Índice general

Capítulo 8

Principios de Estimación Puntual

1. Introducción

De la definición de la suficiencia, es fácil ver que los datos $\mathbf{X}$ es en sí mismo siempre es suficiente. Por lo tanto la suficiencia no sería un concepto particularmente útil a menos que pudiéramos encontrar estadísticos suficientes que realmente representen una reducción de los datos; Sin embargo, a partir de los ejemplos que se han dado en lección anterior, se puede ver que esto es realmente posible. Por lo tanto, el problema real reside en la determinación de sí un estadístico suficiente representa la mejor reducción posible de los datos.
Hay dos nociones de lo que significa la "mejor reducción" de los datos. El primero de ellos es el concepto de la suficiencia mínima; un estadístico suficiente $\mathbf{T}$ es estadístico suficiente mínimo si para cualquier otra estadístico suficiente S, existe una función g tal que $\mathbf{T} = g(S)$. De esta manera, un estadístico suficiente y míni-

mo (minimal) es el estadístico suficiente que representa la máxima reducción de los datos que contiene la mayor cantidad de información acerca del parámetro desconocido como los datos mismos. La segunda noción (y más fuerte) es el concepto de la *completud* del estimador que se discutió en la lección anterior. Si $\mathbf{X} \sim F_\theta$, entonces un estadístico $\mathbf{T} = \mathbf{T}(\mathbf{X})$ es completo si $\mathbf{E}_\theta[g(\mathbf{T})] = 0$ para todo $\theta \in \Theta$ implica que $\mathbf{P}_\theta(g(\mathbf{T}) = 0) = 1$ para todos $\theta \in \Theta$. En particular, si $\mathbf{T}$ es completo entonces $g(\mathbf{T})$ es un estadístico ancillary (no esencial) para θ solo si $g(\mathbf{t})$ es constante; por tanto, un estadístico completo, $\mathbf{T}$, no contiene información ancillary.
Se puede demostrar que si un estadístico $\mathbf{T}$ es suficiente y completa entonces $\mathbf{T}$ es también un estadístico suficiente minimal; Sin embargo, lo contrario no es cierto. Por ejemplo, suponga que $X_1, X_2, \ldots, X_n$ son variables aleatorias iid. cuya función de densidad es

$$f(\mathbf{x}; \theta) = \frac{\exp(\mathbf{x} - \theta)}{[1 + \exp(\mathbf{x} - \theta)]^2}$$

Para este modelo, no existe un estadístico suficiente unidimensional para θ y, de hecho, los estadísticos de orden $(X_{(1)}, \cdots, X_{(\mathbf{n})})$ son estadístico suficientes minimales (pruébelo). Sin embargo, el estadístico $\mathbf{T} = X_{(n)} - X_{(1)}$ es accesorio (ancillary) y así los estadísticos de orden no son completas. Así, a pesar de que los estadísticos de orden,$(X_{(1)}, \cdots, X_{(\mathbf{n})})$ son unos estadísticos suficientes y mínimos, todavía contiene la información "redundante" sobre θ

¿Qué tan importante es la suficiencia en la práctica?. La discusión anterior sugiere que cualquier procedimiento estadístico debe depender solo del estadístico suficiente minimal. De hecho, se verá en las lecciones siguientes que procedimientos estadísticos óptimos (estimadores puntuales, pruebas de hipótesis y así sucesivamente discutidos en esta lección) casi siempre dependen de los estadísticos suficientes minimales. No obstante, los modelos estadísticos sirven realmente solo como aproximaciones a la realidad y por lo que los procedimientos que son nominalmente óptimas puede fracasar miserablemente en la práctica. Por ejemplo, supongamos que $X_1, X_2, \ldots, X_\mathbf{n}$ son variables aleatorios iid. con media μ y varianza σ^2. Es común suponer que las variables X_i tengan una distribución normal, en cuyo caso $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ son estadísticos suficientes y mínimos para (μ, σ^2). Sin embargo, los procedimientos óptimos para la distribución normal puede fallar miserablemente si cada una de las X_i no son normales. Por esta razón, es importante ser flexible en el desarrollo de los métodos estadísticos.

2. Estimación Puntual

Sea $(\Omega, \mathscr{A}, \mathbf{P}_\theta)$ un espacio de probabilidad, $\mathscr{P} = \{\mathbf{P}_\theta : \theta \in \Theta\}$, el modelo paramétrico, $X_1, X_2, \ldots, X_n : (\Omega, \mathscr{A}) \longrightarrow (\mathbb{R}^m, \mathcal{B}^m)$ variables aleatorias, $(\mathscr{X}, \mathcal{B}_\mathscr{X})$ es el espacio muestral, $\varphi : \Theta \longrightarrow \mathbb{R}^P$ una función medible, es decir $\varphi : (\Theta, \mathcal{B}_\Theta) \longrightarrow (\varphi(\Theta), \mathcal{B}_\varphi)$

Definición 2.1. Un **estimador T** es una función medible

$$\mathbf{T} : (\mathscr{X}, \mathcal{B}_\mathscr{X}) \longrightarrow (\varphi(\Theta), \mathcal{B}_\varphi).$$

Naturalmente, se espera que $\mathbf{T}(\mathbf{X})$ tiende a estar cerca del estimando desconocido $\varphi(\theta)$, pero este requisito no forma parte de una definición formal de un estimador.

Una palabra sobre la *fraseología* : un **estimador** es una variable aleatoria $\mathbf{T}(X_1, X_2, \ldots, X_n)$ en discusión antes de que se haya observado una muestra, mientras que una **estimación** es un valor $\mathbf{T}(x_1, x_2, \ldots, x_n)$ después de que se haya observado la muestra.

La pregunta natural es, “¿Qué estadística debemos usar para estimar $\varphi(\theta)$?” Por supuesto. La respuesta es, “el mejor” pero entonces tenemos que explicar lo que queremos decir con “el mejor”, y preguntar, ¿Existen tales estimadores?.

Las propiedades deseables de un estimador son; Insesgabilidad, Consistencia (fuerte, débil, en la r-ésima media), Suficiencia, Normalidad asintótica, Suficiencia mínima, completitud, Invarianza, etc.

A continuación, nos interesan los estimadores insesgados y aprenderemos sobre otro criterio estadístico: **la eficiencia.**

Definición 2.2. Sea $\varphi : \Theta \longrightarrow \mathbb{R}^m$ una función medible, Una estadística $\mathbf{T} : (\mathscr{X}, \mathcal{B}_{\mathscr{X}}) \longrightarrow (\mathbb{R}^m, \mathcal{B}_m)$ se llama insesgado, si

$$\mathbf{E}_\theta(\mathbf{T}) = \varphi(\theta) \quad \forall\, \theta \in \Theta.$$

Cada función φ en Θ, para la cual existe un estimador insesgado, se denomina **función estimable.**

Definición 2.3. Para un estimador **sesgado,**

$$b(\varphi(\theta), \mathbf{T}) = \mathbf{E}_\theta(\mathbf{T}) - \varphi(\theta)$$

se llama **sesgo** del estimador.

Definición 2.4. Un estimador $\mathbf{T}$ se llama **asintóticamente insesgado** para $\varphi(\theta)$ si

$$\lim_{n\to\infty} b(\varphi(\theta), \mathbf{T}_n) = 0$$

Observación 2.1. Si $\mathbf{T}$ es insesgado para $\varphi(\theta)$, entonces en general $g(\mathbf{T})$ es sesgado para $\varphi(\theta)$, a menos que g sea una función lineal.

Observación 2.2. No siempre existen estimadores insesgados.

Observación 2.3. Los estimadores insesgados no siempre son razonables.

En muchos textos son comunes el uso de los siguientes simbolismos:
$X_1, X_2, \ldots, X_{\mathbf{n}}$ es una muestra aleatoria de una distribución (población) con
media poblacional finita $\mu = \mathbf{E}(X)$
varianza poblacional finita $\sigma^2 = \mathbf{E}[(X-\mu)^2]$.
La media muestral y la varianza muestral se define como

$$\overline{X} = \frac{1}{\mathbf{n}}\sum_{i=1}^{\mathbf{n}} X_i \qquad \text{y} \qquad S^2 = \frac{1}{\mathbf{n}-1}\sum_{i=1}^{\mathbf{n}}(X_i - \overline{X})^2$$

Ejercicio 1.
Demuestre que:

(a) $\mathbf{E}(\overline{X}) = \mu$
(b) $\mathbf{E}(X^2) = \sigma^2 + \mu^2$
(c) $\mathbf{E}(\overline{X}^2) = \frac{\sigma^2}{\mathbf{n}} + \mu^2$

Ejemplo 2.1. Sea $f_\theta(x) = \theta^x(1-\theta)^{1-x} I_{\{0,1\}}(x), 0 < \theta < 1$. Sea $\mathbf{T}(X_1, X_2, \ldots, X_{\mathbf{n}}) = \overline{X}$ y $\varphi(\theta) = \theta$. Entonces,

$$\mathbf{E}(\mathbf{T}) = \frac{1}{\mathbf{n}}\sum_{i=1}^{\mathbf{n}} \mathbf{E}_\theta(X_i) = \theta \qquad \forall\, \theta \in (0,1)$$

Luego $\mathbf{T}$ es un estimador insesgado de θ. ■

Ejemplo 2.2. Para el modelo $\mathscr{P} = \{f_\theta(x), x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}\}$. Sea $\varphi(\theta) = F_\theta(\mathbf{t}) = \mathbf{P}_\theta(X \leq \mathbf{t}) = \mathbf{E}_\theta[I(X \leq \mathbf{t})]$ para un $\mathbf{t} \in \mathbb{R}$ fijo.

La *función distribución empírica (muestral)* se define por

$$F_{\mathbf{n}}(s) = \frac{1}{\mathbf{n}} \sum_{i=1}^{\mathbf{n}} I(X_i \leq s) \quad \text{para } s \in \mathbb{R}$$

donde $I(X_i \leq s)$ es la función indicadora definida por

$$I(X \leq s) = \begin{cases} 1 & \text{si} \quad X \leq s \\ 0 & \text{si} \quad X > s \end{cases}$$

Considere es estimador $\mathbf{T}(X_1, X_2, \ldots, X_{\mathbf{n}}) = F_{\mathbf{n}}(\mathbf{t})$. Entonces,

$$\mathbf{E}(T) = \mathbf{E}_\theta[F_{\mathbf{n}}(\mathbf{t})] = \frac{1}{\mathbf{n}} \sum_{i=1}^{\mathbf{n}} \mathbf{E}_\theta[I(X_i \leq \mathbf{t})] = F_\theta(\mathbf{t}) \,\forall\, \theta \in \Theta.$$

La función distribución empírica $F_{\mathbf{n}}$ es un estimador insesgado de la función distribución poblacional F_θ. ■

Ejemplo 2.3. Para cualquier distribución para la cual la media poblacional $\mu = \mathbf{E}_\theta(X)$ sea finita, considere la media muestral $\mathbf{T}(X_1, X_2, \ldots, X_{\mathbf{n}}) = \overline{\mathbf{X}}$ entonces $\mathbf{E}_\theta(\mathbf{T}) = \mathbf{E}_\theta(X)$; esto es, la media muestral $\overline{\mathbf{X}}$ es un estimador insesgado de la media poblacional. ■

Ejemplo 2.4. Para cualquier distribución para la cual la varianza poblacional

$$\sigma^2 = \varphi(\theta) = \mathbf{V}ar_\theta(X) = \mathbf{E}_\theta[(X - \mathbf{E}_\theta(X))^2]$$

sea finita, considere el estadístico $\mathbf{T}(X_1, X_2, \ldots, X_\mathbf{n}) = S^2$. Como

$$\begin{aligned}
\mathbf{T} &= \frac{1}{\mathbf{n}-1}\left\{\sum_{i=1}^{\mathbf{n}} X_i^2 - \mathbf{n}\overline{\mathbf{X}}^2\right\} \\
\mathbf{E}_\theta(\mathbf{T}) &= \frac{1}{\mathbf{n}-1}\left\{\sum_{i=1}^{\mathbf{n}} \mathbf{E}_\theta(X_i^2) - \mathbf{n}\mathbf{E}_\theta(\overline{\mathbf{X}}^2)\right\} \\
&= \frac{1}{\mathbf{n}-1}\left\{\mathbf{n}[\mathbf{V}ar_\theta(X) + (\mathbf{E}_\theta(X))^2] - \mathbf{n}\left[\frac{\mathbf{V}ar_\theta(X)}{\mathbf{n}} + (\mathbf{E}_\theta(X))^2\right]\right\} \\
&= \frac{1}{\mathbf{n}-1}\left\{\mathbf{n}\mathbf{V}ar_\theta(X) - \mathbf{V}ar_\theta(X)\right\} \\
&= \mathbf{V}ar_\theta(X) = \sigma^2
\end{aligned}$$

Este demuestra que varianza muestral S^2 es un estimador insesgado de la varianza poblacional. ■

Ejercicio 2.

¿Los siguientes estimadores son insesgados para la función de parámetro $\varphi(\theta)$ dada?

(a) La función de densidad está dada por

$$f_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad \text{y} \quad \theta = (\mu, \sigma^2)$$

$\varphi(\theta) = \sigma^4$; $\mathbf{T}(X_1, X_2, \ldots, X_\mathbf{n}) = \frac{\mathbf{n}+1}{\mathbf{n}-1} S^4$

(b) Considere el modelo Bernoulli y $\varphi(\theta) = \theta(\theta)$. Sean

$$\mathbf{T}_1(X_1, X_2, \ldots, X_\mathbf{n}) = \frac{\sum_{i=1}^{\mathbf{n}} X_i \left[\mathbf{n} - \sum_{i=1}^{\mathbf{n}} X_i\right]}{\mathbf{n}(\mathbf{n}-1)}; \mathbf{T}_2 = \overline{\mathbf{X}}(1-\overline{\mathbf{X}}).$$

(c) La densidad es $f_\theta(x) = \frac{1}{\theta} e^{-\frac{1}{\theta}x} I(0 < x < \infty)$ para $\theta > 0$ y $\varphi(\theta) = \theta^2$. $\mathbf{T}_1 = S^2$; $\mathbf{T}2 = \overline{X}^2$

3. Error cuadrado medio

Si se usa un estadístico $\mathbf{T}$ como un estimador para $\varphi(\theta)$, cuando $\varphi(\theta)$ es una función real, entonces cada valor $(\mathbf{t}-\varphi(\theta))$ es un "error".

