor estimador, de ahí más información acerca de θ.

Lema 7.3.1 Si $f(x|\theta)$ satisface:

$$\frac{d}{d\theta}\mathrm{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X|\theta)\right] = \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)f(x|\theta)\right)\right]dx$$

lo cual es verdadero para una familia exponencial, entonces:

$$\mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right] = -\mathrm{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right]$$

Ejemplo 7.3.3 Volviendo al ejercicio de la distribución de Poisson:

$$\mathrm{Var}_\lambda\left(W(\mathbf{X})\right) \geq \frac{\left(\frac{d}{d\lambda}\mathrm{E}_\lambda[W(\mathbf{X})]\right)^2}{n\mathrm{E}_\lambda\left[\left(\frac{\partial}{\partial\lambda}\log f(X|\lambda)\right)^2\right]}$$

Si consideramos cualquier estimador insesgado:

$$\begin{aligned}
\mathrm{Var}_\lambda\left(W(\mathbf{X})\right) &\geq \frac{\left(\frac{d}{d\lambda}(\lambda)\right)^2}{n\mathrm{E}_\lambda\left[\left(\frac{\partial}{\partial\lambda}\log f(X|\lambda)\right)^2\right]} \\
&\geq \frac{1}{n\mathrm{E}_\lambda\left[\left(\frac{\partial}{\partial\lambda}\log f(X|\lambda)\right)^2\right]}
\end{aligned}$$

y como la distribución de Poisson pertenece a una familia exponencial:

$$\begin{aligned}
\mathrm{Var}_\lambda\left(W(\mathbf{X})\right) &\geq \frac{1}{-n\mathrm{E}_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log f(X|\lambda)\right]} \\
&\geq \frac{1}{-n(-\frac{1}{\lambda})} \\
&\geq \frac{\lambda}{n}
\end{aligned}$$

Como $\mathrm{Var}_\lambda(\overline{X}) = \lambda/n$, entonces $\overline{X}$ es el mejor estimador insesgado de λ.

Es importante recordar que un supuesto clave en el teorema de Cramér-Rao es la posibilidad de derivar bajo el signo de la integral, algo que es en cierto modo restrictivo pero que se satisface si la densidad pertenece a una familia exponencial.

A pesar de que sea aplicable el teorema de Cramér-Rao esto no garantiza que la varianza del mejor estimador sea igual al limite inferior de la desigualdad.

Ejemplo 7.3.4 Sea $X_1, \cdots, X_n$ una muestra aleatoria obtenida desde la distribución $\mathcal{N}(\mu, \sigma^2)$. Considere el problema de estimar σ^2 cuando μ es conocido. La función de densidad cumple con las condiciones del teorema de Cramér-Rao y el lema 7.3.2, luego:

$$\frac{\partial^2}{\partial(\sigma^2)^2}\log f = \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}$$

$$-\mathrm{E}\left[\frac{\partial^2}{\partial(\sigma^2)^2}\log f\right] = -\mathrm{E}\left[\frac{1}{2\sigma^4} - \frac{(X-\mu)^2}{\sigma^6}\right] = \frac{1}{2\sigma^4}$$

Para todo estimador insesgado W de σ^2 se tiene:

$$\operatorname{Var}(W|\mu,\sigma^2) \geq \frac{2\sigma^4}{n}$$

En el ejemplo 7.3.1 se vió que:

$$\operatorname{Var}(S^2|\mu,\sigma^2) = \frac{2\sigma^4}{n-1}$$

es decir que S^2 no alcanza la cota inferior de Crámer-Rao. La pregunta ahora es si existe algún estimador insesgado de σ^2 cuya varianza alcance la cota mencionada.

Corolario 7.3.2 Sean $X_1, \cdots, X_n$ independientes e identicamente distribuidos como $f(x|\theta)$ tal que satisface las condiciones del teorema Cramér-Rao. Sea $L(\theta|x) = \prod_{i=1}^n f(x_i|\theta)$ denota la función de verosimilitud. Si $W(\mathbf{X}) = W(X_1, \cdots, X_n)$ es cualquier estimador insesgado de $\tau(\theta)$, entonces $W(\mathbf{X})$ alcanza la cota inferior real de Cramér-Rao sí y solo si:

$$\frac{\partial}{\partial\theta}\log L(\theta|\mathbf{x}) = a(\theta)[W(\mathbf{x}) - \tau(\theta)] \tag{7.3.4}$$

para alguna función $a(\theta)$.

