Inferencia Estadística

Marisol García Peña

Departamento de Matemáticas
Pontificia Universidad Javeriana

Bogotá, 2022

Estimación puntual (cont.)

Ejemplo

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población geométrica con parámetro π. Muestre que $\overline{X}$ es una estadística suficiente para π. $P[X_i = x_i] = \pi(1-\pi)^{x-1}, x_i \geq 1$

Ejemplo

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población $U(0, \theta)$. $f(x_i, \theta) = \frac{1}{\theta}$, $0 < x < \theta, \theta > 0$. Muestre que $X_{(n)} = \max\limits_{1 \leq i \leq n} X_i$ es una estadística suficiente para θ.

Teorema de la factorización para estadísticas conjuntamente suficientes

Teorema

Sea $X_1, \ldots, X_n$ una muestra aleatoria de tamaño n de una densidad $f(\bullet, \theta)$, donde el parámetro θ puede ser un vector. Un conjunto de estadísticas $S_1(X_1, \ldots, X_n), \ldots, S_r(X_1, \ldots, X_n)$ es conjuntamente suficiente si y sólo si la densidad conjunta de $X_1, \ldots, X_n$ puede ser factorizada como

$$\begin{aligned} f(x_1, \ldots, x_n; \theta) &= g(s_1(x_1, \ldots, x_n), \ldots, s_r(x_1, \ldots, x_n), \theta) h(x_1, \ldots, x_n) \\ &= g(s_1, \ldots, s_r, \theta) h(\boldsymbol{x}) \end{aligned}$$

donde la función $h(x_1, \ldots, x_n)$ es no negativa y no depende de θ y la función $g(s_1, \ldots, s_r, \theta)$ es no negativa y depende de θ y del valor de la estadísticas $s_1, \ldots, s_r$.

Estimación puntual (cont.)

Ejemplo

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución $N(\mu, \sigma^2)$ donde μ y σ^2 son desconocidos, encuentre las estadísticas suficientes para los parámetros.

$$
\begin{aligned}
f_{\boldsymbol{x},\boldsymbol{\theta}}(x, \boldsymbol{\theta}) &= \prod_{i=1}^{n} f(x_i, \boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}x_i^2 - 2\mu\sum_{i=1}^{n}x_i + n\mu^2\right]\right\} \\
&= \underbrace{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\} \exp\left\{\frac{\mu}{\sigma^2}\sum_{i=1}^{n}x_i - \frac{\sum_{i=1}^{n}x_i^2}{2\sigma^2}\right\}}_{g(\boldsymbol{s},\boldsymbol{\theta})} \underbrace{1}_{h(\boldsymbol{x})}
\end{aligned}
$$

$S(X) = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ son est. conj. suficientes para $\theta = (\mu, \sigma^2)$.

Estimación puntual (cont.)

Teorema

Un estimador de máxima verosimilitud o un conjunto de estimadores de máxima verosimilitud dependen de la muestra a través de cualquier conjunto de estadísticas conjuntamente suficientes.

El máximo de $L(x_1, \ldots, x_n, \theta)$ se obtiene en el mismo punto que el máximo de $g(s_1, \ldots, s_r, \theta)$.

Teorema

Si S es una estadística suficiente para θ y si el EMV de θ es único, entonces es una función de S.

Estadísticas suficientes minimales

- Una estadística S es suficiente minimal si no se puede reducir más sin violar la propiedad de suficiencia.
- La estadística $\boldsymbol{S} = (S_1, \ldots, S_m)$ es suficiente minimal si m es la menor dimensión de $\boldsymbol{S}$ tal que $\boldsymbol{S}$ continue siendo suficiente para θ.
- Si para un problema existe un único estimador de máxima verosimilitud $\widehat{\theta}$ para θ y si $\widehat{\theta}$ es una estadística suficiente entonces $\widehat{\theta}$ es suficiente minimal.

Estimación puntual (cont.)

Un conjunto de estadísticas conjuntamente suficientes es suficiente minimal si y sólo si es una función de cualquier otro conjunto de estadísticas suficientes.