Definición 3.1. El error cuadrado medio absoluto (MAE) (en español **ECMA**) de $\mathbf{T}$ se define por

$$\mathbf{ECMA}_\theta(\mathbf{T}) = \mathbf{E}_\theta\left[|\mathbf{T}-\varphi(\theta)|\right]$$

Definición 3.2. El error cuadrado medio (MSE) (en español ECM) de $\widehat{\theta}$ se define por

$$\mathbf{ECM}_\theta(\mathbf{T}) = \mathbf{E}_\theta\left[(\mathbf{T}-\varphi(\theta))^2\right];$$

es fácil demostrar que $\mathbf{ECM}_\theta(\mathbf{T}) = Var_\theta(\mathbf{T}) + [b(\varphi(\theta),\mathbf{T})]^2$.

El sesgo de $\mathbf{T}$ da alguna indicación si la distribución de la muestra se centra alrededor de $\varphi(\theta)$ mientras $\mathbf{ECMA}_\theta\,(\mathbf{T})$ y $\mathbf{ECM}_\theta\,(\mathbf{T})$ son medidas de la dispersión de la distribución muestral de $\mathbf{T}$ alrededor de $\varphi(\theta)$. El error cuadrado medio absoluto, **ECMA**, y el error cuadrado medio, **ECM**, son medidas adecuadas para la comparación de diferentes estimadores de una función paramétrica $\varphi(\theta)$; como debería ser conveniente que $\mathbf{T}$ esté cerca de $\varphi(\theta)$, es natural preferir estimadores con valores pequeños de **ECMA** o **ECM**. Aunque **ECMA** puede parecer ser una mejor medida para evaluar la exactitud de un estimador, **ECM** habitualmente se prefiere en lugar de **ECMA**. Hay varias razones para preferir **ECM**; la mayoría de éstos se derivan de la descomposición de $\mathbf{ECM}_\theta(\mathbf{T})$ en los componentes de varianza y sesgo:

$$\mathbf{ECM}_\theta(\mathbf{T}) = Var_\theta(\mathbf{T}) + [b(\varphi(\theta),\mathbf{T})]^2.$$

Esta descomposición hace que **ECM** sea mucho más fácil de trabajar que el **MAE**. Por ejemplo, cuando **T** es una función lineal de $(X_1, X_2, \ldots, X_{\mathbf{n}})$, la media y la varianza de **T** (y por tanto su **ECM**) se calculan fácilmente; el cálculo de la **ECMA** es mucho más difícil. Con frecuencia, la distribución muestral de un estimador es aproximadamente normal; Por ejemplo, a menudo es cierto que la distribución de **T** es aproximadamente normal con media θ y varianza media $\sigma^2(\theta)/\mathbf{n}$. En tales casos, la varianza $\sigma^2(\theta)/\mathbf{n}$ se aproxima a menudo razonablemente bien por $\mathbf{ECM}_\theta\,(\mathbf{T})$ por lo que el **ECM** caracteriza esencialmente la dispersión de la distribución muestral de **T**. (Típicamente, el componente de la varianza del **ECM** es mucho más grande que la componente sesgo y así $\mathbf{MSE}_\theta\,(\mathbf{T}) \approx Var_\theta\,(\mathbf{T})$.) Sin embargo, también es importante tener en cuenta que el **ECM** de un estimador puede ser infinito, incluso cuando su distribución muestral sea aproximadamente normal.

Ejemplo 3.1. Suponga que $X_1, X_2, \ldots, X_{\mathbf{n}}$ son variables aleatorias normales independientes e idénticamente distribuidas como normales, $\mathscr{N}(\mu, \sigma^2)$. Un estimador insesgado de σ^2 es

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{\mathbf{n}}(X_i - \overline{X})^2$$

■

Sin embargo, si $S = \sqrt{S^2}$, la desviación estándar muestral, no es un estimador insesgado para desviación estándar poblacional, σ; usando el hecho de que,

$$Y = \frac{(\mathbf{n}-1)S^2}{\sigma^2} \sim \chi^2(\mathbf{n}-1)$$

resulta que,

$$S = \frac{\sigma}{\sqrt{(\mathbf{n}-1)}}Y^{\frac{1}{2}}$$

y tomando la esperanza, se tiene

$$\mathbf{E}_\sigma(S) = \frac{\sigma}{\sqrt{(\mathbf{n}-1)}}\mathbf{E}(Y^{\frac{1}{2}})$$

y como

$$Y \sim \chi^2(\mathbf{n}-1) \Rightarrow f_Y(y) = \frac{1}{2^{\frac{\mathbf{n}-1}{2}}\Gamma(\frac{\mathbf{n}-1}{2})} y^{\frac{\mathbf{n}-1}{2}-1} e^{-\frac{1}{2}y}$$

se tiene

$$\begin{aligned}
\mathbf{E}(Y^{\frac{1}{2}}) &= \int_0^\infty y^{\frac{1}{2}} \frac{1}{2^{\frac{\mathbf{n}-1}{2}}\Gamma(\frac{\mathbf{n}-1}{2})} y^{\frac{\mathbf{n}-1}{2}-1} e^{-\frac{1}{2}y} dy \\
\mathbf{E}(Y^{\frac{1}{2}}) &= \frac{\sqrt{2}\Gamma(\mathbf{n}/2)}{\Gamma((\mathbf{n}-1)/2)}
\end{aligned}$$

finalmente resultando,

$$\mathbf{E}_\sigma(S) = \frac{\sigma}{\sqrt{(\mathbf{n}-1)}} \frac{\sqrt{2}\,\Gamma(\mathbf{n}/2)}{\Gamma((\mathbf{n}-1)/2)} \neq \sigma.$$

Lo que indica que la desviación estándar muestral, S, no es un estimador insesgado para la desviación estándar poblacional σ. Sin embargo, cuando $\mathbf{n}$ tiende para infinito, $\mathbf{n} \to \infty$, es posible demostrar que $E(S) \to \sigma$.

Ejemplo 3.2. Suponga que $X_1, X_2, \ldots, X_\mathbf{n}$ es un vector aleatorio cuyos componentes son variables aleatorias iid con una función de distribución uniforme en $[0, \theta]$. Sea $\mathbf{T} = X_{(\mathbf{n})}$, el estadístico de orden máximo; la función de densidad de $\mathbf{T}$ es

$$f_\mathbf{T}(t; \theta) = \frac{\mathbf{n}}{\theta^\mathbf{n}} t^{\mathbf{n}-1} \qquad \text{para} \qquad 0 \leq t \leq \theta.$$

Note que $\mathbf{T} \leq \theta$ y en consecuencia $\mathbf{E}_\theta\,(\mathbf{T}) \leq \theta$; de hecho es fácil de probar que

$$\mathbf{E}_\theta(\mathbf{T}) = \frac{\mathbf{n}}{\mathbf{n}+1}\theta.$$

La forma de $\mathbf{E}_\theta\left(\mathbf{T}\right)$ permite construir fácilmente un estimador insesgado de θ. Si se define $\mathbf{T}^* = \frac{\mathbf{n+1}}{\mathbf{n}}\mathbf{T}$ entonces evidentemente $\mathbf{T}^*$ es un estimador insesgado de θ. ■

Suponga que $\widehat{\theta}_{\mathbf{n}}$ es un estimador de algún parámetro θ basado en n variables aleatorias $X_1,\ldots,X_{\mathbf{n}}$. Cuando $\mathbf{n}$ tiende a crecer, parece razonable esperar que la distribución muestral de $\widehat{\theta}_{\mathbf{n}}$ debería ser cada vez más concentrada en torno al valor verdadero del parámetro θ. Esta propiedad de la sucesión de estimadores $\{\hat{\theta}_{\mathbf{n}}\}$ se conoce como consistencia.

4. Estimadores insesgados uniformes de varianza mínima (EIUMV)

A continuación, consideramos solo El error cuadrado medio y usamos la varianza como una medida de la calidad de varios estimadores insesgados. En este sentido, el "**mejor**" estimador es dada por la siguiente definición:

En lo sigue se considera el caso en que $\Theta \subset \mathbb{R}$.

Definición 4.1. Sea $\varphi(\theta)$ una función estimable y sea $\mathcal{T}$ el conjunto de todos los estimadores insesgados φ, con $\mathbf{E}_\theta(\mathbf{T}^2) < \infty$ para $\forall\,\theta \in \Theta$, es decir, $\mathcal{T} = \{\mathbf{T} : \mathbf{E}_\theta(\mathbf{T}) = \varphi(\theta) \wedge \mathbf{E}_\theta(\mathbf{T}^2) < \infty, \forall \theta \in \Theta\}$. El estadístico $\mathbf{T} \in \mathcal{T}$ se llama estimador insesgado uniforme de varianza mínima (UMVUE) (EIUMV en español) de $\varphi(\theta)$ si

(a) $\mathbf{E}_\theta(\mathbf{T}) < \infty$ para todo $\theta \in \Theta$
(b) Para cualquier $U \in \mathcal{T}$, $\mathbf{V}ar_\theta(T) \leq \mathbf{V}ar_\theta(U)$ para todo $\theta \in \Theta$.

Para identificar un EIUMV es útil tener otro criterio tal como establece el siguiente Teorema.

Teorema 4.1. Sea $\mathcal{T}_0$ el conjunto de todos los estimadores de cero, esto es, $\mathcal{T}_0 = \{\mathbf{S} : \mathbf{E}_\theta(\mathbf{S}) = 0 \,\forall\, \theta \in \Theta\}$. Un estimador insesgado $\mathbf{T}$ de $\varphi(\theta)$ es un EIUMV si y solo si para todo $\mathbf{S} \in \mathcal{T}_0$, $\mathbf{E}_\theta(\mathbf{TS}) = 0$ para todo $\theta \in \Theta$

Prueba. Primero; si $\mathbf{E}_\theta(\mathbf{T}^2) = 0$ para todo $\theta \in \Theta$, entonces $\mathbf{T} = 0$ casi seguramente. Se deduce que $\mathbf{E}_\theta(\mathbf{TS}) = 0$ y $\varphi(\theta)$ también.
Condición de suficiencia: Asumamos que se cumple $\mathbf{E}_\theta(\mathbf{TS}) = 0$ para todo $\theta \in \Theta$. Sea $\mathbf{T}' \in \mathcal{T}$ (es decir otro estimador insesgado de $\varphi(\theta)$): Entonces, como $\mathbf{E}_\theta(\mathbf{T} - \mathbf{T}') = 0$ para todo $\theta \in \Theta$, entonces $\mathbf{T} - \mathbf{T}' \in \mathcal{T}_0$ de manera que

$$\mathbf{E}_\theta[\mathbf{T}(\mathbf{T} - \mathbf{T}')] = 0 \Leftrightarrow \mathbf{E}_\theta(\mathbf{T}^2) = \mathbf{E}_\theta(\mathbf{TT}')$$

Aplicando la desigualdad de Cauchy-Schwarz resulta

$$\mathbf{E}_\theta(\mathbf{T}^2) = \mathbf{E}_\theta(\mathbf{TT}') \leq (\mathbf{E}_\theta(\mathbf{T}^2))^{\frac{1}{2}}(\mathbf{E}_\theta(\mathbf{T}'^2))^{\frac{1}{2}}$$

lo cual es equivalente a

$$[\mathbf{E}_\theta(\mathbf{T}^2)]^{\frac{1}{2}} \leq [\mathbf{E}_\theta(\mathbf{T}'^2)]^{\frac{1}{2}}$$

consecuentemente

$$\mathbf{V}ar_\theta(\mathbf{T}) \leq \mathbf{V}ar_\theta(\mathbf{T}')$$

Condición de necesidad: Ahora asumamos que $\mathbf{T}$ es un EIUMV. Sea $\mathbf{S} \in \mathcal{T}_0$. para todo $\lambda \in \mathbb{R}$, $\mathbf{T} + \lambda\mathbf{S}$ también es un estimador insesgado de $\varphi(\theta)$. Resulta que

$$\mathbf{V}ar_\theta(\mathbf{T} + \lambda\mathbf{S}) = \mathbf{V}ar_\theta(\mathbf{T}) + \lambda^2\mathbf{V}ar_\theta(\mathbf{S}) + 2\lambda\mathbf{E}_\theta[\mathbf{S}(\mathbf{T} - \varphi(\theta))]. \tag{4.1}$$

Como $\mathbf{T}$ es un EIUMV, se tiene

$$\mathbf{V}ar_{\vartheta}(\mathbf{T}) \leq \mathbf{V}ar_{\theta}[\mathbf{T} + \lambda \mathbf{S}] \tag{4.2}$$

De las ecuaciones (4.1) y (4.2) se deduce que

$$\lambda^2 \mathbf{V}ar_{\theta}(\mathbf{S}) + 2\lambda \mathbf{E}_{\theta}[\mathbf{S}(\mathbf{T} - \varphi(\theta))] \geq 0, \forall \lambda \in \mathbb{R}.$$

Completando al cuadrado en λ, resulta

$$\left(\lambda + \frac{\mathbf{E}_{\theta}[\mathbf{S}(\mathbf{T} - \varphi(\theta))]}{\mathbf{V}ar_{\theta}(\mathbf{S})}\right)^2 \geq \left[\frac{\mathbf{E}_{\theta}[\mathbf{S}(\mathbf{T} - \varphi(\theta))]}{\mathbf{V}ar_{\theta}(\mathbf{S})}\right]^2,$$

por decir, $(\lambda + r)^2 \geq r^2$ para todo λ. Pero, esta desigualdad es falsa para $-r < \lambda < -\frac{r}{2}$ cuando $r > 0$ y para $-\frac{r}{2} < \lambda < -r$ cuando $r < 0$.
Resulta que $\mathbf{E}_{\theta}[\mathbf{S}(\mathbf{T} - \varphi(\theta))]$ debe ser igual a cero, en consecuencia resulta

$$\mathbf{E}_{\theta}[\mathbf{S}(\mathbf{T} - \varphi(\theta))] = 0 \Rightarrow \mathbf{E}_{\theta}[\mathbf{T}S] = \mathbf{E}_{\theta}[\mathbf{S}\varphi(\theta)] = \mathbf{E}_{\theta}[\mathbf{S}]\varphi(\theta) = 0$$

$\square$

Ejemplo 4.1. Recuerde que el modelo Binomial con parámetro $\mathbf{n}$ y θ es completo. Esto significa que el único estimador insesgado de 0 es es $\mathbf{S}(X) = 0$. Por lo tanto, $\mathbf{E}[g(X)\mathbf{S}(X)] = 0$ para todo g, en particular $g(\mathbf{X}) = \mathbf{T}(X) = \frac{X}{\mathbf{n}}$. En mérito del Teorema anterior (Teorema 4.1), $\mathbf{T}(\mathbf{X}) = \frac{\mathbf{X}}{\mathbf{n}}$ es el EIUMV de $\mathbf{E}[\mathbf{T}] = \mathbf{E}\left[\frac{X}{\mathbf{n}}\right] = \theta$. $\blacksquare$

El siguiente lema y ejercicio organizan la desigualdad CBS en una forma necesaria para el siguiente teorema

Lema 4.1. Sean X e Y variables aleatorias conjuntas con segundos momentos finitos. Entonces su covarianza existe y es igual

$$\mathbf{Cov}(X, Y) = \mathbf{E}\big[(X - \mathbf{E}(X))(Y - \mathbf{E}(Y))\big].$$

Si ambas varianzas son positivas, existe su correlación:

$$\rho(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sqrt{\mathbf{V}ar(X)\mathbf{V}ar(Y)}}$$

Teorema 4.2. Existe a lo más un solo EIUMV para una función estimables.

