Estimación Puntual

Cristopher Morales Ubal
e-mail: c.m.ubal@gmail.com

Supongamos que deseamos estudiar la familia de probabilidades $P = \left\{\mathbb{P}_{\underline{\theta}} : \underline{\theta} \in \Theta\right\}$ donde $\mathbb{P}_{\Theta}$ es de forma conocida solo que se desconoce el valor parámetrico $\underline{\theta} \in \Theta$.Un procedimiento para resolve este problema es elegir un vector $\underline{\theta} \in \Theta$ utilizando algun criterio conveniente basado en una muestra de tamaño n seleccionada.

Nota: Estos apuntes estan basados en el libro [1] con el cual aprendi estadistica durante mi epoca de estudiante en la Universidad Técnica Federico Santa María (UTFSM). Han sido escritos en $\LaTeX\,2_{\varepsilon}$de manera de tener un apunte personal para recordar conceptos importantes de estadistica.

Definición: Sea $P = \left\{\mathbb{P}^{X}_{\underline{\theta}} : \underline{\theta} \in \Theta\right\}$ donde $\mathbb{P}_{\Theta}$ una familia de probabilidades. Se dice que $x_1, x_2, \ldots, x_n$ es una muestra aleatoria (m.a.) proveniente de la familia P ssi $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes e idénticamente distribuidas.

Definición: Sea $\chi = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1, x_2, \ldots, x_n \text{ es m.a proveniente de } P\}$ entonces χ se llama espacio de información.

Definición: Sea $\underline{T} : \chi \longrightarrow \tau$ una función medible que no depende del parámetro $\underline{\theta} \in \Theta$. Entonces se dice que $\underline{T} = t\left(\underline{x}\right)$ es una estadistica basada en la información χ.
Observación: se denota $\underline{x} = (x_1, x_2, \ldots, x_n)$ m.a. de la población.

Definición: Dada una información en un espacio χ se define como estimador de $\underline{\theta} \in \Theta$ a toda estadística $\underline{T} = t\left(\underline{x}\right)$, tal que

$$\mathbb{P}\left(t\left(\underline{x}\right) - \underline{\theta}\right) = 0 \ , \ \forall \underline{\theta} \in \Theta$$

Observaciones:

1. Si $R\left(\underline{T}\right) \subset \Theta \Longrightarrow \underline{T} = t\left(\underline{x}\right)$ es un estimador de $\underline{\theta}$.
2. Si $x_1, x_2, \ldots, x_n$ es una m.a. se tiene que la función de densidad conjunta de $\underline{x}$ es $f_{\underline{\theta}}\left(\underline{x}\right) = \Pi_{i=1}^{n} f_{\underline{\theta}}\left(x_i\right)$
3. m.a. $\not\Leftrightarrow$ m.a.s.

Definición: Sea $x_1, x_2, \ldots, x_n$ m.a. proveniente de una familia de densidad $f\left(\bullet, \underline{\theta}\right)$. Se llama *momento muestral de orden r-ésimo* con respecto al origen a la expresión

$$m_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$$

Teorema: Sea $x_1, x_2, \ldots, x_n$ m.a. proveniente de una familia de densidad $f\left(\bullet, \underline{\theta}\right)$ y sea $\overline{x}_n = m_1$ (media muestral). Supongamos además que la población tiene varianza finita σ^2. Entonces

$$\sigma^2_{\overline{x}_n} = \mathbb{E}\left[\left(\overline{x}_n - \mathbb{E}\left(\overline{x}_n\right)\right)^2\right] = \frac{\sigma^2}{n}$$

Definición: Sea $\widehat{\underline{\theta}} = T\left(\underline{x}\right)$ estadistica basada en χ. Se dice que $\widehat{\underline{\theta}}$ es una estadistica suficiente de $\underline{\theta} \in \Theta$ si $\forall\ \underline{\theta}^* = T^*\left(\underline{x}\right)$ estadística de $\underline{\theta}$ se tiene que $f\left(\widehat{\underline{\theta}} \mid \underline{\theta}^*\right)$ es independiente de $\underline{\theta}$.
Obs: una estadística es suficiente cuando tenemos la misma capacidad informativa en χ que en $\widehat{\underline{\theta}} = t\left(\underline{x}\right)$.