Teorema

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una función de densidad o de probabilidad $f(x, \boldsymbol{\theta})$. Se dice que $\boldsymbol{S}(\boldsymbol{X})$ es una estadística suficiente minimal si para qualquier par de puntos muestrales x y y se tiene que

$$\frac{f_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{\theta})}{f_{\boldsymbol{Y}}(\boldsymbol{y}, \boldsymbol{\theta})} \text{ no depende de } \boldsymbol{\theta} \Leftrightarrow \boldsymbol{S}(\boldsymbol{X}) = \boldsymbol{S}(\boldsymbol{Y})$$

Ejemplo

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución $N(\mu, \sigma^2)$ con μ conocido. Encuentre una estadística suficiente minimal para $\sigma^2 = \theta$.

$$f_{\boldsymbol{x}}(\boldsymbol{x}, \theta) = \left(\frac{1}{\sqrt{2\pi}}\sigma\right)^n \exp\left\{\frac{-n\mu^2}{2\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^n x_i\right\}$$

$$f_{\boldsymbol{Y}}(\boldsymbol{y}, \theta) = \left(\frac{1}{\sqrt{2\pi}}\sigma\right)^n \exp\left\{\frac{-n\mu^2}{2\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n y_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^n y_i\right\}$$

$$\frac{f_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{\theta})}{f_{\boldsymbol{Y}}(\boldsymbol{y}, \boldsymbol{\theta})} = \frac{\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^n x_i\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n y_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^n y_i\right\}} = \frac{\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i\right]\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n y_i^2 - 2\mu\sum_{i=1}^n y_i\right]\right\}}$$

$$= \exp\left\{\frac{1}{2\sigma^2}\left[\sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i^2\right] + \frac{\mu}{\sigma^2}\left[\sum_{i=1}^n x_i - \sum_{i=1}^n y_i\right]\right\}$$

No depende de $\sigma^2 \Leftrightarrow S_1(X) = \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ y $S_2(X) = \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$. Entonces $\boldsymbol{S}(\boldsymbol{X}) = (S_1(\boldsymbol{X}), S_2(\boldsymbol{X}))$ es una estadística suficiente minimal para σ^2.

Familia exponencial

Una familia de densidades de un parámetro (θ unidimensional) $f(x,\theta)$ que se puede expresar como

$$f(x,\theta) = c(\theta)h(x)\exp[Q(\theta)T(x)]$$

para $-\infty < x < \infty$ para todo θ y para una elección adecuada de funciones $c(\bullet), h(\bullet), Q(\bullet)$ y $T(\bullet)$, pertenece a la *familia exponencial* o *clase exponencial*.

Ejemplo

Sea $f(x,\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x)$. ¿Pertenece a la familia exponencial?

Con $c(\theta) = \theta, h(x) = I_{(0,\infty)}(x), Q(\theta) = -\theta, T(x) = x$, $f(x,\theta)$ pertenece a la familia exponencial.

Ejemplo

$f(x,\theta) = f(x,\lambda)$ es Poisson. ¿Pertenece a la familia exponencial?

$$
\begin{aligned}
P[X = x] &= \frac{e^{-\lambda}\lambda^x}{x!} I_{\{0,1,\dots\}}(x) \\
&= e^{-\lambda} \left(\frac{1}{x!} I_{\{0,1,\dots\}}(x) \right) \exp(x \log \lambda)
\end{aligned}
$$

$c(\lambda) = e^{-\lambda}, h(x) = (1/x!) I_{\{0,1,\dots\}}(x), Q(\lambda) = \log\lambda$ y $T(x) = x$, por tanto pertenece a la familia exponencial.

Estimación puntual (cont.)

Si $f(x,\theta) = c(\theta)h(x)\exp[Q(\theta)T(x)]$, entonces

$$\prod_{i=1}^{n} f(x_i,\theta) = c^n(\theta)\left[\prod_{i=1}^{n} h(x_i)\right]\exp\left[Q(\theta)\sum_{i=1}^{n} T(x_i)\right]$$

por el criterio de factorización $\sum_{i=1}^{n} T(X_i)$ es una estadística suficiente.

Estimación puntual (cont.)

Familia exponencial k-paramétrica

Una familia de densidades $f(\bullet, \theta_1, \ldots, \theta_k)$ que puede ser expresada como

$$f(x, \theta_1, \ldots, \theta_k) = c(\theta_1, \ldots, \theta_k) h(x) \exp\left\{ \sum_{j=1}^{k} Q_j(\theta_1, \ldots, \theta_k) T_j(x) \right\}$$

pertenece a la familia exponencial k-paramétrica.