Prueba. Supongamos que $\mathbf{T}$ y $\mathbf{S}$ ambos EIUMV de $\varphi(\theta)$; entonces

$$\mathbf{E}_\theta[\mathbf{T}-\mathbf{S}] = 0 \quad \text{y} \quad \mathbf{V}ar_\theta(\mathbf{T}) = \mathbf{V}ar_\theta(\mathbf{S}).$$

Ahora

$$\mathbf{Cov}_\theta[\mathbf{T},\mathbf{S}] = \mathbf{E}_\theta[(\mathbf{T}-\varphi(\theta))(\mathbf{S}-\varphi(\theta))] = \mathbf{E}_\theta[\mathbf{TS}] - \varphi^2(\theta).$$

Por el Teorema 4.1, se tiene

$$\mathbf{E}_\theta[\mathbf{T}(\mathbf{T}-\mathbf{S})] = 0 \quad \text{o} \quad \mathbf{E}_\theta[\mathbf{T}^2] = \mathbf{E}[\mathbf{TS}].$$

Resulta que

$$\mathbf{Cov}(\mathbf{T},\mathbf{S}) = \mathbf{E}_\theta[\mathbf{T}^2] - \varphi^2(\theta) = \mathbf{V}ar_\theta(\mathbf{T})$$

y por lo tanto $\rho_\theta(\mathbf{TS}) = 1$ para todo $\theta \in \Theta$ por ejercicio 2.
$\mathbf{T} = \alpha\mathbf{S} + \beta$ casi con probabilidad 1 para
$\alpha = \mathbf{Cov}(\mathbf{T}.\mathbf{S})/\mathbf{V}ar_\theta(\mathbf{T}) = 1$. Esto es, $\mathbf{T} = \mathbf{S}+\beta$ casi seguramente.
Pero como $\mathbf{E}_\theta(\mathbf{T}) = \mathbf{E}_\theta(\mathbf{S}) = \varphi(\theta)$, entonces β debe ser igual a cero y por lo tanto $\mathbf{P}_\theta(\mathbf{T}=\mathbf{S}) = 1$ para todo $\theta \in \Theta$

□

Ejemplo 4.2. Supongamos que X es una variable aleatoria binomial con parámetros $\mathbf{n}$ y θ, donde $0 < \theta < 1$ se desconoce. Deseamos encontrar un estimador insesgado de $\varphi(\theta) = \frac{1}{\theta}$; es decir, tenemos que encontrar un estadístico $T(X)$ tal que $E_\theta[T(X)] = \frac{1}{\theta}$ para todo $0 < \theta < 1$. Entonces, tenemos que encontrar $T(0)$, $T(1)$,..., $T(n)$ tal que

$$\sum_{x=0}^{\mathbf{n}} T(x)\binom{\mathbf{n}}{x}\theta^x(1-\theta)^{n-x} = \frac{1}{\theta} \tag{4.3}$$

Multiplicando ambos lados por θ, obtenemos

$$\sum_{x=0}^{\mathbf{n}} T(x)\binom{\mathbf{n}}{x}\theta^{x+1}(1-\theta)^{\mathbf{n}-x} = \sum_{k=1}^{n+1} a(k)\theta^k = 1$$

donde $a(1)$,...,$a(n+1)$ dependen de $T(0)$, $T(1)$,..., $T(n)$; es decir, $a(k) = T(k-1)\binom{\mathbf{n}}{k-1}(1-\theta)^{\mathbf{n}-(k-1)}$. De ello se deduce que toda elección de $a(1)$,...,$a(n+1)$, la igualdad

$$a(1)\theta + \cdots + a(n+1)\theta^{n+1} = 1$$

se cumpla para a lo más, $n+1$ valores de θ entre 0 y 1 y que no se cumple para todo θ. Por lo tanto no existe un estimador insesgado de $\frac{1}{\theta}$.
Otra forma de analizar: que no existe tal estimador; $T(X)$, puede verse, por ejemplo, por el hecho de que cuando $\theta \to 0$, el lado izquierdo de la igualdad (4.3) tiende a $T(0)$ y el lado derecho a ∞. Sin embargo, existen estimadores de $\frac{1}{\theta}$ que (para $\mathbf{n}$ no demasiado pequeños) están cerca de $\frac{1}{\theta}$ con alta probabilidad. Por ejemplo, dado que $\frac{X}{\mathbf{n}}$ tiende a estar cerca de θ, $\frac{n}{X}$ (con algún ajuste cuando $X = 0$) tenderá a estar cerca de $\frac{1}{\theta}$. ■

Ejemplo 4.3. Suponga que Xes una variable aleatoria Poisson con media λ. Se desea encontrar un estimador insesgado de $e^{-\lambda}$ basado en X. Si $T(X)$ es este estimador, se debe tener

$$E_\lambda[T(X)] = \sum_{x=0}^{\infty} T(x)\frac{e^{-\lambda}\lambda^x}{x!} = e^{-2\lambda}$$

para todo $\lambda > 0$. Multiplicando ambos miembros de la ecuación por e^{λ}, resulta

$$\sum_{x=0}^{\infty} T(x)\frac{\lambda^x}{x!} = e^{-\lambda}$$
$$= \sum_{x=0}^{\infty} (-1)^x \frac{\lambda^x}{x!},$$

la cual implica que $T(x) = (-1)^x$. Por lo tanto el único estimador insesgado de $e^{-2\lambda}$ es $(-1)^X$; puesto que $0 < e^{-2\lambda} <$ para $\lambda > 0$, esto es claramente un estimador ridículo. Sin embargo, cabe señalar, que si $X_1, X_2, \ldots, X_n$ son variables aleatorias iid. Poisson con parámetro λ entonces $(1-2/n)^T$ (con $T = \sum_{i=1}^n X_i$) es insesgado; este estimador es algo más sensato sobre todo para valores grandes de n. ■

Los ejemplos 4.2 y 4.3, a pesar de que en muchos casos la clase de estimadores insesgados es no trivial.

Ejemplo 4.4. Suponga que X_1,...,X_n son variables aleatorias iid. con función de densidad

$$f(x;\mu) = e^{-(x-\mu)} \quad \text{para} \quad x \geq \mu$$

(donde $(-\infty < \mu < \infty)$. Dos posibles estimadores insesgados de μ son

$$\widehat{\mu}_1 = X_{(1)} - \frac{1}{n} \quad \text{y} \quad \widehat{\mu}_2 = \overline{X} - 1;$$

tenga en cuenta también que $t\widehat{\mu}_1 + (1-t)\widehat{\mu}_2$ es también estimador insesgado para cualquier t. realizando cálculos simples resulta que

$$Var(\widehat{\mu}_1) = \frac{1}{n^2} \quad \text{y} \quad Var(\widehat{\mu}_2) = \frac{1}{n}$$

de modo que $\widehat{\mu}_1$ tiene la menor varianza.Es interesante tener en cuenta que $\widehat{\mu}_1$ depende solamente de un estadístico suficiente de una sola dimensión $X_{(1)}$. ■

¿El estimador $\widehat{\mu}_1$ tiene la varianza mínima entre todos los estimadores insesgados de μ en los ejemplos anteriores?
El siguiente Teorema indica que en la investigación de EIUMV debe basarse en una función del estadístico suficiente.

Teorema 4.3. (**Teorema de Rao-Blackwell**)
Sea $\mathscr{P} = \{f_\theta(x), x \in \mathcal{X} \subseteq \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}\}$ y $\mathbf{T}(X_1, X_2, \ldots, X_\mathbf{n})$ sea un estadístico suficiente para $\varphi(\theta)$ donde $\varphi : \Theta \to \mathbb{R}$. Además asuma que $\mathbf{S}(X_1, X_2, \ldots, X_\mathbf{n})$ es un estimador in sesgado para $\varphi(\theta)$ con $\mathbf{E}_\theta(\mathbf{S}^2) < \infty$ para todo $\theta \in \Theta$ y que $\mathbf{S}$ no es una función de $\mathbf{T}$ solo. Considere la función definida por $U(\mathbf{t}) = \mathbf{E}_\theta[\mathbf{S}|\mathbf{T} = \mathbf{t}]$ (ambos θ y $\mathbf{T}$ pueden ser vectores.) Entonces

(a) $U(\mathbf{T}) = \mathbf{E}[\mathbf{S}|\mathbf{T}]$ es un estadístico y una función $\mathbf{T}$;
(b) $U(T)$ es un estimador insesgado de $\varphi(\theta)$;
(c) $\mathbf{V}ar_\theta[U(\mathbf{T})] \leq \mathbf{V}ar_\theta(\mathbf{S})$

Prueba . (a) Como T es un estadístico suficiente $U(T)$ no depende de θ y por lo tanto $U(\mathbf{T})$ es un estadístico y una función de T.

(b) Por la propiedad fundamental de la esperanza condicional $\mathbf{E}_\theta[\mathbf{E}(\mathbf{S}|\mathbf{T})] = \mathbf{E}_\theta[\mathbf{S}]$,

$$\mathbf{E}_\theta[U(\mathbf{T})] = \mathbf{E}_\theta[\mathbf{E}(\mathbf{S}|\mathbf{T})] = \mathbf{E}_\theta[\mathbf{S}] = \varphi(\theta), \forall\, \theta \in \Theta$$

(c) Se sabe que $\mathbf{V}ar(Y) = \mathbf{V}ar[\mathbf{E}(Y|X)] + \mathbf{E}[\mathbf{V}ar(Y|X)]$; luego,

$$\mathbf{V}ar[U(T)] = \mathbf{V}ar[\mathbf{E}(\mathbf{S}|\mathbf{T})] \leq \mathbf{V}ar_\theta(\mathbf{S})$$

□

La interpretación del Teorema es el siguiente: Si por aluna razón el investigador está interesado en hallar un estadístico con varianza más pequeña posible dentro de la clase de estimadores insesgados de θ, entonces puede restringir su investigación a la subclase de los estadístico insesgados que dependen solo de $\mathbf{T}$ (con probabilidad uno). Este es debido, a que si un estadístico U ya no es una función de $\mathbf{T}$ solo (con probabilidad uno), entonces se convierte así condicionándolo con respecto a $\mathbf{T}$. La varianza del estadístico resultante será más pequeña que la varianza del estadístico con que se comenzó el ítem (c) del Teorema.

Ejemplo 4.5. Considere el modelo exponencial

$$f(x;\theta) = \frac{1}{\theta}e^{-\frac{1}{\theta}x}\mathbf{I}(0 < x < \infty) \quad \text{con} \quad \varphi(\theta) = \theta > 0.$$

Para una muestra aleatoria de tamaño $\mathbf{n}$, sea $\mathbf{T} = \displaystyle\sum_{i=1}^{\mathbf{n}} X_i$. primero necesitamos saber si $\mathbf{T}$ es un estadístico suficiente, para ello utilizamos el Teorema de factorización, así

$$\begin{aligned} f(x_1, x_2, \ldots, x_{\mathbf{n}};\theta) &= \prod_{I=1}^{\mathbf{n}} \frac{1}{\theta}e^{-\frac{1}{\theta}x_i}\mathbf{I}(0 < x_i < \infty) \\ &= \underbrace{\theta^{-\mathbf{n}}e^{-\frac{1}{\theta}\sum_{i=1}^{\mathbf{n}} x_i}}_{g(\mathbf{t};\theta)} \underbrace{\prod_{i=1}^{\mathbf{n}} \mathbf{I}(0 < x_i < \infty)}_{h(x_1,x_2,\ldots,x_{\mathbf{n}})}, \end{aligned}$$

luego $\mathbf{T} = \displaystyle\sum_{i=1}^{\mathbf{n}} X_i$. es un estadístico suficiente para θ. Ahora buscamos un estimador insesgado para θ. Sea $\mathbf{S}(X_1, X_2, \ldots, X_{\mathbf{n}}) = X_1$ ¿es $\mathbf{S}$ insesgado?, veamos,

$$\mathbf{E}(\mathbf{S}) = \mathbf{E}(X_1) = \int_0^{\infty} x\frac{1}{\theta}e^{-\frac{1}{\theta}x}dx = \theta$$

Luego $\mathbf{S}$ es estimador insesgado para *theta*. Ahora por el Teorema de Rao-Blackwell

$$U(\mathbf{T}) = \mathbf{E}[\mathbf{S}|\mathbf{T}] = \mathbf{E}[X_1|\mathbf{T}]$$

es un estimador insesgado de θ con varianza más pequeña que $\mathbf{X}_1$.
$\blacksquare$

Ejercicio 3.
Por un cálculo directo verificar la desigualdad de varianza , calcule $\mathbf{V}ar_\theta(X_1)$, $\mathbf{V}ar_\theta(U(\mathbf{T}))$ en el ejemplo anterior. Indicación:

$$\mathbf{E}(\mathbf{T}|\mathbf{T}) = \mathbf{T} = \sum_{i=1}^{\mathbf{n}} \mathbf{E}[X_i|\mathbf{T}] \quad \text{implica} \quad \mathbf{E}[X_1|T] = \frac{\mathbf{T}}{\mathbf{n}}$$

Ejercicio 4.
Sean $X_1, X_2, \ldots, X_\mathbf{n}$ una muestra aleatoria de la distribución Bernoulli con parámetro θ. Pruebe que

$$\mathbf{E}\left[X_1 \middle| \sum_{i=1}^{\mathbf{n}} X_i\right]$$

es un estimador insesgado de θ con menor varianza que la varianza de X_1.

Según la hipótesis del Teorema de Rao-Blackwell $U(\mathbf{T}) = \mathbf{E}[\mathbf{S}|\mathbf{T}]$ es un estimador insesgado con varianza no mayor que de $\mathbf{S}$; Sin embargo el Teorema no dice que $U(\mathbf{T})$ es un EIUMV para $\varphi(\theta)$ Pero, el siguiente Teorema sí dice que, si en adición, el estadístico suficiente $\mathbf{T}$ es completo, entonces $\mathbf{E}[\mathbf{S}|\mathbf{T}]$ es de hecho un EIUMV.