Ejemplo 7.3.5 Retomando el ejemplo 7.3.4 se tiene:

$$L(\mu,\sigma^2|\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}$$

y además:

$$\frac{\partial}{\partial\sigma^2}\log L(\mu,\sigma^2|\mathbf{x}) = \frac{n}{2\sigma^4}\left(\sum_{i=1}^n \frac{(x_i-\mu)^2}{n} - \sigma^2\right)$$

Luego:

$$a(\sigma^2)\left(\sum_{i=1}^n \frac{(x_i-\mu)^2}{n} - \sigma^2\right) = \frac{n}{2\sigma^4}\left(\sum_{i=1}^n \frac{(x_i-\mu)^2}{n} - \sigma^2\right)$$

entonces:

$$a(\sigma^2) = \frac{n}{2\sigma^4}$$

Es decir, el mejor estimador insesgado de σ^2 es $\sum(x_i-\mu)^2/n$ y puede ser calculado solo si μ es conocido, en caso contrario no puede alcanzarse la cota.

Lo discutido hasta el momento nos deja algunas preguntas sin respuesta. Primero, ¿qué hacer cuando $f(x|\theta)$ no satisface las condiciones del teorema de Cramér-Rao? y segundo, ¿qué hacer cuando no sea posible alcanzar la cota inferior con los estimadores disponibles?

Se requiere buscar un método que permita estudiar a los mejores estimadores insesgados desde otro punto de vista, usando el concepto de *suficiencia.* Recordar que si X y Y son dos variables aleatorias cualesquiera entonces, siempre que los esperados existan, se tiene:

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}[\mathrm{E}[X|Y]] \\ \mathrm{Var}(X) &= \mathrm{Var}\left(\mathrm{E}[X|Y]\right) + \mathrm{E}[\mathrm{Var}(X|Y)] \end{aligned}$$

Los resultados anteriores permiten establecer el siguiente teorema.

Teorema 7.3.2 (Rao-Blackwell) Sea W un estimador insesgado de $\tau(\theta)$ *y* T una estadística suficiente para θ. Se define $\phi(T) = \mathrm{E}[W|T]$. Entonces $\mathrm{E}_\theta[\phi(T)] = \tau(\theta)$ y $\mathrm{Var}_\theta\left(\phi(T)\right) \leq \mathrm{Var}_\theta\left(W\right)$ para todo θ, es decir, $\phi(T)$ es el mejor estimador insesgado uniforme de $\tau(\theta)$.

Para hallar el mejor estimador de $\tau(\theta)$ se requiere considerar solamente estimadores basados en estadísticos suficientes.

Teorema 7.3.3 Si W es el mejor estimador insesgado de $\tau(\theta)$, entonces W es único.

Ejemplo 7.3.6 Sea $X_1, \cdots, X_n$ una muestra aleatoria obtenida desde la distribución $\mathcal{BI}(k,\theta)$ donde k es conocido. Hallar el mejor estimador insesgado para la probabilidad de obtener un éxito, es decir:

$$\tau(\theta) = \Pr_\theta(X=1) = k\theta(1-\theta)^{k-1}$$

El siguiente teorema permite establecer un vínculo entre los conceptos de suficiencia, completez y unicidad.

Teorema 7.3.4 (Lehmann-Scheffé) Sea T una estadística suficiente y completa para el parámetro θ y $\phi(T)$ un estimador basado solo en T. Entonces $\phi(T)$ es el único mejor estimador insesgado de su valor esperado.

Se presenta la aplicación práctica del teorema anterior. Si T es una estadística suficiente y completa para el parámetro θ y $h(X_1, \cdots, X_n)$ un estimador insesgado de $\tau(\theta)$, entonces $\phi(T) = E(h(X_1, \cdots, X_n)|T)$ es el único mejor estimador insesgado de $\tau(\theta)$.

Ejemplo 7.3.7 Sea $X_1, \cdots, X_n$ una muestra aleatoria obtenida desde la distribución:

$$f(x|\theta) = \theta(1+x)^{-(1+\theta)} \qquad x > 0$$

para $\theta > 0$. Hallar el mejor estimador insesgado para θ^{-1}.

7.3.3. Consistencia

Definición 7.3.4 Una secuencia de estimadores $W_n = W_n(X_1, \cdots, X_n)$ es una *secuencia consistente de estimadores* del parámetro θ si para todo $\epsilon > 0$ y todo $\theta \in \Theta$:

$$\lim_{n\to\infty} \Pr_\theta(| W_n - \theta |< \epsilon) = 1 \tag{7.3.5}$$

Ejemplo 7.3.8 Sean $X_1, \cdots, X_n$ independientes e identicamente distribuidas según $\mathcal{N}(\theta, 1)$. Considere la secuencia:

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

recodando que $\overline{X}_n \sim \mathcal{N}(\theta, \frac{1}{n})$ se tiene:

$$\begin{aligned}
\Pr_\theta\left(| \overline{X}_n - \theta |< \epsilon\right) &= \int\limits_{\bar{x}_n=\theta-\epsilon}^{\bar{x}_n=\theta+\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{n}{2}(\bar{x}_n-\theta)^2} d\bar{x}_n \\
&= \int\limits_{y=-\epsilon}^{y=\epsilon} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{n}{2}y^2} dy \quad (y = \overline{x}_n - \theta) \\
&= \int\limits_{t=-\epsilon\sqrt{n}}^{t=\epsilon\sqrt{n}} \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}t^2} dt \quad (t = y\sqrt{n}) \\
&= \Pr_\theta(-\epsilon\sqrt{n} < Z < \epsilon\sqrt{n}) \to 1
\end{aligned}$$

cuando $n \to \infty$. Luego $\overline{X}_n$ es una secuencia consistente de estimadores de θ.

Teorema 7.3.5 Si W_n es una secuencia de estimadores del parámetro θ que satisfacen

a. $\lim_{n\to\infty} \operatorname{Var}_\theta[W_n] = 0$.

b. $\lim_{n\to\infty} \operatorname{Sesgo}_\theta[W_n] = 0$.

entonces W_n es una secuencia consistente de estimadores de θ.

Ejemplo 7.3.9 Como:

$$\mathrm{E}_\theta[\overline{X}_n] = \theta \;\text{ y }\; \operatorname{Var}_\theta(\overline{X}_n) = \frac{1}{n}$$

satisfacen las condiciones del teorema 7.3.5, luego la secuencia $\overline{X}_n$ es consistente. Además, del teorema 5.2.2, $\overline{X}_n$ proveniene de un muestreo independiente e identicamente distribuido de cualquier población con media θ por lo que es consistente para dicho parámetro siempre que la varianza sea finita.

Ejemplo 7.3.10 Sea $X_1, \cdots, X_n$ una muestra aleatoria obtenida desde la distribución con función de densidad:

$$f(x|\theta) = \frac{x}{\theta} \exp\left\{-\frac{x^2}{2\theta}\right\} \qquad x > 0$$

¿Es $T = \frac{1}{2n}\sum_{i=1}^{n} X_i^2$ un estimador consistente de θ?

Teorema 7.3.6 Si W_n una secuencia consistente de estimadores del parámetro θ. Sean $a_1, a_2, \cdots$ y $b_1, b_2, \cdots$ secuencias de constantes que satisfacen:

a. $\lim_{n\to\infty} a_n = 1$.

b. $\lim_{n\to\infty} b_n = 0$.

entonces, $U_n = a_n W_n + b_n$ es una secuencia consistente de estimadores de θ.

Teorema 7.3.7 (Consistencia de los estimadores de máxima verosimilitud) Sean $X_1, \cdots, X_n$ variables aleatorias independientes e identicamente distribuidas obtenidas de $f(x|\theta)$, y sea $L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta)$ la función de verosimilitud. Sea $\hat{\theta}$ el estimador de máxima verosimilitud de θ y $\tau(\theta)$ una función continua de θ. Bajo ciertas condiciones de reguralidad sobre $f(x|\theta)$, y por consiguiente $L(\theta|\mathbf{x})$, para todo $\epsilon > 0$ y $\theta \in \Theta$:

$$\lim_{n\to\infty} \Pr_\theta\left(\mid \tau(\hat{\theta}) - \tau(\theta) \mid\geq \epsilon\right) = 0$$

es decir $\tau(\hat{\theta})$ es un estimador consistente de $\tau(\theta)$.

7.4. Otras consideraciones

7.4.1. Variancia asintótica de los estimadores de máxima verosimilitud

Definición 7.4.1 Una secuencia de estimadores W_n es *asintóticamente eficiente* para un parámetro $\tau(\theta)$ si:

$$\lim_{n\to\infty} \frac{\mathrm{Var}_\theta\left(W_n\right)}{\left[\frac{[\tau'(\theta)]^2}{n\mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right]}\right]} = 1$$

es decir W_n alcanza la cota inferior de Crámer-Rao conforme $n \to \infty$.

7.4.2. Aproximación por series de Taylor

Definición 7.4.2 Si una función $g(x)$ tiene derivadas de orden r, es decir que existe $g^{(r)}(x) = \frac{d^r}{dx^r}g(x)$, entonces para cualquier constante a, la *polinomial de Taylor de orden r alrededor de a* es:

$$T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(x-a)^i$$

Teorema 7.4.1 (Taylor) Si $g^{(r)}(a) = \frac{d^r}{dx^r}g(x)\mid_{x=a}$ existe, entonces:

$$\lim_{x\to a} \frac{g(x)-T_r(x)}{(x-a)^r} = 0$$