Estimación puntual (cont.)

Ejemplo

Sea $X \sim Bin(n, \theta), 0 < \theta < 1$. Muestre que la distribución pertenece a la familia exponencial.

$$\begin{aligned} P[X = x] &= \binom{n}{x} \theta^x (1-\theta)^{n-x} I_{(0,\ldots,n)}(x) \\ &= \binom{n}{x} \exp[\log(\theta^x)] \exp[\log(1-\theta)^{n-x}] I_{(0,\ldots,n)}(x) \\ &= \underbrace{\binom{n}{x} I_{(0,\ldots,n)}(x)}_{h(x)} \underbrace{(1-\theta)^n}_{c(\theta)} \exp[\underbrace{x}_{T_1(X)} \underbrace{\log(\theta/(1-\theta))}_{Q_1(\theta)}] \end{aligned}$$

Estimación puntual (cont.)

Teorema

Sea $f(x, \boldsymbol{\theta})$ una función de densidad o de probabilidad de la familia exponencial k-paramétrica

$$f(x, \boldsymbol{\theta}) = c(\boldsymbol{\theta}) h(x) \exp\left\{ \sum_{j=1}^{k} Q_j(\boldsymbol{\theta}) T_j(x) \right\}$$

si se considera una muestra aleatoria de tamaño n de esa familia, se tiene que

$$P_j(X) = \sum_{i=1}^{n} T_j(X_i)$$

son conjuntamente suficientes para θ.

Ejemplo

En el ejemplo anterior de la Binomial, $j = 1$ solo un parámetro θ.

$$P_i(X) = \sum_{i=1}^{n} T_1(X_i) = \sum_{i=1}^{n} X_i$$

Entonces $P_1(X) = \sum_{i=1}^{n} X_i$ es una estadística suficiente para θ

Si

$$
\begin{aligned}
&f(x, \theta_1, \ldots, \theta_k) \\
&\qquad = c^n(\theta_1, \ldots, \theta_k)\left[\prod_{i=1}^{n} h(x_i)\right] \exp\left[\sum_{j=1}^{k} Q_j(\theta_1, \ldots, \theta_k) \sum_{i=1}^{n} T_j(x_i)\right]
\end{aligned}
$$

por el criterio de factorización $\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i)$ es un conjunto de estadísticas suficientes y minimales.

Estimación insesgada

- $ECM[T(\tau(\theta))] = E[[T - \tau(\theta)]^2] = V[T] + \{\tau(\theta) - E[T]\}^2.$
- Si T es un estimador insesgado de $\tau(\theta) \Longrightarrow E[T] = \tau(\theta)$ y $ECM[T(\tau(\theta))] = V[T].$
- Estimador de error cuadrático medio uniformemente mínimo entre los estimadores insesgados.
- Estimador con varianza uniformemente mínima entre los estimadores insesgados.

Estimación puntual (cont.)

Estimador insesgado de varianza uniformemente mínima - UMVUE

Sea $X_1, \ldots, X_n$ es una muestra aleatoria de $f(\bullet, \theta)$. Un estimador $T^*(X_1, \ldots, X_n)$ de $\tau(\theta)$ es definido como un estimador insesgado de varianza uniformemente mínima de $\tau(\theta)$ si y sólo si:

1. $E[T^*] = \tau(\theta) \Longrightarrow$ estimador insesgado.
2. $V[T^*] \leq V[T]$ para cualquier otro estimador $T(X_1, \ldots, X_n)$ de $\tau(\theta)$ que también es insesgado, $E[T] = \tau(\theta)$.

Uniformly minimum-variance unbiased estimator - UMVUE

Estimación puntual (cont.)