Teorema 4.4. (**Lehmann-Scheffé**) Sean $X_1, X_2, \ldots, X_\mathbf{n}$ una muestra aleatoria de $\mathscr{P} = \{f_\theta(x), x \in \mathcal{X} \subseteq \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}\}$, y sea $\varphi(\theta)$

una función estimable . Si $\mathbf{T} = \mathbf{T}(X_1, X_2, \ldots, X_\mathbf{n})$ es un estadístico suficiente y completo y para todo $\theta \in \Theta$, $U = U(\mathbf{T})$ es un estimador insesgado de $\varphi(\theta)$ con varianza finita, entonces U es el único EIUMV de $\varphi(\theta)$.

Prueba. Por el Teorema de Rao-Blackwell, la investigación para un EIUMV de $\varphi(\theta)$ puede restringirse a la clase de estimadores insesgados de $\varphi(\theta)$ que son funciones del estadístico suficiente $\mathbf{T}$. pero la completitud implica que la clase de estimadores que tienen solo un elemento en el sentido *casi seguramente* (c.s.): si $V(\mathbf{T})$ es otro estimador insesgado de $\varphi(\theta)$, entonces para todo $\theta \in \Theta$

$$\mathbf{E}_\theta[U(\mathbf{T}) - V(\mathbf{T})] = 0 \quad \text{implica que } U(\mathbf{T}) - V(\mathbf{T}) = 0 \text{ c.s..}$$

Tenga en cuenta que el Teorema de Lehmann-Scheffé proporciona *condiciones suficiente* para un EIUMV pero no *condiciones necesarias*; esto es el Teorema proporciona un método de construcción de un EIUMV cuando un estadístico suficiente y completo exista. Como un estadístico suficiente y completo es también un estadístico minimal; después de obtener $\mathbf{E}[\mathbf{S}] = \varphi(\theta)$, la investigación puede procederse observando para estadísticos suficientes que son minimales ; luego, puede emplearse el condicionamiento. Pero puede suceder que incluso si no existe un estadístico completo y suficiente, existe un EIUMV, la búsqueda de un EIUMV en tales condiciones se discutirá en la próxima lección. $\square$

Ejemplo 4.6. Para $\theta \in \Theta = \{2, 3, \ldots\}$ y $A = \{1, 2, \ldots, \theta\}$, sea

$$f(x; \theta) = \frac{1}{\theta}\mathbf{I}_A(x)$$

Sea $g : A \to \mathbb{R}$ definida por

$$g(x) = \begin{cases} \frac{1}{2} & \text{si} \quad x = 1 \\ -\frac{1}{2} & \text{si} \quad x = 2 \\ 0 & \text{si} \quad x \geq 3 \end{cases}$$

Se observa que $\mathbf{E}_\theta[g(X)] = 0$ para todo $\theta \in \Theta$, pero $g \neq 0$ c.s., es decir el modelo no es completo ■

Ejercicio 5.
Continúe con el ejemplo anterior. Pruebe que

$$(a) \quad \mathbf{E}_\theta[g(X)] = 0 \qquad (b) \quad \mathbf{E}_\theta[X] = \frac{\theta + 1}{2}$$

Ejemplo 4.7. Tomando el enunciado del ejemplo 4.6, considere el estadístico dado por

$$\mathbf{T}(X) = \begin{cases} 2X - 1 & \text{para} \quad X \geq 3 \\ 2 & \text{para} \quad X = 1, 2 \end{cases}$$

Como

$$\begin{aligned}
\mathbf{E}_\theta[\mathbf{T}(X)] &= \sum_{x=1}^{\theta} T(x)f(x;\theta) = \frac{1}{\theta}\sum_{x=1}^{\theta} T(x)\mathbf{I}_A(x) \\
&= \frac{1}{\theta}[T(1)\mathbf{I}_A(1) + T(2)\mathbf{I}_A(2) + \sum_{x=3}^{\theta}(2x-1)\mathbf{I}_A(x)] \\
&= \frac{1}{\theta}[2 + 2 + \sum_{x=3}^{\theta}(2x-1)\mathbf{I}_A(x)] \\
&= \frac{1}{\theta}[1 + 3 + \sum_{x=3}^{\theta}(2x-1)\mathbf{I}_A(x)] \\
&= \frac{1}{\theta}[\sum_{x=1}^{\theta}(2x-1)\mathbf{I}_A(x)] \\
&= \mathbf{E}[2X-1] = \theta
\end{aligned}$$

$\mathbf{T}(X)$ es un estimador insesgado de θ. Ahora, sea $\mathbf{S}(X)$ un estimador tal $\mathbf{E}[\mathbf{S}(X)] = 0$ para todo $\theta \in \Theta$; esto es

$$\mathbf{E}[\mathbf{S}(X)] = \sum_{x=1}^{\theta}\mathbf{S}(x)f(x;\theta) = \frac{1}{\theta}\sum_{x=1}^{\theta}\mathbf{S}(x) = 0, \text{para } \theta \in \Theta = \{2, 3, \ldots, \theta\}$$

veamos

$$\begin{aligned}
&\text{paras } \theta = 2\text{:} \quad \frac{1}{2}[\mathbf{S}(1) + \mathbf{S}(2)] = 0 \Rightarrow \mathbf{S}(1) = -\mathbf{S}(2) \\
&\text{paras } \theta = 3\text{:} \quad \frac{1}{2}[\mathbf{S}(1) + \mathbf{S}(2) + \mathbf{S}(3)] = 0 \Rightarrow \mathbf{S}(3) = 0
\end{aligned}$$

En efecto, resukta que para todo $\theta \geq 3$, $\mathbf{S}(x) = 0$, para todo $x \geq 3$. Por lo tanto, para algún $\alpha > 0$ se tiene

$$\mathbf{S}(x) = \begin{cases} \alpha & \text{si} \quad x = 1 \\ -\alpha & \text{si} \quad x = 2 \\ 0 & \text{si} \quad x \geq 3 \end{cases}$$

Por rl Teorema 4.1 se tiene

$$\begin{aligned}\mathbf{E}_\theta[T\mathbf{S}] &= \frac{1}{\theta}\sum_{x=1}^{\theta}\mathbf{T}(x)\mathbf{S}(x)\\ &= \mathbf{T}(1)\mathbf{S}(1) + \mathbf{T}(2)\mathbf{S}(2) = 0 \text{ para todo } \theta \in \Theta.\end{aligned}$$

lo que implica que $\mathbf{T}$ es un EIUMV de θ. ■

Ejemplo 4.8. Sea $X_1, X_2, \ldots, X_\mathbf{n}$ una muestra aleatoria de una distribución Bernoulli con parámetro $\theta = \mathbf{P}(X_i = 1)$. Tenga en cuenta que $\mathbf{T} = \sum_{i=1}^{\mathbf{n}} X_i$ es un estadístico suficiente y completo para θ. Se tiene:

(a) Como $\overline{\mathbf{X}} = \frac{1}{\mathbf{n}}\sum_{i=1}^{\mathbf{n}} X_i$ es un estimador insesgado de θ y es una función del estadístico suficiente y completo $\mathbf{T}$, $\overline{\mathbf{X}}$ es el EIUMV de θ.

(b) sea $\varphi(\theta) = \theta(1-\theta)$ y

$$\mathbf{S}^2(X_1, X_2, \ldots, X_\mathbf{n}) = \frac{1}{\mathbf{n}-1}\sum_{i=1}^{\mathbf{n}}(X_i - \overline{\mathbf{X}})^2$$

como $\mathbf{V}ar_\theta(\mathbf{X}) = \theta(1-\theta)$, se observa que S^2 es un estimador insesgado de $\varphi(\theta)$. por el Teorema Lehmann-Scheffé , el EIUMV de $\varphi(\theta)$ es $U(\mathbf{T}) = \mathbf{E}_\theta[\mathbf{S}|T]$.

(c) Calculemos esta esperanza condicional. Como $X_i \in \{0, 1\}$ se tiene que $X_i^2 = X_i$, $i = 1, 2, \ldots, \mathbf{n}$ y

$$\sum_{i=1}^{\mathbf{n}}(X_i - \overline{\mathbf{X}})^2 = \sum_{i=1}^{\mathbf{n}} X_i^2 - \frac{\left(\sum_{i=1}^{\mathbf{n}} X_i\right)^2}{\mathbf{n}}$$

$$\begin{aligned}
\mathbf{E}_\theta[\mathbf{S}^2|\mathbf{T}] &= \mathbf{E}_\theta\left[\frac{1}{\mathbf{n}-1}\sum_{i=1}^{\mathbf{n}}(X_i-\overline{\mathbf{X}})^2\,\middle|\,\sum_{i=1}^{\mathbf{n}}X_i\right]\\
&= \mathbf{E}_\theta\left[\frac{1}{\mathbf{n}-1}\left\{\sum_{i=1}^{\mathbf{n}}X_i^2-\frac{\left(\sum_{i=1}^{\mathbf{n}}X_i\right)^2}{\mathbf{n}}\right\}\,\middle|\,\sum_{i=1}^{\mathbf{n}}X_i\right]\\
&= \mathbf{E}_\theta\left[\frac{1}{\mathbf{n}-1}\left\{\sum_{i=1}^{\mathbf{n}}X_i-\frac{\left(\sum_{i=1}^{\mathbf{n}}X_i\right)^2}{\mathbf{n}}\right\}\,\middle|\,\sum_{i=1}^{\mathbf{n}}X_i\right]\\
&= \frac{1}{\mathbf{n}-1}\left\{\sum_{i=1}^{\mathbf{n}}X_i-\frac{\left(\sum_{i=1}^{\mathbf{n}}X_i\right)^2}{\mathbf{n}}\right\}\\
&= \frac{1}{\mathbf{n}-1}\left\{\mathbf{T}-\frac{(\mathbf{T})^2}{\mathbf{n}}\right\}
\end{aligned}$$

■

Ejercicio 6.

Sea $X_1, X_2, \ldots, X_{\mathbf{n}}$ una muestra aleatoria de una distribución normal con media $\theta \in \mathbb{R}$ varianza conocida σ^2.

(a) Demuestre que $\mathbf{E}[X_1|\overline{\mathbf{X}}]$ es un EIUMV de θ.

(b) Demuestre que $\mathbf{E}[X_1|\overline{\mathbf{X}}] = \overline{\mathbf{X}}$ c.s.

Ejercicio 7.

Considere el modelo Poisson con parámetro $\theta > 0$ y $\varphi(\theta) = \theta e^{-\theta}$. Para una muestra aleatoria simple $X_1, X_2, \ldots, X_{\mathbf{n}}$. Demuestre:

(a) que

$$\mathbf{T} = \begin{cases} 1 & \text{si} \quad X_1 = 1 \\ 0 & \text{si} \quad X_1 \neq 1 \end{cases}$$

es un estimador insesgado para θ;

(b) que $\sum_{j=1}^{\mathbf{n}} X_j$ es un estadístico suficiente y completo para θ.

(c) Derive el EIUMV de $\varphi(\theta)$.

5. Estimadores eficientes

Cuando no se dispone de estadísticos suficientes y completos, para encontrar "mejores estimadores" uno debe investigar otras técnicas; la siguiente definición impone algunas condiciones adicionales. Aquí X es una variable aleatoria de tipo continuo (v.a.); si X es una variable aleatoria de tipo discreto, la integración es reemplazada por la sumatoria.

Definición 5.1. Sea X una variable aleatoria general con rango $\mathcal{X}$ y densidad$f(x;\theta)$ para $\theta \in \Theta$; sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de esta población de modo que el espacio muestral es

$$\mathscr{X} = \mathcal{X}^n = \underbrace{\mathcal{X} \times \mathcal{X} \times \cdots \times \mathcal{X}}_{n \text{ veces}}.$$

Definición 5.2. Un modelo $\mathscr{P} = \{f_\theta(x) : x \in \mathbb{R}, \theta \in \Theta \subset \mathbb{R}\}$ se llama *modelo regular* si cumple las siguientes condiciones:

(R_1) El soporte $f(\cdot;\theta)$ es un conjunto $\mathscr{X}$ independiente de θ. ($\mathscr{X}$ es el rango de la variable aleatoria X).

(R_2) Θ es un intervalo abierto (finito o infinito) de $\mathbb{R}$.

(R_3) $f(x;\cdot)$ es diferenciable en Θ para $x \in \mathscr{X}$ excepto posiblemente en el conjunto $\mathcal{A}$ con $P_\theta(\mathcal{A}) = 0$ para todo $\theta \in \Theta$.

(R_4) La información de Fisher, $I(\theta) = \mathbf{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\ln f(\cdot;\theta)\right\}^2\right]$ cumple $0 < I(\theta) < \infty$.

(R_5) $0 = \frac{\partial}{\partial\theta}(1) = \frac{\partial}{\partial\theta}\int_{\mathscr{X}} f(x;\theta)dx = \int_{\mathscr{X}}\left[\frac{\partial}{\partial\theta}f(x;\theta)\right]dx$

Los siguientes ítems son consecuencias simples de la definición 5.2.

Primero, consideremos una muestra aleatoria simple, $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ de un modelo probabilístico $\mathscr{P}$, sea la función de verosimilitud

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

(a) Como $\frac{\partial}{\partial\theta}\ln L(\theta|\mathbf{x}) = \displaystyle\sum_{i=1}^{\mathbf{n}} \frac{\partial}{\partial\theta}\ln f(x_i|\theta)$,

$$I_{\mathbf{n}}(\theta) = \mathbf{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\ln L(\theta|\mathbf{X})\right\}^2\right] < \infty \text{ por ítem } (R_4)$$

$I_n(\theta)$ se llama la información de Fisher en la muestra $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.

