Inferencia Estadística

Marisol García Peña

Departamento de Matemáticas
Pontificia Universidad Javeriana

Bogotá, 2022

Estimación Puntual

Estimación Puntual

Estimador Estadística derivada de una muestra para inferir el valor de un parámetro poblacional.

Estimación Valor del estimador en una muestra particular.

Notación Parámetro poblacional $\Longrightarrow \mu, \sigma, \pi$

Estimador Muestral $\Longrightarrow \hat{\mu}, \hat{\sigma}, \hat{\pi}$ o $\overline{x}, s, p$

Estimación puntual (cont.)

- Objetivo. Estimar un parámetro desconocido θ o el valor de una función $\tau(\theta)$ de un parámetro desconocido de una población dada.
- Sea $X_1, X_2, \ldots, X_n$ una m.a.s de una población. Sea θ un parámetro de interés de esa población. Un estimador puntual (estimador) de θ es cualquier función de la m.a.s.: $T(X_1, X_2, \ldots, X_n)$.
- Estimador puntual $\Longrightarrow$ estadístico cuya función es estimar (aproximar) el valor de un parámetro. Un estimador es una v.a.
- Una realización particular del estimador, es decir, el valor del estimador en una muestra particular, $T(X_1, X_2, \ldots, X_n)$, se llama estimación.

Estimación puntual (cont.)

Media muestral	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$	x_i, i-ésimo valor de los datos n tamaño de muestra.
Proporción muestral	$p = \frac{X}{n}$	X "éxitos" en la muestra, n tamaño de muestra.
Desvest muestral	$s = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}}$	X_i i-ésimo valor de los datos, n tamaño de muestra.

Estimación puntual (cont.)

- Estimación por el método de los momentos.
- Estimación por máxima verosimilitud.
- Otros métodos.

Estimación por método de los momentos

- k-ésimo momento de v.a. $\Longrightarrow \mu'_k = E[X^k]$.
- k-ésimo momento muestral $\Longrightarrow M'_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$.
- Momentos muestrales $\Longrightarrow$ Momentos poblacionales.
- Resolver el sistema de ecuaciones $\mu'_k = M'_k$, para $k = 1, 2, \ldots, t$, donde t es el número de parámetros por estimar.

Estimación puntual (cont.)

- Los estimadores no son únicos.
- También se pueden usar los momentos alrededor de la media.
- Los estimadores obtenidos por cualquiera de los momentos usados pueden no ser los mismos.

Ejemplo

Una m.a. de n observaciones, $X_1, \ldots, X_n$ se selecciona de una población en la que X_i para $i = 1, \ldots, n$, posee una f.d.p. uniforme en el intervalo $(0, \theta)$. Estimar θ.

- $\mu_1' = \mu = \frac{\theta}{2}$.
- P.M.M $\Longrightarrow M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$.
- Igualando $\Longrightarrow \mu_1' = \frac{\theta}{2} = \overline{X}$.
- $\widehat{\theta} = 2\overline{X}$.

Ejemplo

Sea $X_1, X_2, \ldots, X_n$ una m.a. de una población normal $N(\mu, \sigma^2)$, encuentre el estimador de los parámetros.

Ejemplo

Sea $X_1, X_2, \ldots, X_n$ una m.a. de una población con distribución uniforme en el intervalo $(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)$. Encuentre los estimadores de momentos para μ y para σ.

Estimación por máxima verosimilitud

- Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población X con función de probabilidad P_θ (o con función de densidad f_θ). Para cada muestra particular $(x_1, \ldots, x_n)$, la **función de verosimilitud** se define como la función de probabilidad (o de densidad) conjunta de $(X_1, \ldots, X_n)$ evaluada en $(x_1, \ldots, x_n)$.

$L(\theta) = L(x_1, \ldots, x_n; \theta) = P_\theta(X_1 = x_1, \ldots, X_n = x_n)$ si X es discreta.
$L(\theta) = L(x_1, \ldots, x_n; \theta) = f_\theta(x_1, \ldots, x_n)$ si X es continua.

Estimación puntual (cont.)