Límite inferior para varianza

Sea $X_1, \ldots, X_n$ es una variable aleaotria con función de densidad o probabilidad $f(\bullet, \theta)$. Sea $T = (X_1, \ldots, X_n)$ un estimador insesgado de $\tau(\theta)$. $f(\bullet, \theta)$ satisface las siguientes *condiciones de regularidad* (caso continuo):

1. $\frac{\partial}{\partial\theta} \log f(x, \theta)$ existe $\forall x$ y $\forall \theta$.
2. $$\frac{\partial}{\partial\theta} \int \cdots \int \prod_{i=1}^{n} f(x_i, \theta) dx_1 \cdots dx_n = \int \cdots \int \frac{\partial}{\partial\theta} \prod_{i=1}^{n} f(x_i, \theta) dx_1 \cdots dx_n.$$
3. $$\frac{\partial}{\partial\theta} \int \cdots \int t(x_1, \ldots, x_n) \prod_{i=1}^{n} f(x_i, \theta) dx_1 \cdots dx_n$$ $$= \int \cdots \int t(x_1, \ldots, x_n) \frac{\partial}{\partial\theta} \prod_{i=1}^{n} f(x_i, \theta) dx_1 \cdots dx_n.$$
4. $0 < E\left[\left[\frac{\partial}{\partial\theta} \log f(X, \theta)\right]^2\right] < \infty, \forall \theta.$

Teorema

Desigualdad de Cramér-Rao

Bajo las condiciones (i) a (iv) se tiene

$$V[T] \geq \frac{[\tau'(\theta)]^2}{nE\left[\left[\frac{\partial}{\partial\theta}\log f(X,\theta)\right]^2\right]}$$

donde $T(X_1,\ldots,X_n)$ *es un estimador insesgado de* $\tau(\theta)$*. La igualdad se obtiene si y sólo si existe una función* $K(\theta, n)$ *tal que*

$$\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X,\theta) = K(\theta,n)[t(x_1,\ldots,x_n)-\tau(\theta)].$$

La parte derecha se conoce como Cota inferior de Cramér-Rao para la varianza de estimadores insesgados de $\tau(\theta)$*.*

Estimación puntual (cont.)

- Proporciona una cota inferior para la varianza de estimadores insesgados.
- Si $T(X)$ es un estimador insesgado de $\tau(\theta)$ y si $V[T(X)] =$ cota de Cramér-Rao $\Longrightarrow$ $T(X)$ es un UMVUE de $\tau(\theta)$.
- Si $T(X)$ es un estimador insesgado de θ y si $V[T(X)] = \frac{1}{I_F(\theta)} \Longrightarrow$ $T(X)$ es un UMVUE de θ, $I_F(\theta)$ para una muestra.
- Definición alternativa para la cota de Cramér-Rao

$$V[T] \geq \frac{\left\{\frac{\partial}{\partial\theta} E[T(X)]\right\}^2}{I_F(\theta)}$$

con $I_F(\theta)$ para una muestra.

Si la estimación de máxima verosimilitud de θ, $\widehat{\theta}$ está dada por la solución de la ecuación

$$\frac{\partial}{\partial\theta}\log L(\theta, x_1, \ldots, x_n) \equiv \frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i, \theta) = 0$$

y si $T^* = t^*(X_1, \ldots, X_n)$ es un estimador insesgado de $\tau^*(\theta)$ cuya varianza coincide con la cota inferior de Cramér-Rao, entonces $t^*(X_1, \ldots, X_n) = \tau^*(\widehat{\theta})$.

Bajo estas condiciones un estimador de máxima verosimilitud es UMVUE.

Estimación puntual (cont.)

Si $T^* = t(X_1, \ldots, X_n)$ es un estimador insesgado de algún $\tau^*(\theta)$ cuya varianza coincide con la cota inferior de Cramér-Rao, entonces $f(\bullet, \theta)$ es miembro de la clase/familia exponencial y viceversa, si $f(\bullet, \theta)$ es miembro de la familia exponencial, entonces existe un estimador insesgado, T^* de alguna función $\tau^*(\theta)$, cuya varianza coincide con la cota inferior de Cramér-Rao.

Ejemplo

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población Bernoulli(θ), $0 < \theta < 1$. Se pide:

- Encuentre la información de Fisher.
- Sea $T(X) = \overline{X}$, muestre que $T(X)$ es un UMVUE de θ.

Ejemplo

Sea X una variable aleatoria con distribución Poisson(θ), $\theta > 0$. Sea la estadística $T(X) = \begin{cases} 1, & \text{si } x = 0 \\ 0, & \text{e.o.c.} \end{cases} \implies T(X) = 1_{[X=0]}$. Se pide calcular la cota inferior de Cramér-Rao y la $V[T(X)]$. ¿Qué se puede concluir?