(b) La variables aleatorias

$$\left\{\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right\}$$

son iid porque son funciones de variables aleatorias disjuntas iid $\{X_i\}$ y tiene media cero por (R_5); esto es,

$$\begin{aligned}
E_\theta\Big[\Big\{\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\Big\}\Big] &= \int_{\mathscr{X}} \frac{\partial}{\partial\theta}\ln f(x_i;\theta)\cdot f(x_i,\theta)dx \\
&= \int_{\mathscr{X}} \frac{\frac{\partial}{\partial\theta}f(x_i;\theta)}{\cancel{f(x_i;\theta)}}\cancel{f(x_i;\theta)}dx \\
&= \int_{\mathscr{X}} \frac{\partial}{\partial\theta}f(x_i;\theta)dx \\
&= \frac{\partial}{\partial\theta}\int_{\mathscr{X}} f(x;\theta)dx \\
&= \frac{\partial}{\partial\theta}(1) \\
&= 0\ \forall\theta\in\Theta.
\end{aligned}$$

(c) Resulta que la varianza de cada variable aleatoria $\frac{\partial}{\partial\theta}\ln f(x_i;\theta)$ es $I(\theta)$; esto es,

$$\begin{aligned}
Var_\theta\Big(\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\Big) &= E_\theta\Big[\Big\{\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\Big\}^2\Big] - \Big\{\cancelto{0}{E\Big[\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\Big]}\Big\}^2 \\
&= E_\theta\Big[\Big\{\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\Big\}^2\Big] \\
&= I(\theta)
\end{aligned}$$

en consecuencia $I_n(\theta) = nI(\theta)$ es la varianza de la muestra por tanto

$$\frac{\partial}{\partial\theta}\ln L(\theta|\mathbf{X}) \sim (0, n\ I(\theta))$$

Ejercicio 8.

Asuma en adición, $f(x;\cdot)$ es doblemente diferenciable con respecto a θ y que

$$\frac{\partial^2}{\partial\theta^2}\int_{\mathscr{X}} f(x|\theta) = \int_{\mathscr{X}} \frac{\partial^2}{\partial\theta^2}f(x;\theta)dx$$

Demuestre que

$$I(\theta) = -E_\theta\Big[\frac{\partial^2}{\partial\theta^2}\ln f(x|\theta)\Big]$$

El siguiente teorema contiene la *desigualdad de información* o *límite inferior* desarrollado en diferentes formas por Cramer, Rao, Dini, Fréchet. Usamos la notación

$$\mathscr{X} = \underbrace{\mathcal{X} \times \mathcal{X} \times \cdots \times \mathcal{X}}_{(n \text{ veces})}$$

es el producto cartesiano de un conjunto $\mathcal{X}$ consigo mismo n veces.

Teorema 5.1. Sea $X_1, \ldots, X_n$ una muestra aleatoria simple de un modelo regular $\mathscr{P}$, de modo que $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathscr{X}$. Sea $\mathcal{T}(\theta)$ una función estimable de θ; sea $T(\mathbf{X}) = T(X_1, X_2, \ldots, X_n)$ un estimador insesgado de $\mathcal{T}(\theta)$ $\forall\theta \in \Theta$ con $\mathbf{V}ar_\theta(T) < \infty \forall\theta \in \Theta$. Si

$$\mathcal{T}'(\theta) = \frac{\partial}{\partial\theta} E(T(X_1, X_2, \ldots, X_n)) = \int_{\mathbf{X}} T(\mathbf{x}) \frac{\partial}{\partial\theta} f(\mathbf{x}|\theta) d\mathbf{x},$$

entonces

$$\mathbf{V}ar_\theta(T(\mathbf{X})) \geq \frac{[\mathcal{T}'(\theta)]^2}{n\ I(\theta)}.$$

Demostración. Por la desigualdad de Cauchy-Schwarz, resulta

$$\Big[E_\theta\Big[\{T(\mathbf{X}) - \mathcal{T}(\theta)\}\Big\{\frac{\partial}{\partial\theta} \ln L(\theta|\mathbf{X})\Big\}\Big]\Big]^2 \leq Var_\theta(T(\mathbf{X})) \times I_n(\theta) \tag{5.1}$$

Como $E_\theta[\frac{\partial}{\partial\theta}\ln L(\theta|\mathbf{X})] = 0$,

$$\begin{aligned}
E_\theta\Big[\{T(\mathbf{X}) - \mathcal{T}(\theta)\}\Big\{\frac{\partial}{\partial\theta}\ln L(\theta|\mathbf{X})\Big\}\Big] &= E_\theta\Big[T(\mathbf{X})\Big\{\frac{\partial}{\partial\theta}\ln L(\theta|\mathbf{X})\Big\}\Big] \\
&= \int_{\mathbf{X}} T(\mathbf{X})\Big\{\frac{\partial}{\partial\theta}\ln L(\theta|\mathbf{X})\Big\}\cdot f(\mathbf{X}|\theta)d\mathbf{X} \\
&= \int_{\mathbf{X}} T(\mathbf{X})\frac{\frac{\partial}{\partial\theta}f(\mathbf{X}|\theta)}{\cancel{f(\mathbf{X}|\theta)}}\cdot \cancel{f(\mathbf{X}|\theta)}d\mathbf{X} \\
&= \int_{\mathbf{X}} T(\mathbf{X})\frac{\partial}{\partial\theta}f(\mathbf{X}|\theta)d\mathbf{X} \\
&= \frac{\partial}{\partial\theta}\int_{\mathbf{X}} T(\mathbf{X})f(\mathbf{X}|\theta)d\mathbf{X} \\
&= \frac{\partial}{\partial\theta}E_\theta(T(\mathbf{X})) = \frac{\partial}{\partial\theta}\mathcal{T}(\theta) = \mathcal{T}'(\theta)
\end{aligned}$$

De esta manera el lado izquierdo de la ecuación (5.1) es $(\mathcal{T}'(\theta))^2$ y la conclusión es inmediata. $\square$

Cuando la igualdad en realidad se cumple, la varianza del estimador insesgado $\mathbf{T}$ es tan pequeña como la varianza de cualquier estimador insesgado de modo que entonces $\mathbf{T}$ puede ser un EIUMV.

Ejercicio 9.

Sea X una variable aleatoria discreta con soporte $\mathcal{X} = a_1, a_2, \ldots, a_m$ y función de probabilidad

$$P_i(\theta) = P_\theta(X = a_i),\ i = 1, 2, \ldots m;\ \theta \in [\alpha, \beta].$$

Suponga que todos los P_i sean diferenciables en $[\alpha, \beta]$. Sean $\lambda_1, \ldots, \lambda_m$ números reales y $H(\theta) = \sum_{i=1}^m \lambda_i P_i(\theta)$. Demuestre que el límite inferior de $\sum_{i=1}^m \Big(\lambda_i - H(\theta)\Big)^2 P_i(\theta)$ es

$$[H'(\theta)]^2 + \left\{\sum_{i=1}^m P_i(\theta)\left[\frac{\partial}{\partial\theta}\log P_i(\theta)\right]^2\right\}$$

Ejercicio 10.
Considere el modelo Bernoulli con densidad

$$f(x;\theta) = \theta^x(1-\theta)^{1-x}; \ x \in \{0,1\} \text{ y } \theta \in (0,1).$$

Encuentre $\mathbf{E}_\theta(X)$, $\mathbf{E}_\theta(X^2)$, $\mathbf{E}_\theta[(1-X)^2]$.

Ejemplo Continúe con el modelo Bernoulli; sea $\mathcal{T}(\theta) = \theta$ y una muestra aleatoria $X_1, X_2, \ldots, X_n$. Tenga en cuenta que $\Theta = (0,1)$, el espacio paramétrico, es un conjunto abierto (realmente un intervalo) de $\mathbb{R}$. Obviamente, $\frac{\partial}{\partial\theta} f(x;\theta)$ existe para cada x y la derivada dentro de cualquier signo de sumatoria finita es válido. Sea

$$\mathbf{T} = \mathbf{T}(\mathbf{X}) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

(a) $\mathbf{T}$ es un estimador insesgado de $\mathcal{T}(\theta) = \theta$ tal que

$$E_\theta(\mathbf{T}) = E_\theta(\overline{X}) = \theta \ \forall \ \theta \in (0,1).$$

(b) Como

$$\frac{\partial}{\partial\theta} \ln f(X;\theta) = \frac{X}{\theta} - \frac{1-X}{1-\theta}$$

y

$$\frac{\partial^2}{\partial\theta^2} \ln f(X;\theta) = -\frac{X}{\theta^2} - \frac{1-X}{(1-\theta)^2},$$

entonces

$$\begin{aligned} E_\theta\Big[\frac{\partial^2}{\partial\theta^2} \ln f(x;\theta)\Big] &= -\frac{E(X)}{\theta^2} - \frac{1-E_\theta(X)}{(1-\theta)^2} \\ &= -\frac{1}{\theta} - \frac{1-\theta}{(1-\theta)^2} \\ &= -\frac{1}{\theta(1-\theta)} \end{aligned}$$

$$I(\theta) = -E_\theta\Big[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\Big] = -\Big(-\frac{1}{\theta(1-\theta)}\Big) = \frac{1}{\theta(1-\theta)}$$
$$I(\theta) = \frac{1}{\theta(1-\theta)}$$

Resulta que "información" límite inferior de la varianza es $\frac{\theta(1-\theta)}{n}$.

(c) La varianza de **T** es $\frac{\theta(1-\theta)}{n}$ que es iguala al límite inferior de Rao-Cramer y de modo que **T** es un EIUMV

Ejemplo 5.1. Sea $\mathscr{P}$ un modelo normal con media $\theta \in \mathbb{R}$, varianza σ^2 conocida. Sea $\varphi(\theta) = \theta$ y $\mathbf{T} = \overline{X}$. Como

$$\frac{\partial}{\partial\theta}\ln f(X;\theta) = \frac{X-\theta}{\sigma^2},\ I(\theta) = E_\theta\Big[\frac{1}{\sigma^4}(X-\theta)^2\Big] = \frac{1}{\sigma^2}$$

El límite inferior de "información" Rao-Cramer es $\frac{\sigma^2}{n}$ que es precisamente igual a $\mathbf{V}ar_\theta(\overline{X})$; en consecuencia $\overline{X}$ es un EIUMV de θ. ■

Definición 5.3. [Estimador eficiente] Un estimador insesgado **T** de $\mathcal{T}(\theta)$ tal que

$$\mathbf{V}ar_\theta(T) = \frac{\Big[\mathcal{T}'(\theta)\Big]^2}{n\ I(\theta)}\ \forall\ \theta \in \Theta$$

se llama un estimador eficiente.

La expresión $\frac{\Big[\mathcal{T}'(\theta)\Big]^2}{n\ \mathbf{I}(\theta)}$ se llama **cota Inferior de Cramér-Rao** o límite inferior de información.

Cuando las asunciones del teorema anterior se cumplen, un estimador eficiente necesariamente es un EIUMV; debe haberse notado que los estadísticos en los ejemplos anteriores, fueron también completos. Sin embargo, un EIUMV no necesariamente es un estadístico eficiente.

El siguiente ejercicio está preparado para un ejemplo de esta naturaleza.

Ejercicio 11.
Considere el modelo de Poisson con función de densidad

$$f(x;\theta)=\begin{cases} e^{-\theta}\frac{\theta^x}{x!} & \text{si } x=0,1,\ldots \\ 0 & \text{c.c.} \end{cases}$$

Demuestre que:

(a) $\mathbf{S}=\mathbf{S}(\mathbf{X})=\sum_{j=1}^{n} X_j \sim \text{Poisson}(n\theta)$
(b) $\mathbf{S}$ es un estadístico suficiente y completo para θ
(c) $E_\theta(\mathbf{t}^{X_i})=e^{-\theta(1-\mathbf{t})}\ \forall \mathbf{t}\in\mathbb{R}$
(d) $I_n(\theta)=\dfrac{n}{\theta}$

Ejemplo 5.2. Continúe con el ejercicio y tome $\mathcal{T}(\theta)=e^{-\theta}$. Entonces

$$\frac{\partial}{\partial\theta}\ln f(X|\theta)=\frac{X}{\theta}-1 \quad \text{y} \quad \frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)=-\frac{X}{\theta^2}$$

entonces

$$I(\theta)=-E_\theta\Big[\frac{\partial^2}{\partial\theta^2}\ln f(x|\theta)\Big]=-E\Big[-\frac{X}{\theta^2}\Big]=\frac{E(X)}{\theta^2}=\frac{\theta}{\theta^2}=\frac{1}{\theta}.$$

(a) Sea $\mathbf{T} = \mathbf{T}(\mathbf{X}) = \begin{cases} 1 & \text{si } X_1 = 0 \\ 0 & \text{si } X_1 \geq 1 \end{cases}$ Luego

$$\begin{aligned} E_\theta(t) &= 1P(X_1 = 0) + 0P(X_1 \geq 1) \\ &= e^{-\theta} = \mathcal{T}(\theta) \end{aligned}$$

Como $\mathbf{S}(\mathbf{X}) = \sum_{i=1}^{\mathbf{n}} X_i$ es un estadístico suficiente y completo, el EIUMV de $\mathcal{T}(\theta)$ es $U(\mathbf{S}) = E_\theta(\mathbf{T}|\mathbf{S}) = P_\theta\left(X_1 = 0|\mathbf{S}\right)$ (por el Teorema de Lehmann-Scheffé).

(b) Necesitamos solo esta probabilidad condicional, pero es igual de fácil encontrar la distribución condicional de de X_1 dado $\mathbf{S}$. La distribución conjunta de X_1 y $\mathbf{S}$ es

$$\begin{aligned} P_\theta(X_1 = k, \mathbf{S} = m) &= P_\theta(X_1 = k, \sum_{j=2}^{n} X_j = m - k) \\ &= P_\theta(X_1 = k)P_\theta\Big(\sum_{j=2}^{n} X_j = m - k\Big) \\ &= \frac{e^{-\theta}\theta^k}{k!} \cdot \frac{((n-1)\theta)^{m-k}}{(m-k)!} e^{-(n-1)\theta} \\ &= \frac{e^{-n\theta}}{k!(m-k)!}\theta^k (n-1)^{m-k}\theta^{m-k} \\ P_\theta(X_1 = k, \mathbf{S} = m) &= \frac{\theta^m e^{-n\theta}}{k!(m-k)!}(n-1)^{m-k} \end{aligned}$$

Como

$$P_\theta(\mathbf{S} = m - k) = (n\theta)^{m-k} \frac{e^{-n\theta}}{(m-k)!}$$

$$
\begin{aligned}
P_\theta(X_1 = k|S = m) = \frac{P_\theta(X_1 = k, S = m)}{P(S = m)} &= \frac{\frac{\theta^m e^{-n\theta}}{k!(m-k)!}(n-1)^{m-k}}{\frac{e^{-n\theta}(n\theta)^m}{m!}} \\
&= \frac{m!\theta^m (n-1)^{m-k}}{k!(m-k)!(n\theta)^m} \\
&= \frac{m!}{k!(m-k)!}\frac{(n-1)^{m-k}}{n^m} \\
&= \binom{m}{k}\frac{(n-1)^m}{n^m}(n-1)^{-k} \\
&= \binom{m}{k}(1-\frac{1}{n})^m n^{-k}(1-\frac{1}{n})^{-k} \\
&= \binom{m}{k}\frac{1}{n^k}(1-\frac{1}{n})^{m-k} \\
&= \binom{m}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{m-k}
\end{aligned}
$$

Lo que significa que la distribución condicional de X_1 dado $\mathbf{S} = m$ es una distribución binomial con parámetros m y $1/n$; es decir

$X_1|\mathbf{S} = m \sim \text{Bin}\big(m, \frac{1}{n}\big)$.