El caso más frecuente, es aquel en el que las v.a. son i.i.d.

- X v.a. discreta $\sim P(\cdot;\theta)$. $X_1,\ldots,X_n$ m.a.s de X

$$\begin{aligned} L(x_1,\ldots,x_n;\theta) &= P(X_1 = x_1,\ldots,X_n = x_n;\theta) \\ &= P(X_1 = x_1;\theta)\cdot\ldots\cdot P(X_n = x_n;\theta) \\ &= \prod_{i=1}^{n} P(X = x_i;\theta) \end{aligned}$$

- X v.a. continua $\sim f(\cdot;\theta)$. $X_1,\ldots,X_n$ m.a.s de X

$$\begin{aligned} L(x_1,\ldots,x_n;\theta) &= f(x_1,\ldots,x_n;\theta) \\ &= f(x_1;\theta)\cdot\ldots\cdot f(x_n;\theta) \\ &= \prod_{i=1}^{n} f(x_i;\theta) \end{aligned}$$

- Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población X con función de verosimilitud $L(\theta)$. Para cada muestra particular $(x_1, \ldots, x_n)$, la **estimación de máxima verosimilitud** de θ es el valor $\hat{\theta}_{MV}$ que maximiza la verosimilitud.

$$L(x_1, \ldots, x_n; \hat{\theta}_{MV}) = \max_{\theta} L(x_1, \ldots, x_n; \theta)$$

- El **estimador de máxima verosimilitud**, $\hat{\theta}_{MV}(X_1, \ldots, X_n)$, es aquel que evaluado en cada muestra particular nos da la estimación de máxima verosimilitud $(\hat{\theta}_{MV}(x_1, \ldots, x_n))$.

Sea $\theta = (\theta_1, \ldots, \theta_k)$. Procedimiento para obtener el E.M.V. de θ_j dada una muestra particular $(x_1, \ldots, x_n)$

- Escribir la verosimilitud: $L(\theta) = L(x_1, \ldots, x_n; \theta)$.
- Escribir el logaritmo de la verosimilitud: $\ell(\theta) = \ln L(\theta)$, $\ell(\theta)$ se llama **función soporte** o **log-verosimilitud**.
- Obtener el θ_j tal que

$$\frac{\partial}{\partial \theta_j}\ell(\theta) = 0 \text{ denotarlo por } \hat{\theta}_j.$$

- Comprobar que realmente es un máximo, es decir, que

$$\frac{\partial^2}{\partial \theta_j^2}\ell(\theta)\Big|_{\theta_j = \hat{\theta}_j} < 0$$

Estimación puntual (cont.)

Ejemplo

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma)$. Obtener el E.M.V. de μ.

- Función de verosimilitud

$$L(\mu) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

- Función soporte

$$\ell(\mu) = \ln L(\mu) = -n\ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

- Obtener el valor de μ tal que $\frac{\partial}{\partial\mu}\ell(\mu) = 0$.

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

$$\frac{\partial}{\partial\mu}\ell(\mu) = 0 \Leftrightarrow \frac{1}{\sigma^2}\left(-n\mu + \sum_{i=1}^{n} x_i\right) = 0 \Leftrightarrow -n\mu + \sum_{i=1}^{n} x_i = 0$$

$$\Leftrightarrow n\mu = \sum_{i=1}^{n} x_i \Leftrightarrow \mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Por tanto $\hat{\mu} = \bar{x}$.

Estimación puntual (cont.)

- Comprobar que realmente es un máximo, es decir, que $\frac{\partial^2}{\partial \mu^2}\ell(\mu)|_{\mu=\hat{\mu}} < 0$.
$$\frac{\partial^2}{\sigma \mu^2}\ell(\mu) = \frac{-n}{\sigma^2} < 0 \quad \forall_{\mu} \Rightarrow \frac{\partial^2}{\partial \mu^2}\ell(\mu)|_{\mu=\hat{\mu}} < 0$$

Por tanto la estimación máximo verosímil para una muestra dada es $\hat{\mu}_{MV} = \bar{x}$, y el estimador máximo verosímil es $\hat{\mu}_{MV} = \overline{X}$.