Teorema (criterio de factorización de Neyman-Fisher)

Sea $\left\{\left(\chi\ ,\ \mathcal{B}_{\chi}\ ,\ \mathbb{P}^{X}_{\underline{\theta}}\right) : \underline{\theta} \in \Theta\right\}$ una familia de espacios de probabilidad y sea $f_{\underline{\theta}}$ función de densidad (cuantía) correspondiente a $P^{X}_{\underline{\theta}}$. Sea $\underline{T} = t\left(\underline{x}\right)$ una estadistica. Si existe una factorización del tipo:

$$f_{\underline{\theta}}\left(\underline{x}\right) = g\left(\underline{\theta}, t\left(\underline{x}\right)\right) \cdot h\left(\underline{x}\right) \ \ \forall\ \left(\underline{\theta}, \underline{x}\right) \in \Theta \times \chi$$

entonces T es una estadística suficiente para $\underline{\theta} \in \Theta$ (suficiente para la familia).

Ejemplo: Sea $x_1, x_2, \ldots, x_n$ m.a. de tamaño n proveniente de una familia $N(\mu, 1)$.
luego

$$\begin{aligned}
f(x_1, x_2, \ldots, x_n, \mu) &= \prod_{i=1}^{n} f(x_i, \mu) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_i - \mu}{1}\right)^2\right) \\
&= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right) \\
&= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 - 2\mu\sum_{i=1}^{n} x_i + n\mu^2\right)\right) \\
&= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\left(n\mu^2 - 2\mu\sum_{i=1}^{n} x_i\right)\right) \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) \\
&= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{n\mu}{2}\left(\mu - 2\cdot\frac{\sum_{i=1}^{n} x_i}{n}\right)\right) \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right)
\end{aligned}$$

definiendo

$$\widehat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$$

podemos escribir

$$f(x_1, x_2, \ldots, x_n, \mu) = g(\mu, \widehat{\mu}) \cdot h(x_1, x_2, \ldots, x_n)$$

donde

$$g(\mu, \widehat{\mu}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{n\mu}{2}(\mu - 2\widehat{\mu})\right) \ \wedge \ h(x_1, x_2, \ldots, x_n) = \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right)$$

por lo tanto

$$\widehat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$$

es una estadistica suficiente para μ.

∎

Método de los Momentos

Sea $\left\{\left(\chi\ ,\ \mathcal{B}_\chi\ ,\ \mathbb{P}_{\underline{\theta}}^{X}\right) : \underline{\theta} \in \Theta\right\}$ espacio de probabilidad y sea $f_{\underline{\theta}}(\underline{x})$ una función de densidad asociada a $\mathbb{P}_{\underline{\theta}}^{\underline{x}}$. El método de estimación de los momentos consite en igualar los momentos muestrales con los poblacionales:

$$\mu_r = m_r \iff \mathbb{E}(X^r) = \frac{\sum_{i=1}^{n} x_i^r}{n}\ ,\ r = 1, 2, ..., k$$

donde $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ es el vector parametrico a determinar, resultando de esta forma un sistema de k ecuaciones y k incognitas $(\theta_1, \theta_2, \ldots, \theta_k)$

Ejemplo: Sea $x_1, x_2, \ldots, x_n$ una m.a. de tamaño n proveniente de una familia $N(\mu, \sigma^2)$.
en este caso

$$\underline{\theta} = (\mu, \sigma^2)$$

luego aplicando el métodos de los momentos se tiene que:

$$\mu_1 = m_1 \implies \mathbb{E}(X) = \frac{\sum_{i=1}^n x_i}{n} \iff \mu = \frac{\sum_{i=1}^n x_i}{n} \tag{1}$$

$$\mu_2 = m_2 \implies \mathbb{E}\left(X^2\right) = \frac{\sum_{i=1}^n x_i^2}{n} \iff \sigma^2 + \mu^2 = \frac{\sum_{i=1}^n x_i^2}{n} \tag{2}$$

asi resolviendo las ecuaciones (1) y (2) para μ y σ^2 se obtiene

$$\widehat{\mu} = \frac{\sum_{i=1}^n x_i}{n} = \overline{X}_n$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2 = \overline{X^2}_n - \left(\overline{X}_n\right)^2$$

los cuales son los estimadores obtenidos por el método de los momentos para μ y σ^2.