En particular resulta,

$$U = U(\mathbf{S}) = P_\theta(X_1 = 0|\mathbf{S}) = \left(1 - \frac{1}{n}\right)^{\mathbf{S}}$$

(c) Ahora podemos calcular la varianza de $\mathbf{U}$, utilizando el ítem (c) del ejercicio 4.

$$E_\theta(\mathbf{U}) = E_\theta\left[\left(1-\frac{1}{n}\right)^{\mathbf{S}}\right] = e^{-n\theta\left(1-\left\{1-\frac{1}{n}\right\}\right)} = e^{-\theta}$$

$$\mathbf{U}^2 = \left(1-\frac{1}{n}\right)^{2\mathbf{S}} \Longrightarrow E_\theta[\mathbf{U}^2] = E_\theta\left(1-\frac{1}{n}\right)^{2\mathbf{S}} = e^{-n\theta\left[1-\left(1-\frac{1}{\mathbf{n}}\right)^2\right]}$$

$$E_\theta[\mathbf{U}^2] = e^{-2\theta+\frac{\theta}{\mathbf{n}}}$$

$$\mathbf{V}ar_\theta(U) = E[\mathbf{U}^2] - [E(\mathbf{U})]^2 = e^{-2\theta}\left(e^{\frac{\theta}{n}} - 1\right)$$

(d) El límite inferior de información Rao-Cramer, resulta

$$\frac{(\mathcal{T}'(\theta))^2}{n\mathbf{I}(\theta)} = \frac{(-e^{-\theta})^2}{n/\theta} = \frac{\theta e^{-2\theta}}{n}$$

Ahora, un resultado matemático establece que: $\forall x > 0,\ e^x - 1 > x$, tenemos para $\theta > 0$ y $n > 0$ entonces

$$\left(e^{\frac{\theta}{n}} - 1\right) > \frac{\theta}{n}$$

multiplicando ambos miembros de la desigualdad por $e^{-2\theta}$ resulta

$$e^{-2\theta}\left(e^{\frac{\theta}{n}} - 1\right) > \frac{\theta}{n}e^{-2\theta}, \quad \forall\theta > 0$$

Así $\mathbf{U} = (1 - \frac{1}{n})^{\mathbf{S}}$ es el EIUMV de $\mathcal{T}(\theta) = e^{-\theta} > 0$.
Como la $\mathbf{V}ar_\theta(\mathbf{T})$ no es igual al límite inferior de Rao-Cramer, $\mathbf{U}$ no es un estimador eficiente para $\mathcal{T}(\theta) = e^{-\theta}$.

■

Ejercicio 12.
Sea $X \sim b(n, \theta)$. Demuestre que $\mathbf{T}(X) = \frac{X}{n}$ es un EIUMV de θ y también es eficiente.

Si un modelo no satisface las condiciones de regularidad (R_1) a (R_5), entonces un EIUMV puede tener una varianza que realmente es menor que el límite inferior de información (cuando éste existe).

Ejercicio 13.
Considere el modelo uniforme

$$\mathscr{P} = \{f_\theta : \theta \in \Theta > 0\}$$

donde

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} & \text{si } 0 < x < \theta,\ \theta > 0,\ x \in \mathbb{R} \\ 0 & \text{si c.c.} \end{cases}$$

(a) ¿La familia satisface las condiciones de regularidad?
(b) Calcule $\mathbf{I}(\theta)$
(c) Sea $\mathbf{X}$ una muestra aleatoria extraída de $\mathcal{P}$ con estadístico de orden máximo $\mathbf{T} = \text{máx}\{X_1, \ldots, X_n\}$. Demuestre que $\mathbf{U} = \frac{n+1}{n}\mathbf{T}$ es un estimador insesgado de $\mathcal{T}(\theta) = \theta$.
(d) Calcule $\mathbf{V}ar_\theta(\mathbf{U})$
(e) Use el teorema de Lehmann-Scheffé para demostrar que $\mathbf{U}$ es el EIUMV de θ
(f) Compare $\mathbf{V}ar_\theta(\mathbf{U})$ con el límite inferior de "información". Explique.

Concluimos esta parte sobre la estimación insesgada del parámetro real uni-dimensional al mostrar una relación entre la estimación eficiente y las familias exponenciales.

Según las condiciones de regularidad, vemos que cuando

$$\mathbf{V}ar_\theta(T(\mathbf{X})) = \frac{(\mathcal{T}'(\theta))^2}{nI(\theta)} \text{ y finito}$$

La desigualdad de Cauchy-Schwarz realmente es una desigualdad, de modo que,

$$\mathbf{T}(\mathbf{X}) - (\mathcal{T})(\theta) = K_\theta \frac{\partial}{\partial\theta} \ln f(\mathbf{x};\theta)$$

donde k_θ es alguna función de θ pero no de $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Integrando

$$\frac{1}{k_\theta}\left(\mathbf{T}(\mathbf{x}) - \mathcal{T}(\theta)\right) = \frac{\partial}{\partial\theta} \ln f(\mathbf{x};\theta)$$

resulta: $Q(\theta)\mathbf{T}(\mathbf{x}) + B(\theta) = \ln f(\mathbf{x};\theta)h(\mathbf{x})$ para algunas funciones Q, B, h.

Resulta que

$$f(\mathbf{x};\theta) = h(\mathbf{x})e^{Q(\theta)\mathbf{T}(\mathbf{x})+B(\theta)}$$

que es una familia exponencial de un parámetro y en consecuencia $\mathbf{T}(\mathbf{X})$ es un estadístico suficiente. En este sentido, solo las familias exponenciales tienen estadísticas eficientes (suficientes).

Definición 5.4. Sea $\mathbf{T}_1$ y $\mathbf{T}_2$ dos estimadores insesgados para un parámetro $\mathcal{T}(\theta) = \theta$. Suponga que $E_\theta(\mathbf{T}_1^2) < \infty$, $E_\theta(\mathbf{T}_2^2) < \infty$. Se define la eficiencia de $\mathbf{T}_1$ relativo a $\mathbf{T}_2$ por

$$Ef_\theta.(\mathbf{T}_1|\mathbf{T}_2) = \frac{\mathbf{V}ar_\theta(T_1)}{\mathbf{V}ar_\theta(T_2)}$$

y se dice que $\mathbf{T}_1$ es más eficiente que $\mathbf{T}_2$ si $Ef_\theta.(\mathbf{T}_1|\mathbf{T}_2) < 1$

Definición 5.5. Asuma que las condiciones de regularidad de la desigualdad Cramer-Rao son satisfechas por la familia de densidades $\mathscr{P} = \{f_\theta(x) : x \in \mathbb{R}, \theta \in \Theta \subset \mathbb{R}\}$. Entonces se dice que un estimador insesgado $\mathbf{T}$ para para una función paramétrica $\mathcal{T}(\theta) = \theta$ es más eficiente para la familia $\mathscr{P}$ si

$$\mathbf{V}ar_\theta(T) = \frac{1}{I_n} = \frac{1}{E_\theta\left\{\frac{\partial \ln f(x|\theta)}{\partial\theta}\right\}^2}$$

Definición 5.6. Sea $\mathbf{T}$ el estimador más eficiente para la familia de densidades $\mathscr{P} = \{f_\theta(x) : x \in \mathbb{R}, \theta \in \Theta \subset \mathbb{R}\}$. Entonces la eficiencia de cualquier estimador insesgado $\mathbf{T}_1$ de θ es definida como

$$Ef_\theta(\mathbf{T}_1|\mathbf{T}) = \frac{\mathbf{V}ar_\theta(T_1)}{\mathbf{V}ar_\theta(T)}$$

Evidentemente la eficiencia del estimador más eficiente es 1, y la eficiencia de cualquier estimador insesgado $\mathbf{T}_1$ es mayor que 1

Definición 5.7. Se dice que un estimador T_1 es asintóticamente eficiente si

$$\lim_{\mathbf{n}\to\infty} Ef_\theta(T_1) = 1$$

y $\mathbf{T}_1$ es al menos asintóticamente insesgado en el sentido que

$$\lim_{\mathbf{n}\to\infty} E_\theta(T_1) = \theta$$

6. Estimación Insesgado : caso vector

En esta lección, extendemos los resultados de algunos procedimientos al caso donde el espacio de parámetros Θ es un subconjunto de $\mathbb{R}$ para $k > 1$. Hemos introducido símbolos especiales para vectores de parámetros; Dejamos que el contexto determine su dimensionalidad. Recuerde primero que se dice que una variable aleatoria de valor real X se dice que es integrable (con respecto a $\mathbf{P}$) si $\mathbf{E}[|X|]$

Definición 6.1. Para $\mathbf{x} \in \mathbb{R}^m$, el vecto columna $\mathbf{x} = (x_1, x_2, \ldots, x_{\mathbf{m}})^t$ donde t denota transpuesta. La norma de $\mathbf{x}$ se define por

$$\|\mathbf{x}\| = \left[\sum_{i=1}^{\mathbf{m}} X_i^2\right]^{\frac{1}{2}} = (\mathbf{x}'\mathbf{x})^{\frac{1}{2}}$$

UN vector $\mathbf{X} = (X_1, X_2, \ldots, X_{\mathbf{m}})^t$ es una función medible de $(\Omega, \mathscr{A}, \mathbf{P})$ para $(\mathbb{R}^{\mathbf{m}}, \mathfrak{B}^{\mathbf{m}})$ donde cada X_j es una variable aleatoria real. La $\|\mathbf{x}\|$ es la variable aleatoria real $(\mathbf{x}'\mathbf{x})^{\frac{1}{2}}$. Si cada componente X_j en $\mathbf{X}$ es integrable, en media (vector $\mathbf{m}\times 1$) $\mathscr{E}[\mathbf{X}] = (\mathbf{E}[X_1], \mathbf{E}[X_2], \ldots, \mathbf{E}[X_{\mathbf{m}}])^t$

Ejercicio 14.

Pruebe los siguiente corolarios

(a) $\|\mathbf{x}\|^2 = \mathbf{x}^t\mathbf{x} = \operatorname{tr}(\mathbf{x}'\mathbf{x}) = \operatorname{tr}(\mathbf{x}\mathbf{x}') = \operatorname{tr}(\mathbf{x}\mathbf{x}')$ donde

$$\operatorname{tr}(B) = \sum_{i=1}^{\mathbf{m}} b_{ii}$$

es la traza de la matriz $B = [b_{ii}]$ de orden $\mathbf{m} \times \mathbf{m}$.

(b) $\mathscr{E}[\|\mathbf{X}\|] < \infty$ si y solo si $\mathbf{E}(|X_j|) < \infty$ para $j = 1(1)m$
Sugerencia: primero demuestre que
$\|\mathbf{X}\| \leq |X_1| + |X_2| + \cdots + |X_{\mathbf{m}}| \leq \mathbf{m}\|\mathbf{X}\|$.

(c) Cuando $\mathscr{E}[\|\mathbf{X}\|^2] < \infty$ También lo es $\mathscr{E}[\|\mathbf{X}\|]$
Sugerencia: use la desigualdad Cauchy-Schwarz

(d) Cuando $\mathscr{E}[\|\mathbf{X}\|^2] < \infty$, $\mathscr{E}[\|\mathbf{X} - \mathscr{E}[\mathbf{X}]\|^2] < \infty$ y
$\mathbf{E}[(X_j - \mathbf{E}(X_j))^2] < \infty$ para $j = 1(1)\mathbf{m}$

Definición 6.2. La matriz covarianza del vector aleatorio $\mathbf{X}$ se define por

$$\begin{aligned} \mathbf{Cov}(\mathbf{X}) = \Sigma_{\mathbf{X}} &= \mathscr{E}[(\mathbf{X} - \mathscr{E}[\mathbf{X}])(\mathbf{X} - \mathscr{E}[\mathbf{X}])^t] \\ &= [\sigma_{ij}] \quad \text{para } ij = 1, \ldots, \mathbf{m} \end{aligned}$$

donde $\sigma_{ij} = \mathbf{E}[(X_i - \mathbf{E}(X_i))(X_j - \mathbf{E}(X_j))]$ son todas finitas. Si $\mathbf{X}$ es “sobreentendido” $\Sigma_{\mathbf{X}}$ puede ser escrito como $\Sigma = [\sigma_{ij}]$

Ejercicio 15.

Para definición anterior Demuestre que
$\Sigma_{\mathbf{X}} = \mathscr{E}(\mathbf{X}\mathbf{X}^t) - (\mathscr{E}[\mathbf{X}])(\mathscr{E}[\mathbf{X}])^t$.