Ejemplo

Suponga que una muestra aleatoria de tamaño n es seleccionada de una distribución Bernoulli con $P[X = x] = p^x q^{1-x} I_{\{0,1\}}(x)$, $0 \leq p \leq 1$ y $q = 1 - p$. Encuentre el EMV de p.

Propiedad de invarianza de los estimadores de máxima verosimilitud

Sea $T(X_1, X_2, \ldots, X_n)$ el estimador de máxima verosimilitud de θ en la densidad $f(x, \theta)$, donde θ se asume unidimensional. Si $\tau(\bullet)$ es una función con inversa de valor único, entonces el estimador de máxima verosimilitud de $\tau(\theta)$ es $\tau(T)$.

Ejemplo

En la distribución normal con μ_0 conocido, el estimador de máxima verosimilitud de σ^2 es $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2$. Por la propiedad de invarianza, el estimador de máxima verosimilitud de σ es $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2}$.

Similarmente, el estimador de máxima verosimilitud de $\log\sigma^2$ es $\log\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2\right]$

Extensión propiedad de invarianza

- θ k-dimensional.
- Se elimina el supuesto de que $\tau(\bullet)$ tiene una inversa de valor único.

Sea $\widehat{\boldsymbol{\Theta}} = (\widehat{\Theta}_1, \ldots, \widehat{\Theta}_k)$, donde $\widehat{\Theta}_j = T_j(X_1, \ldots, X_n)$, un estimador de máxima verosimilitud de $\theta = (\theta_1, \ldots, \theta_k)$ en la densidad $f(\bullet, \theta_1, \ldots, \theta_k)$. Si $\tau(\theta) = (\tau_1(\theta), \ldots, \tau_r(\theta))$ para $1 \leq r \leq k$ es una transformación del espacio del parámetro, entonces el estimador de máxima verosimilitud de $\tau(\theta) = (\tau_1(\theta), \ldots, \tau_r(\theta))$ es $\tau(\widehat{\boldsymbol{\Theta}}) = (\tau_1(\widehat{\boldsymbol{\Theta}}), \ldots, \tau_r(\widehat{\boldsymbol{\Theta}}))$.

Ejemplo

Se obtiene los estimadores de máxima verosimilitud para μ y σ^2, pero para $\mu + \sigma^2$ que no es una función uno a uno de μ y σ^2 y con el principio de invarianza no sería posible encontrar su estimador de máxima verosimilitud. Con la extensión del principio, se obtiene $\overline{x} + \frac{1}{n}\sum(x_i - \overline{x})^2$

Estimación puntual (cont.)

La aproximación matemática para maximizar una función es derivarla, igualarla a 0 y resolverla.

Cuando no es posible resolver analíticamente $\Longrightarrow$ Software

- Usar software para optimizar la función de log-verosimilitud $\Longrightarrow$ `optimize` in R.
- Derivar la función de log-verosimilitud y usar software para resolverla $\Longrightarrow$ `uniroot` in R.

Obtención de los estimadores de máxima verosimilitud por métodos iterativos

Una función $f(x)$ puede ser expandida en serie de Taylor alrededor de un punto x_0, es decir,

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)(x - x_0)^2}{2!} + \cdots + \frac{f^n(x_0)(x - x_0)^n}{n!}$$

La idea principal es expandir $\frac{\partial \ell(\theta)}{\partial \theta}\Big|_{\theta=\widehat{\theta}}$ en serie de Taylor alrededor de una estimación inicial de $\theta^{(0)}$.

$$\begin{aligned}
\ell(\theta) &= \log(L(\theta, x)) \\
\frac{\partial \ell(\theta)}{\partial \theta}\Big|_{\theta=\widehat{\theta}} &= \frac{\partial \ell(\theta)}{\partial \theta}\Big|_{\theta=\theta^{(0)}} + \frac{\partial^2 \ell(\theta)}{\partial \theta^2}\Big|_{\theta=\theta^{(0)}} (\widehat{\theta} - \theta^{(0)})
\end{aligned}$$