■

Método de Máxima Verosimilitud

Definición: sea $x_1, x_2, \ldots, x_n$ m.a.de tamaño n, proveniente de la familia de espacios de probabilidades $\left(\chi\ ,\ \mathcal{B}_\chi\ ,\ \mathbb{P}_{\underline{\theta}}^X\right)$, $\underline{\theta} \in \Theta$ se llama *Función de Verosimilitud* a $L_{\underline{x}} : \Theta \longrightarrow \mathbb{R}$ definida por:

$$L_{\underline{x}}(\underline{\theta}) = f_{\underline{\theta}}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^n f(x_i, \theta)\ ,\ \underline{\theta} \in \Theta$$

Definición: Sea $x_1, x_2, \ldots, x_n$ una m.a. de tamaño n, proveniente de una familia $f(\underline{x}, \underline{\theta})$ y $L_{\underline{x}}(\underline{\theta})$ su función de verosimilitud.
Se llama *Estimador de Maxima Verosimilitud* de $\underline{\theta}$ y se denota por $\widehat{\underline{\theta}}$ al $\widehat{\underline{\theta}} = t(x_1, x_2, \ldots, x_n)$ que maximiza a $L_{\underline{x}}(\underline{\theta})$.
Es decir $\widehat{\underline{\theta}}$ es el estimador de máxima verosimilitud de $\underline{\theta}$ ssi

i) $\widehat{\underline{\theta}}$ es un estimador de $\underline{\theta}$.

ii) $L_{\underline{x}}\left(\widehat{\underline{\theta}}\right) = \sup_{\underline{\theta} \in \Theta} L_{\underline{x}}(\underline{\theta})\ ,\ \forall \underline{x} \in \chi$

Obs:

1. En general si se satisfacen cierta condiciones de regularidad $L_{\underline{x}}(\underline{\theta})$ es maximo ssi $\nabla L_{\underline{x}}(\underline{\theta}) = \overrightarrow{0}$. es decir

$$\frac{\partial L_{\underline{x}}(\underline{\theta})}{\partial \theta_1} = 0\ ,\ \forall i = 1, 2, \ldots, k$$

2. El problema de maximización

$$\max_{\underline{\theta} \in \Theta} L_{\underline{x}}(\underline{\theta})\ ,\ (*)$$

es equivalente al problema de maximización

$$\max_{\underline{\theta} \in \Theta} \ln(L_{\underline{x}}(\underline{\theta}))\ ,\ (**)$$

asi en general, se resuelve el problema $(**)$ pues es más facil de manejar algebraicamente.

Propiedades de los Estimadores

1. Sea $\widehat{\underline{\theta}} = t\left(\underline{x}\right)$ un estimador de $\underline{\theta}$. Se dice que $\underline{\theta}$ es un *Estimador Insesgado* de $\widehat{\underline{\theta}}$ ssi $\mathbb{E}\left(\widehat{\underline{\theta}}\right) = \underline{\theta}$.

2. Sea $\widehat{\underline{\theta}} = t\left(\underline{x}\right)$ un estimador de $\underline{\theta}$. Se dice que $\underline{\theta}$ es un *Estimador Asintoticamente Insesgado* de $\widehat{\underline{\theta}}$ ssi $\lim\limits_{n\to\infty} \mathbb{E}\left(\widehat{\underline{\theta}}\right) = \underline{\theta}$.

3. Sea $\left\{\widehat{\underline{\theta}}_n\right\}_{n\in\mathbb{N}}$ una sucesión de estimadores de $\underline{\theta}$. Esta sucesión se dice *Consistente en Error Cuadratico Medio* ssi

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\widehat{\underline{\theta}}_n - \underline{\theta}\right)^2\right] = 0$$

Proposición:

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\widehat{\underline{\theta}}_n - \underline{\theta}\right)^2\right] = 0 \iff \lim_{n\to\infty} \mathbb{V}\left(\widehat{\underline{\theta}}_n\right) = 0 \ \wedge \ \lim_{n\to\infty} \mathbb{E}\left(\widehat{\underline{\theta}}_n\right) = 0$$