Lema 6.1. *La matriz covarianza $\Sigma_{\mathbf{X}}$ es simétrica. Además $\Sigma_{\mathbf{X}}$ es no negativa : $y \in \mathbb{R}^m$, $y'\Sigma_{\mathbf{X}} y \geq 0$.*

Prueba. Como $\mathscr{E}[X_iX_j] = \mathscr{E}[X_jX_i]$, la simetría es obvia. Ahora,

$$\begin{aligned} y'\Sigma_{\mathbf{X}}y &= y^t\mathscr{E}[(\mathbf{X}-\mathscr{E}[\mathbf{X}])(\mathbf{X}-\mathscr{E}[\mathbf{X}])^t]y \\ &= \mathscr{E}[y^t(\mathbf{X}-\mathscr{E}[\mathbf{X}])(\mathbf{X}-\mathscr{E}[\mathbf{X}])^ty] \\ &= \mathscr{E}[y^t(\mathbf{X}-\mathscr{E}[\mathbf{X}])(y^t(\mathbf{X}-\mathscr{E}[\mathbf{X}]))^t] \\ &= \mathscr{E}[(y^t(\mathbf{X}-\mathscr{E}[\mathbf{X}]))^2] \geq 0 \end{aligned}$$

La no-negatividad se puede simbolizar por $\Sigma_{\mathbf{X}} \geq 0$ esto también permita escribir para vectores aleatorios $\mathbf{X}_1, \mathbf{X}_2$, $\Sigma_{\mathbf{X}_1} \leq \Sigma_{\mathbf{X}_2}$ si

$$y^t(\Sigma_{\mathbf{X}_1} - \Sigma_{\mathbf{X}_2})y \geq 0 \quad \text{para tod } y \in \mathbb{R}^{\mathbf{m}}$$

$\square$

Ejercicio 16.
Demuestre que este “orden parcial” cumple con la propiedad de transitividad

$$\Sigma_{\mathbf{X}_1} \leq \Sigma_{\mathbf{X}_2} \wedge \Sigma_{\mathbf{X}_2} - \Sigma_{\mathbf{X}_3} \Rightarrow \Sigma_{\mathbf{X}_1} \leq \Sigma_{\mathbf{X}_3}$$

Definición 6.3. Sea el modelo paramétrico

$$\mathscr{P} = \{f(x;\theta) : \mathbf{x} \in \mathbb{R}^{\mathbf{m}}, \theta \in \Theta \subseteq \mathbb{R}^k\}$$

con una muestra aleatoria $X^1, X^2, \ldots, X^{\mathbf{n}}$. Sea $\Phi : \Theta \to \Phi(\Theta)$; entonces $\mathbf{T} = \mathbf{t}(X^1, X^2, \ldots, X^{\mathbf{n}})$ es un estimador de $\Phi(\Theta)$ si $\mathbf{t}(X^1, X^2, \ldots, X^{\mathbf{n}}) \in \Phi(\Theta)$ para todo punto muestrales $(X^1, X^2, \ldots, X^{\mathbf{n}})$ (c.s.). El estadístico $\mathbf{T}$ es también insesgado si $\mathbf{E}_\theta[\mathbf{T}] = \Phi(\theta)$ para todo $\theta \in \Theta$. A diferencia del caso unidimensional, aquí restringimos el rango de $\mathbf{T}$ para $\Phi(\Theta)$ para eliminar estimadores “absurdos”. El siguiente ejemplo muestra que las propiedades básicas de la media y la varianza muestral se transfieren al caso multidimensional.

Ejemplo 6.1. Continúe de la definición; escriba $\sum_{i=1}^{\mathbf{n}}$ como $\sum$.

(a) La media muestral es un estimador insesgado de la media poblacional. Suponga que $\Phi(\theta) = \mathscr{E}_\theta[\mathbf{X}]$ es finito en $\mathbb{R}^{\mathbf{m}}$. Entonces, $\mathbf{T}(X^1, X^2, \ldots, X^{\mathbf{m}}) = \overline{\mathbf{X}} = \frac{1}{\mathbf{n}}\sum X^i$ está en $\mathbb{R}^{\mathbf{m}}$ y $\mathscr{E}_\theta[\mathbf{T}] = \frac{1}{\mathbf{n}}\sum \mathscr{E}_\theta[\mathbf{X}^i] = \mathscr{E}_\theta[\mathbf{X}]$.

(b) Suponga que la matriz covarianza existe:

$$\Phi(\theta) = \Sigma_\theta.$$

Sea

$$\begin{aligned}\mathbf{S} &= \mathbf{S}(X^1, X^2, \ldots, X^{\mathbf{m}})\\ &= \frac{1}{\mathbf{n}-1}\sum[(X^i - \mathbf{T})(X^i - \mathbf{T})'] \text{ y } \mathscr{E}_\theta[X] = \mathscr{E}[X^i] = \boldsymbol{\mu}.\end{aligned}$$

Ahora

$$\begin{aligned}&= \sum[(X^i - \boldsymbol{\mu})(X^i - \boldsymbol{\mu})']\\ &= \sum[(X^i - \mathbf{T} + \mathbf{T} - \boldsymbol{\mu})(X^i - \mathbf{T} + \mathbf{T} - \boldsymbol{\mu})']\\ &= \sum[(X^i - \mathbf{T})(\mathbf{x}^i - \mathbf{T})' + \sum(X^i - \mathbf{T})(\mathbf{T} - \boldsymbol{\mu})\\ &\quad + \sum(\mathbf{T} - \boldsymbol{\mu})(X^i - \mathbf{T})' + \sum(\mathbf{T} - \boldsymbol{\mu})(\mathbf{T} - \boldsymbol{\mu})'\end{aligned}$$

como

$$\sum(X^i - \mathbf{T})(\mathbf{T} - \boldsymbol{\mu})' = \left[\sum(X^i - \mathbf{T})\right](\mathbf{T} - \boldsymbol{\mu})' = 0.$$

Lo anterior puede ser escrito como

$$(\mathbf{n}-1)\mathbf{S} = \sum(X^i - \boldsymbol{\mu})(X^i - \boldsymbol{\mu})' - \mathbf{n}(\mathbf{T} - \boldsymbol{\mu})(\mathbf{T} - \boldsymbol{\mu})'$$

Tomando la esperanza a ambos miembros de esta última igualdad se tiene

$$(\mathbf{n}-1)\mathscr{E}_\theta[\mathbf{S}] = \mathbf{nCov}(\mathbf{X}) - \mathbf{nCov}(\mathbf{T}) \tag{6.1}$$

(c) La covarianza de la media muestral de la media muestral es $\frac{1}{\mathbf{n}}\Sigma_\theta$, es decir la matiz covarianza e la población dividido entre $\mathbf{n}$. Como

$$\begin{aligned}
T-\boldsymbol{\mu} &= \frac{\sum X^i}{\mathbf{n}} - \boldsymbol{\mu} = \frac{1}{\mathbf{n}}\left[\sum (X^i - \boldsymbol{\mu})\right],\\
\mathbf{Cov}(\mathbf{T}) &= \mathscr{E}_\theta[(\mathbf{T}-\boldsymbol{\mu})(\mathbf{T}-\boldsymbol{\mu})']\\
&= \mathscr{E}_\theta\left[\left\{\frac{1}{\mathbf{n}}\sum (X^i-\boldsymbol{\mu})\right\}\left\{\frac{1}{\mathbf{n}}\sum (X^i-\boldsymbol{\mu})\right\}'\right]\\
&= \frac{1}{\mathbf{n}^2}\sum_{i=1}^{\mathbf{n}}\sum_{j=1}^{\mathbf{n}} \mathscr{E}_\theta\left[(X^i-\boldsymbol{\mu})(X^j-\boldsymbol{\mu})'\right]
\end{aligned}$$

Cuando $i \neq j$, X^i y X^j son independientes de modo que

$$\mathscr{E}_\theta\left[(X^i-\boldsymbol{\mu})(X^j-\boldsymbol{\mu})'\right] = \begin{cases} 0 & \text{si} \quad i \neq j \\ \Sigma_\theta & \text{si} \quad i = j \end{cases}$$

$$\text{Por lo tanto },\mathbf{Cov}(\mathbf{T}) = \frac{1}{\mathbf{n}}\mathbf{Cov}(\mathbf{X}) = \frac{1}{\mathbf{n}}\Sigma_\theta \tag{6.2}$$

(d) La covarianza muestral $\mathbf{S}$ es un estimador insesgado de la matriz covarianza poblacional. De la igualda (6.1) y ((c)), resulta

$$(\mathbf{n}-1)\mathscr{E}_\theta[\mathbf{S}] = \mathbf{n}\Sigma_\theta - \mathbf{n}\frac{\Sigma_\theta}{\mathbf{n}} = \Sigma_\theta$$

■

Ejercicio 17.
Sea A una matriz de constantes de orden $P \times \mathbf{m}$; sea X un vector aleatorio de orden $\mathbf{m}$ dimensional, Demuestre que:

(a) $\mathscr{E}[\mathbf{AX}] = \mathbf{A}\mathscr{E}[\mathbf{X}]$ cuando $\mathscr{E}[\mathbf{X}]$ exista.

(b) $\mathbf{Cov}(\mathbf{AX}) = \mathbf{A}\mathbf{Cov}(\mathbf{X})\mathbf{A}'$ cuando $\mathbf{Cov}(\mathbf{X})$ exista.

A menos que se indique lo contrario, tomaremos la *función de pérdida cuadrática* en el caso del vector como

$$L(\boldsymbol{\theta}, \mathbf{a}) = \|\mathbf{X} - \mathbf{a}\|^2 = (\boldsymbol{\theta} - \mathbf{a})'(\boldsymbol{\theta} - \mathbf{a}) \text{ para } \boldsymbol{\theta}, \mathbf{a} \in \mathbb{R}^k$$

En general se puede considerar

$$(\boldsymbol{\theta} - \mathbf{a})'\mathbf{A}(\theta)(\boldsymbol{\theta} - \mathbf{a}) \text{ para cualquier } \mathbf{A}(\boldsymbol{\theta}) \geq 0,$$

Por lo tanto en particular, para la matriz identidad $\mathbf{I}$ en $\mathbb{R}^k$.
En cualquier caso, Para una función pérdida L dada, El riesgo de un estimador $\mathbf{T}$ se define como

$$\boldsymbol{R}(\Theta, \mathbf{T}) = \mathscr{E}_\theta \left[\|L(\theta, \mathbf{T})\| \right].$$

Este es $\mathscr{E}_\theta \left[\|\boldsymbol{\theta} - \mathbf{T}\|^2 \right]$ cuando $\Phi(\boldsymbol{\theta}) = \boldsymbol{\theta}$ y $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{I}$

Lema 6.2. Sea $\mathbf{T}$y $\mathbf{S}$ estimadores insesgados de $\Phi(\boldsymbol{\theta})$ tal que $\Sigma_\theta(\mathbf{T}) \leq \Sigma_\theta(\mathbf{S})$. Entonces $\boldsymbol{R}(\theta, \mathbf{T}) = \mathscr{E}_\theta \left[\|\Phi(\boldsymbol{\theta}) - \mathbf{T}\|^2 \right] \leq \boldsymbol{R}(\boldsymbol{\theta}, \mathbf{S})$

Prueba. (a) Por hipótesis se tiene,

$$y'(\Sigma_\theta(\mathbf{S}) - \Sigma_\theta(\mathbf{T}))y \geq 0 \,\forall\, y \in \mathbb{R}^k$$

En particular, para $y = (1, 0, 0, \ldots, 0)$ resulta
$\sigma_{11}(\mathbf{S}) - \sigma_{11}(\mathbf{T}) \geq 0$
Para $y = (0, 1, 0 \ldots, 0)$, se obtiene $\sigma_{22}(\mathbf{S}) - \sigma_{22}(\mathbf{T}) \geq 0$; etc.
Resulta que

$$\operatorname{tr}\left[\Sigma_\theta(\mathbf{S}) - \Sigma_\theta(\mathbf{T})\right] \geq 0$$

(b) Ahora $\mathscr{E}_\theta\left[\|\Phi(\boldsymbol{\theta})-\mathbf{T}\|^2\right]=\mathscr{E}_\theta\left[(\boldsymbol{\theta}-\mathbf{T})'(\Phi(\boldsymbol{\theta})-\mathbf{T})\right]$

$$
\begin{aligned}
&= \mathscr{E}_\theta\left[\operatorname{tr}\left\{(\boldsymbol{\theta}-\mathbf{T})'(\Phi(\boldsymbol{\theta})-\mathbf{T})\right\}\right]\\
&= \mathscr{E}_\theta\left[\operatorname{tr}\left\{(\boldsymbol{\theta}-\mathbf{T})(\Phi(\boldsymbol{\theta})-\mathbf{T})'\right\}\right]\\
&= \operatorname{tr}\mathscr{E}_\theta\left\{(\boldsymbol{\theta}-\mathbf{T})(\Phi(\boldsymbol{\theta})-\mathbf{T})'\right\}\\
&= \operatorname{tr}\left[\Sigma_\theta(\mathbf{T})\right].
\end{aligned}
$$

Entonces $\boldsymbol{R}(\boldsymbol{\theta},\mathbf{S})-\boldsymbol{R}(\boldsymbol{\theta},\mathbf{T})=\operatorname{tr}\left[\Sigma_\theta(\mathbf{S})\right]-\operatorname{tr}\left[\Sigma_\theta(\mathbf{T})\right]$
$=\operatorname{tr}\left[\Sigma_\theta(\mathbf{S})-\Sigma_\theta(\mathbf{T})\right]$.
Esta es no-negativa por el ítem(a).
Este sugiere que se puede restringir la selección de estimadores $\mathbf{T}$ a aquellos que son insesgados con matriz covarianza "mínima":para cada estimador insesgado $\mathbf{S}$
, $\Sigma_\theta(\mathbf{S})-\Sigma_\theta(\mathbf{T})$

□

Ejercicio 18.
Sea $\Phi(\boldsymbol{\theta})=(\Phi_1(\boldsymbol{\theta}),\Phi_2(\boldsymbol{\theta}).\dots,\Phi_k(\boldsymbol{\theta}))\in\mathbb{R}^k$ y
$\mathbf{T}(X^1,X^2,\dots,X^{\mathbf{n}})=(\mathbf{T}_1(X^1,X^2,\dots,X^{\mathbf{n}}),\dots,\mathbf{T}_{\mathbf{n}}(X^1,X^2,\dots,X^{\mathbf{n}}))$
sea un estimador insesgado de $\Phi(\boldsymbol{\theta})$. Demuestre que si $\mathbf{T}$tiene una matriz covarianza mínima entre todos los estimadores insesgados de $\Phi(\boldsymbol{\theta})$, entonces para cada $j=1,2,\dots,k$, $\mathbf{T}_j$ es un estimador insesgado de $\Phi_j(\boldsymbol{\theta})$ con varianza mínima.

El Teorema que se da a continuación (ver por ejemplo Roa, 1965) contiene extensiones para el caso vector de los principales resultados para parámetros reales.

Teorema 6.1. Sea $\mathbf{T}$ un estimador insesgado de $\Phi(\boldsymbol{\theta})\in\mathbb{R}^k$. Sea $\mathbf{S}$ un estadístico suficiente para $\boldsymbol{\theta}\in\mathbb{R}^k$. Si Φ es diferenciable, sea $D(\boldsymbol{\theta})$ una matriz $k\times k$ de las derivadas parciales $\frac{\partial\Phi_i}{\partial\Phi_j}$,

(a) $\mathscr{E}[\mathbf{T}|\mathbf{S}]$ es un estimador insesgado de $\varPhi(\boldsymbol{\theta})$ con matriz covarianza menor que la varianza de $\mathbf{T}$.

(b) Si $\mathbf{S}$ es también completo, entonces $\mathscr{E}[\mathbf{T}|\mathbf{S}]$ es un estimador insesgado de $\varPhi(\boldsymbol{\theta})$ con matriz covarianza mínima.