De esta manera se tiene que:

$$-\frac{\partial \ell(\theta)}{\partial \theta}\Big|_{\theta=\theta^{(0)}} = \frac{\partial^2 \ell(\theta)}{\partial \theta^2}\Big|_{\theta=\theta^{(0)}} (\widehat{\theta} - \theta^{(0)})$$

$$-\frac{\partial \ell(\theta)}{\partial \theta}\Big|_{\theta=\theta^{(0)}} \left[\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\Big|_{\theta=\theta^{(0)}}\right]^{-1} = \widehat{\theta} - \theta^{(0)}$$

$$\widehat{\theta} = \theta^{(0)} - \frac{\partial \ell(\theta)}{\partial \theta}\Big|_{\theta=\theta^{(0)}} \left[\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\Big|_{\theta=\theta^{(0)}}\right]^{-1} \maltese$$

Método de Newton Raphson (algoritmo)

Consiste en resolver iterativamente la ecuación de máxima verosimilitud a través de la expresión (✠) usando la estimación del paso anterior.

1. Se define un valor inicial $\theta^{(0)}$.
2. Se calcula $\widehat{\theta}^{(1)} = \theta^{(0)} - \theta^{(0)} - \left.\frac{\partial \ell(\theta)}{\partial \theta}\right|_{\theta=\theta^{(0)}} \left[\left.\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right|_{\theta=\theta^{(0)}} \right]^{-1}$.
3. Se calcula $\widehat{\theta}^{(2)} = \theta^{(1)} - \left.\frac{\partial \ell(\theta)}{\partial \theta}\right|_{\theta=\theta^{(1)}} \left[\left.\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right|_{\theta=\theta^{(1)}} \right]^{-1}$.

y así sucesivamente hasta obtener convergencia.
Un critério de convergencia muy usado es detener el algoritmo cuando $\left| \widehat{\theta}^{(n)} - \widehat{\theta}^{(n-1)} \right| < \varepsilon = 0.00001$, en este caso $\widehat{\theta}^{(n)}$ será el EMV.

Estimación puntual (cont.)

Método de Score

Es semejante al método de Newton Raphson, la diferencia está en sustituir

$$\left.\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right|_{\theta=\theta^{(0)}} \quad \text{por} \quad E\left[\left.\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right|_{\theta=\theta^{(0)}}\right]$$

Generalizando para un vector de parámetros

Considerando un vector de parámetros $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$. Definiendo la función score

$$\boldsymbol{u}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix}_{p \times 1}$$

Matriz de información observada $(-I(\boldsymbol{\theta}))$

$$-I(\boldsymbol{\theta}) = \frac{\partial^2 \ell(\theta)}{\partial \boldsymbol{\theta} \partial \theta} = \begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_1^2} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_p \partial \theta_1} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_p \partial \theta_2} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_p^2} \end{bmatrix}_{p \times p}$$

Ahora la idea es expandir $\boldsymbol{u}(\widehat{\boldsymbol{\theta}})$ en serie multivariada de Taylor hasta el primer orden alrededor del punto inicial $\theta^{(0)}$.

$$\begin{aligned} \mathbf{0} = \boldsymbol{u}(\widehat{\theta}) &= \boldsymbol{u}(\boldsymbol{\theta}^{(0)}) + \boldsymbol{u}'(\boldsymbol{\theta}^{(0)})(\widehat{\theta} - \theta^{(0)}) \\ &= \boldsymbol{u}(\boldsymbol{\theta}^{(0)}) - I(\boldsymbol{\theta}^{(0)})(\widehat{\theta} - \theta^{(0)}) \end{aligned}$$

Asi, se tiene

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(0)} + I^{-1}(\boldsymbol{\theta}^{(0)})\boldsymbol{u}(\theta^{(0)})$$

Como ya se mencionó, el método de Newton Raphson consiste en obtener una secuencia $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(m)}$, tal que

$$\widehat{\boldsymbol{\theta}}^{(m)}_{p\times 1} = \boldsymbol{\theta}^{(m-1)}_{p\times 1} + I^{-1}_{p\times p}(\boldsymbol{\theta}^{(m-1)})\boldsymbol{u}_{p\times 1}(\boldsymbol{\theta}^{(m-1)})$$

El EMV $\widehat{\theta}$ es el valor $\widehat{\theta}^{(m)}$ tal que $\|\widehat{\boldsymbol{\theta}}^{(m)} - \widehat{\boldsymbol{\theta}}^{(m-1)}\| \leq 0.0001$.