4. Sea $\left\{\widehat{\underline{\theta}}_n\right\}_{n\in\mathbb{N}}$ una sucesión de estimadores de $\underline{\theta}$. Esta sucesión se dice *Consistente Simple* ssi

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\widehat{\underline{\theta}}_n - \underline{\theta}\right| > \epsilon\right) = 0$$

5. Sea $\widehat{\underline{\theta}} = t\left(\underline{x}\right)$ un estimador de $\underline{\theta}$. Se dice que $\widehat{\underline{\theta}}$ es un *Estimador Insesgado de Varianza Minima* ssi

   i) $\mathbb{E}\left(\widehat{\underline{\theta}}\right) = \underline{\theta}$

   ii) $\mathbb{V}\left(\widehat{\underline{\theta}}\right) \leq \mathbb{V}\left(\widehat{\underline{\theta}}^*\right)$, $\forall\, \widehat{\underline{\theta}}^*$ estimador de $\underline{\theta}$.

Definición: Sean $\widehat{\underline{\theta}}$ y $\widetilde{\underline{\theta}}$ dos estimadores de $\underline{\theta}$. Se define como *Eficiencia Relativa* de $\widehat{\underline{\theta}}$ c/r a $\widetilde{\underline{\theta}}$ a la cantidad

$$e\left(\widehat{\underline{\theta}}\ ,\ \widetilde{\underline{\theta}}\right) = \frac{\mathbb{E}\left[\left(\widehat{\underline{\theta}} - \underline{\theta}\right)^2\right]}{\mathbb{E}\left[\left(\widetilde{\underline{\theta}} - \underline{\theta}\right)^2\right]}$$

obs: si $e\left(\widehat{\underline{\theta}}\ ,\ \widetilde{\underline{\theta}}\right) < 1 \implies \widehat{\underline{\theta}}$ es mas *eficiente* que $\widetilde{\underline{\theta}}$.

Definición: Se llama *Cantidad de Información de Fisher* dada por $\underline{x}$ sobre el parametro $\underline{\theta}$ a la cantidad

$$I_n\left(\underline{\theta}\right) = \mathbb{E}\left[\left(\frac{\partial \ln\left(L_{\underline{x}}\left(\underline{\theta}\right)\right)}{\partial \underline{\theta}}\right)^2\right]$$

Proposición:

$$I_n\left(\underline{\theta}\right) = \mathbb{V}\left(\frac{\partial \ln\left(L_{\underline{x}}\left(\underline{\theta}\right)\right)}{\partial \underline{\theta}}\right) \ \vee \ I_n\left(\underline{\theta}\right) = -\mathbb{E}\left[\frac{\partial^2 \ln\left(L_{\underline{x}}\left(\underline{\theta}\right)\right)}{\partial \underline{\theta}^2}\right]$$

Teorema: Si los valores muestrales son independientes y $I\left(\underline{\theta}\right)$ es la cantidad de información de Fisher dada para cada x_i sobre el parametro $\underline{\theta}$, entonces

$$I_n\left(\underline{\theta}\right) = nI\left(\underline{\theta}\right)\ ,\ \text{Donde } I_n\left(\underline{\theta}\right) = \mathbb{E}\left[\left(\frac{\partial \ln\left(L_{\underline{x}}\left(\underline{\theta}\right)\right)}{\partial \underline{\theta}}\right)^2\right]$$

Teorema(Cota de Cramer-Rao): Si el dominio de χ de $\underline{x}$ no depende del parametro $\underline{\theta}$, para todo estimador $\widehat{\underline{\theta}}$ de $\underline{\theta}$ se tiene que:

$$\mathbb{V}\left(\widehat{\underline{\theta}}\right) \geq \frac{\left(1 + b'\left(\widehat{\underline{\theta}}\right)\right)^2}{I_n\left(\underline{\theta}\right)}$$

Donde $b\left(\widehat{\underline{\theta}}\right) = \mathbb{E}\left(\widehat{\underline{\theta}}\right) - \underline{\theta}$ es el sesgo de $\widehat{\underline{\theta}}$

Obs: Si $\widehat{\underline{\theta}}$ alcanza la cota de Cramer-Rao, luego se dice que $\widehat{\underline{\theta}}$ es *Eficiente*.

References

[1] H. Allende. *Probabilidades y estadística*. Universidad Técnica Federico Santa María, 1982.