(c) Según extensiones apropiadas de las condiciones de regularidad, la matriz de información de Fisher la matriz de orden $k \times k$

$$\mathbf{I}(\boldsymbol{\theta}) = \left[\mathscr{E}_\theta \left\{ \Psi \frac{\partial}{\partial \theta_i} \log(f(\mathbf{X};\boldsymbol{\theta}))\varPhi . \Psi \frac{\partial}{\partial \theta_j} \log(f(\mathbf{X};\boldsymbol{\theta}))\varPhi \right\}\right].$$

Si este existe, $\mathbf{I}(\boldsymbol{\theta})$ es definida positiva y no-singular.

(d) Resulta del ítem (c) que $\Sigma_\theta(\mathbf{T}) \geq D(\boldsymbol{\theta})\,[\mathbf{I}(\boldsymbol{\theta})]^{-1}\,D'(\boldsymbol{\theta})$.

7. Estadísticos Suficientes y Completos

Supongamos que S_1 y S_2 son dos estimadores insesgados de $g(\theta)$ (con varianzas finitas) y supongamos que T es un estadístico suficiente para θ. Podemos definir $S_1^* = E(S_1|T)$ y $S_2^* = E(S_2|T)$. Aunque $Var_\theta(S_i^) \leq Var_\theta(S_i)$ (para $i = 1, 2$), no hay manera de saber a priori si $^*_1$ o S_2^* tendrán una menor varianza. Sin embargo, para una elección adecuada de T, nos gustaría tener

$$Var_\theta(S_1^*) = Var_\theta(S_2^*) \text{para todo}\, \theta, \text{o}\, P_\theta(S_1^* = S_2^*) = 1 \text{para todo}\, \theta.$$

Más precisamente, para esta elección particular de T, sólo habrá un estimador insesgado que sea una función de T y, si T es un estadístico minimal suficiente, este estimador insesgado tendrá varianza mínima.

Ejemplo 7.1. Supongamos que $X_1,\ldots,X_n$ son variables aleatorias iid. Poisson con media λ y considere estimadores insesgados de

$g(\lambda) = exp(-\lambda) = P_\lambda(X_i = 0)$. Dos estimadores insesgados de $g(\lambda)$ son

$$S_1 = I(X_1 = 0) \quad \text{y} \quad S_2 = \frac{1}{n}\sum_{i=1}^{n} I(X_i = 0).$$

Un estadístico minimal y suficiente para λ es $T = \sum_{i=1}^{n} X_i$. Para encontrar los estimadores "Rao-Blackwelizados", debemos hallar las distribuciones condicionales de S_1 y $_S2$ dado $T = t$. Es fácil ver que

$$\begin{aligned} P(S_1 = 1|T = t) &= \frac{P_\lambda(S_1 = s, T = t)}{P_\lambda(T = t)} \\ &= \left(1 - \frac{1}{n}\right)^t \end{aligned}$$

y en consecuencia

$$S_1^* = E(S_1|T) = \left(1 - \frac{1}{n}\right)^T$$

Para S_2, definamos $U = \sum_{i=1}^{n} I(X_i = 0)$; obtenemos

$$P(U = u|T = t) = \binom{n}{u}\left(1 - \frac{1}{n}\right)^t$$

para $u = \max(n - t, 0), \ldots, \max(n - 1, n - t)$. Resulta entonces que

$$E(U|T = t) = \sum_u u\binom{n}{u}\left(1 - \frac{1}{n}\right)^t = n\left(1 - \frac{1}{n}\right)^t$$

y así $S_2^* = (1 - 1/n)^T = S_1^*$. Por lo tanto, en este ejemplo, el "Rao-Blackwelización" S_1 y S_2 conduce al mismo estimador. ■

Definición 7.1. Supongamos que $X_1, \ldots, X_n$ son variables aleatorias cuya distribución conjunta depende de algún parámetro θ desconocido. Un estadístico $T = T(X_1, \ldots, X_n)$ (posiblemente vector de valores) se dice que es completo para θ si para cualquier función g, $E_\theta[g(T)] = 0$ para todo θ implica que $P_\theta[g(T) = 0] = 1$ para todo θ.

La completitud de un estadístico T significa esencialmente que T no contiene información "accesoria" (sin sentido) sobre θ. Por ejemplo, si T es completo para θ entonces $g(T)$ es un estadístico auxiliar para θ si, y sólo si, $g(T)$ es constante sobre el rango de T. (Más precisamente, para alguna constante k, $P_\theta[g(T) = k] = 1$ para todo θ.) se puede demostrar que si un estadístico T es suficiente y completa entonces T es también un estadístico suficiente y mínimal. Sin embargo, un estadístico suficiente y mínimal no tiene que ser completo.

Ejemplo 7.2. Supongamos que $X_1, \ldots, X_n$ son variables aleatorias iid. Poisson con media λ y definimos $T = \sum_{i=1}^n X_i$. T por supuesto, es un estadístico suficiente para λ; para ver si T es completo, tenemos que ver si existe una función g tal que $E_\lambda[g(T)] = 0$ para todo $\lambda > 0$. Puesto que T tiene una distribución de Poisson con parámetro media $n\lambda$, tenemos

$$E_\lambda[g(T)] = \sum_{x=0}^{\infty} g(x) \frac{e^{-n\lambda}(n\lambda)^x}{x!}$$

y entonces $E_\lambda[g(T)] = 0$ para todo λ si, y solo si,

$$\sum_{x=0}^{\infty} g(x) \frac{(n\lambda)^x}{x!} = 0 \text{ para todo } \lambda > 0.$$

Como $\sum_{k=0}^{\infty} c_k \lambda^k = 0$ para todo $a < \lambda < b$ si, t solo si, $c_k = 0$ para $k \geq 0$, resulta que $E_\lambda[g(T)] = 0$ para todo $\lambda > 0$ si, y solo si $g(x)n^x/x! = 0$ para $x = 0, 1, \ldots$. En consecuencia $E_\lambda[g(T)] = 0$ para todo $\lambda > 0$ implica que $P_\lambda(g(T) = 0) = 1$ para todo $\lambda > 0$. ■

Ejemplo 7.3. Supóngase que $X_1,\ldots,X_n$ son variables aleatorias iid. discretas con función de probabilidad

$$f(x;\theta) = \begin{cases} \theta, & \text{para} x = -1 \\ (1-\theta)^2\theta^x, & \text{para} x = 0, 1, 2, \ldots \end{cases}$$

donde $0 < \theta < 1$. La función de probabilidad conjunta de $X = (X_1, \ldots, X_n)$ puede ser escrita como una familia exponencial de dos parámetros:

$$f(x;\theta) = e^{[c_1(\theta)T_1(x)+c_2(\theta)T_2(x)+2\ln(1-\theta)]}$$

donde $c_1(\theta) = \ln(\theta) - 2\ln(1-\theta)$, $c_2(\theta) = \ln(\theta)$ y

$$T_1(x) = \sum_{i=1}^{n} I(x_i = -1)$$
$$T_2(x) = \sum_{i=1}^{n} I(x_i x_i \geq 0).$$

Por lo tanto $(T_1(X), T_2(X))$ es un estadístico suficiente (de hecho estadístico suficiente minimal) para θ. Sin embargo, este estadístico suficiente no es completo. Para ver esto, tenga en cuenta que

$$E_\theta(X_i) = -\theta + \sum x(1-\theta)^2\theta^x = 0$$

Como

$$\begin{aligned}\sum_{i=1}^{n} X_i &= \sum_{i=1}^{n} I(x_i x_i \geq 0) - \sum_{i=1}^{n} I(x_i = -1) \\ &= T_2 - T_1,\end{aligned}$$

Resulta que $E_\theta(T_2 - T_1) = 0$ para todo θ. Como $P_\theta(T_2 = T_1) < 1$ para todo θ, resulta que el estadístico suficiente (T_2, T_1) no es completo. ■

El siguiente resultado da una condición para que un estadístico suficiente sea completa en una familia exponencial.

Teorema 7.1. Suponga que $X = (X_1, \ldots, X_n)$ tiene densidad conjunta o función de frecuencia conjunta que es una familia exponencial de p-parámetro:

$$f(x;\theta) = e^{\left[\sum_{i=1}^{p} c_i(\theta)T_i(x) - d(\theta) + S(x)\right]} \text{ para } x \in A$$

Defina $C = \{(c_1(\theta), \ldots, c_p(\theta)) : \theta \in \Theta\}$. Si el conjunto C contiene un conjunto abierto (rectángulos) de la forma $(a1, b1) \times \cdots \times (a_p, b_p)$, entonces el estadístico $(T_1(X), \ldots, T_p(X))$ es completo así como suficiente para θ.

La la prueba de este Teorema queda fuera del alcance de este libro; una prueba puede encontrarse en Lehmann (1991). Esencialmente este resultado es una consecuencia de la unicidad de funciones características. Este dice, hablando más o menos, que un estadístico suficiente p-dimensional en un familia exponencial de p-parámetros también es completo siempre que la dimensión del espacio de parámetros sea p.

Ejemplo 7.4. Supongamos $X_1, \ldots, X_n$ son Variables aleatorias iid. normales con medias μ y varianza σ^2. La densidad conjunta puede ser escrita como una familia exponencial de dos parámetros

$$f(x;\mu,\sigma) = e^{\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i + \frac{\mu}{\sigma^2} x_i - d(\mu,\sigma)\right]}$$

donde $d(\mu, \sigma) = n\mu^2/(2\sigma^2) + n\ln(\sigma) + n\ln(2\pi)/2$. Claramente el rango de la función es

$$c(\mu,\sigma) = \left(-\frac{1}{2\sigma}, \frac{\mu}{\sigma^2}\right)$$

para $-\infty < \mu < \infty$ y $\sigma > 0$ contiene un rectángulo en $\mathbb{R}$ de modo que el estadístico suficiente

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

es completo como suficiente para (μ, σ). Además, como con suficiencia, cualquier función uno-a-uno de un estadístico completo será también será completo; Así, por ejemplo, $(\overline{X}, \sum_{i=1}^{n}(X_i - \overline{X})^2)$ también es completo. ■

El siguiente resultado da un criterio simple para la existencia de un estimador **UMVU** cuando existe un estadístico completo y suficiente.

Definición 7.2. Una sucesión de estimadores $\{\widehat{\theta}_{\mathbf{n}}\}$ se dice que es una sucesión consistente para el parámetro θ si $\{\widehat{\theta}_{\mathbf{n}}\}$ converge en probabilidad para θ, esto es, si

$$\lim_{\mathbf{n}\to\infty} P_\theta[|\widehat{\theta}_{\mathbf{n}} - \theta| > \epsilon] = 0$$

para cada $\epsilon > 0$ y cada θ.

Aunque estrictamente hablando, la consistencia se refiere a una sucesión de estimadores, que frecuentemente decimos que $\widehat{\theta}_{\mathbf{n}}$ es un estimador consistente de θ si este es claro que $\widehat{\theta}_{\mathbf{n}}$ pertenece a una sucesión de estimadores bien definidas;un ejemplo de esto ocurre cuando $\widehat{\theta}_{\mathbf{n}}$ está basada en $\mathbf{n}$ variables aleatorias iid.

Ejemplo 7.5. Supóngase que $X_1,\ldots,X_{\mathbf{n}}$ son variables aleatorias iid. con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$. Se define

$$S_{\mathbf{n}}^2 = \frac{1}{\mathbf{n}-1}\sum_{i=1}^{\mathbf{n}}(X_i - \overline{X}_{\mathbf{n}})^2,$$

la cual es un estimador insesgado de σ^2. Para demostrar que $S^2_{\mathbf{n}}$ es un estimador consistente(o más correctamente $\{S^2_{\mathbf{n}}\}$ es una sucesión consistente de estimadores), tenga en cuenta que ■

$$\begin{aligned} S^2_{\mathbf{n}} &= \frac{1}{\mathbf{n}-1}\sum_{i=1}^{\mathbf{n}}(X_i-\mu+\mu-\overline{X}_{\mathbf{n}})^2 \\ &= \frac{1}{\mathbf{n}-1}\sum_{i=1}^{\mathbf{n}}(X_i-\mu)^2+\frac{\mathbf{n}}{\mathbf{n}-1}(\overline{X}_{\mathbf{n}}-\mu)^2 \\ &= \left(\frac{\mathbf{n}}{\mathbf{n}-1}\right)\frac{1}{\mathbf{n}-1}\sum_{i=1}^{\mathbf{n}}(X_i-\mu)^2+\frac{\mathbf{n}}{\mathbf{n}-1}(\overline{X}_{\mathbf{n}}-\mu)^2. \end{aligned}$$

Por la ley débil de los grandes números tenemos

$$\frac{1}{\mathbf{n}-1}\sum_{i=1}^{\mathbf{n}}(X_i-\mu)^2 \xrightarrow{P} \sigma^2 \qquad \text{y} \qquad \overline{X}_{\mathbf{n}} \xrightarrow{P} \mu$$

y de este modo por el teorema de Slutsky, resulta que

$$S^2_{\mathbf{n}} \xrightarrow{P} \sigma^2.$$

Debe tenerse en cuenta que $S^2_{\mathbf{n}}$ será un estimador consistente de $\sigma^2 = Var(X_i)$ para cualquier distribución con varianza finita.

Ejemplo 7.6. Supóngase que $X_1,\ldots,X_{\mathbf{n}}$ son variables aleatorias independientes con

$$E_\beta(X_i) = \beta t_i \qquad \text{y} \qquad Var_\beta(X_i) = \sigma^2.$$

donde $t_1,\ldots,t_{\mathbf{n}}$ son constantes conocidas y β, σ^2 son parámetros desconocidos. Un posible estimador de β es

$$\widehat{\beta}_{\mathbf{n}} = \frac{\displaystyle\sum_{i=1}^{\mathbf{n}} t_i X_i}{\displaystyle\sum_{i=1}^{\mathbf{n}} t_i^2}.$$

Es fácil demostrar que $\widehat{\beta}_{\mathbf{n}}$ es un estimador insesgado de β para cada $\mathbf{n}$ y en consecuencia para demostrar que $\widehat{\beta}_{\mathbf{n}}$ es consistente, es suficiente demostrar que $Var_{\beta}(\widehat{\beta}_{\mathbf{n}}) \to 0$. Debido a la independencia de las variables aleatorias X_i resulta que

$$Var_{\beta}(\widehat{\beta}_{\mathbf{n}}) = \frac{\sigma^2}{\displaystyle\sum_{i=1}^{\mathbf{n}} t_i^2}$$

De esta manera $\widehat{\beta}_{\mathbf{n}}$ es un estimador consistente supuesto que $\displaystyle\sum_{i=1}^{\mathbf{n}} t_i^2 \to \infty$ cuando $n \to \infty$. ■