Observaciones

- El método supone un valor inicial $\theta^{(0)}$.
- Caso $\widehat{\theta}$ este en la frontera del espacio del parámetro $\Longrightarrow$ cuidado con el método. Se debe reparametrizar.
- Cuidado con la inversa de la matriz de información observada.
- Substituyendo $\boldsymbol{I}(\boldsymbol{\theta})$ por $E[\boldsymbol{I}(\boldsymbol{\theta})]$ se obtiene el método de escore, que usualmente es más rápido.

Algoritmo EM (Expectation-Maximization)

El algoritmo EM consiste en 2 etapas

- **Etapa E** Calcula el valor esperado, es decir, $E[\ell(\theta)]$.
- **Etapa M** Determina $\widehat{\theta}^{(m+1)}$ maximizando $E[\ell(\theta)]$.

El algoritmo se detiene, por ejemplo cuando

$$\frac{\|\boldsymbol{\theta}^{(m)} - \boldsymbol{\theta}^{(m-1)}\|}{\|\boldsymbol{\theta}^{(m)}\|} < \varepsilon = 0.0001$$

Estimación puntual (cont.)

Ejemplo

El tiempo de vida de un neumático de vehículo está medido en distancia (millas) más que en tiempo. Suponga que una compañía produce tres versiones de neumáticos:

- Neumáticos estándar, cuyo tiempo de vida, X_s tiene distribución exponencial con media $1/\lambda$
- Neumáticos económicos cuyo tiempo de vida, X_e tiene distribución exponencial con media $0.77/\lambda$
- Neumáticos premium X_p con distribución exponencial con media $1.25/\lambda$.

Suponga que un neumático de cada tipo es seleccionado de manera aleatoria e independientemente y probado para encontrar el tiempo de vida de cada uno, $x_s = 28, x_e = 25, x_p = 31$ en miles de millas. Encuentre el estimador de máxima verosimilitud de λ.

Se tiene: $X_s \sim exp(\lambda), X_e \sim exp(1.3\lambda)$ y $X_p \sim exp(0.8\lambda)$.

$$\begin{aligned} L(\lambda) &= f_s(X_s;\lambda) f_e(X_e;\lambda) f_p(X_p;\lambda) \\ &= \lambda e^{-\lambda X_s}(1.3\lambda e^{-1.3\lambda X_e})(0.8\lambda e^{-0.8\lambda X_p}) \end{aligned}$$

$$\begin{aligned} \ell(\lambda) = ln(L(\lambda)) &= \ln(1.04) + 3\ln(\lambda) - (X_s + 1.3X_e + 0.8X_p)\lambda \\ \frac{\partial \ell(\lambda)}{\partial \lambda} &= \frac{3}{\lambda} - (X_s + 1.3X_e + 0.8X_p) \end{aligned}$$

Igualando la derivada a 0 y resolviendo, $\widehat{\lambda} = \frac{3}{X_s + 1.3X_e + 0.8X_p}$. Para los datos que se tienen, $\widehat{\lambda} = 0.0395$.

Estimación puntual (cont.)

- Para neumáticos estándar, el tiempo de vida promedio estimado es $\widehat{E}[X_s] = \frac{1}{\widehat{\lambda}} = 25.3$ miles de millas.
- Para neumáticos económicos, el tiempo de vida promedio estimado es $\widehat{E}[X_e] = \frac{1}{1.3\widehat{\lambda}} = 19.5$ miles de millas.
- Para neumáticos premium, el tiempo de vida promedio estimado es $\widehat{E}[X_p] = \frac{1}{0.8\widehat{\lambda}} = 31.6$ miles de millas.