

MAT-206: Sesión 6, Verosimilitud y propiedades

Felipe Osorio

`fosorios.mat.utfsm.cl`

Departamento de Matemática, UTFSM



Definición 1 (Función de verosimilitud):

Para una observación x fijada de un vector aleatorio \mathbf{X} con densidad $f(\cdot; \theta)$. La función de verosimilitud

$$L(\cdot; x) : \Theta \rightarrow \mathbb{R}_+,$$

es definida como

$$L(\theta; x) = f(x; \theta), \quad \theta \in \Theta.$$

Observación:

La verosimilitud corresponde a la **densidad conjunta** de los datos que se desea analizar.



Ejemplo:

Sea X_1, \dots, X_n variables aleatorias IID con distribución $N(\theta, 1)$. Entonces,

$$\begin{aligned} L(\theta; \mathbf{x}) &= \prod_{i=1}^n (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2} (x_i - \theta)^2 \right\} = (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2 \right\}. \end{aligned}$$

Ahora,

$$\sum_{i=1}^n (x_i - \theta)^2 = \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \theta)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \theta)^2,$$

pues $\sum_{i=1}^n (x_i - \bar{x})(\bar{x} - \theta) = 0$. De este modo,

$$\begin{aligned} L(\theta; \mathbf{x}) &\propto \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \theta)^2 \right] \right\} \\ &\propto \exp \left\{ -\frac{n}{2} (\bar{x} - \theta)^2 \right\}. \end{aligned}$$

Es decir, $L(\theta; \mathbf{x})$ es proporcional a una densidad $N_1(\bar{x}, 1/n)$.



Observación:

Es conveniente usar la función de log-verosimilitud dada por

$$\ell(\boldsymbol{\theta}; \mathbf{x}) = \log L(\boldsymbol{\theta}; \mathbf{x}) = \log f(\mathbf{x}; \boldsymbol{\theta}).$$

Ejemplo:

Sea $X \sim N(\theta, 1)$. Entonces,

$$\ell(\theta; X) = -\frac{1}{2} \log 2\pi - \frac{1}{2}(X - \theta)^2,$$

de este modo, la variable aleatoria

$$-2\ell(\theta; X) - \log 2\pi = (X - \theta)^2 \sim \chi^2(1).$$



Función de verosimilitud

Considere $\mathbf{x}_1, \dots, \mathbf{x}_n$, vectores aleatorios IID. Entonces la verosimilitud de los datos combinados es:

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^n f(\mathbf{x}_i; \boldsymbol{\theta}), \quad \ell_n(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(\mathbf{x}_i; \boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}).$$

Ejemplo:

Sea $\mathbf{y}_1, \dots, \mathbf{y}_n$ vectores aleatorios independientes $N_{m_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i(\boldsymbol{\phi}))$ con \mathbf{X}_i matriz $m_i \times p$. Así,

$$\ell_n(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(\mathbf{y}_i; \boldsymbol{\theta}) = -\frac{1}{2} \sum_{i=1}^n \log |2\pi \boldsymbol{\Sigma}_i(\boldsymbol{\phi})| - \frac{1}{2} \sum_{i=1}^n D_i^2(\boldsymbol{\theta}),$$

con

$$D_i^2(\boldsymbol{\theta}) = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\phi})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}),$$

para $i = 1, \dots, n$, con $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top)^\top$.



Lema 1 (Principio de verosimilitud):

Sea θ_0 el parámetro subyacente (o verdadero). Entonces,

$$\mathbb{E}_{\theta_0} \{\ell(\theta; \mathbf{X})\} \leq \mathbb{E}_{\theta_0} \{\ell(\theta_0; \mathbf{X})\}, \quad \forall \theta_0, \theta \in \Theta.$$

Demostración:

En efecto,¹

$$\begin{aligned} \mathbb{E}_{\theta_0} \{\log f(\mathbf{x}; \theta_0)\} - \mathbb{E}_{\theta_0} \{\log f(\mathbf{x}; \theta)\} &= -\mathbb{E}_{\theta_0} \left\{ \log \frac{f(\mathbf{x}; \theta)}{f(\mathbf{x}; \theta_0)} \right\} \\ &\geq -\log \mathbb{E}_{\theta_0} \left\{ \frac{f(\mathbf{x}; \theta)}{f(\mathbf{x}; \theta_0)} \right\} = \log \int \frac{f(\mathbf{x}; \theta)}{f(\mathbf{x}; \theta_0)} f(\mathbf{x}; \theta_0) d\mathbf{x} \\ &= \log \int f(\mathbf{x}; \theta) d\mathbf{x} = \log 1 = 0. \end{aligned}$$

¹Recuerde que, para h convexa $\mathbb{E}(h(Y)) \geq h(\mathbb{E}(Y))$.

Supuesto A1:

Las distribuciones $\{P_\theta : \theta \in \Theta\}$ tienen **soporte común**, de modo que el conjunto

$$A = \{\mathbf{x} : f(\mathbf{x}; \theta) \geq 0\},$$

no depende de θ .

Contraejemplos:

Considere $X \sim U(0, \theta)$, con $\theta \in (0, \infty)$ cuya densidad es

$$f(x; \theta) = \frac{1}{\theta} I_{[0, \theta]}(x).$$

También la familia de distribuciones exponencial con dos parámetros $Y \sim \text{Exp}(a, b)$,

$$f(y; a, b) = \frac{1}{b} \exp\left(-\frac{(x-a)}{b}\right) I_{[a, \infty)}(y), \quad a, b > 0.$$



Supuesto A2:

El espacio paramétrico $\Theta \subset \mathbb{R}^p$ es un conjunto abierto.

Supuesto A3:

Para todo $x \in A$ la función de log-verosimilitud es 3-veces continuamente diferenciable con respecto a $\theta = (\theta_1, \dots, \theta_p)^\top$.

Definición 2 (Función score):

Suponga las condiciones A1 a A3 para todo $x \in A$, se define la **función score** como el vector de derivadas parciales de la log-verosimilitud

$$U(\theta; x) = \frac{\partial \ell(\theta; x)}{\partial \theta} = \left(\frac{\partial \ell(\theta; x)}{\partial \theta_1}, \dots, \frac{\partial \ell(\theta; x)}{\partial \theta_p} \right)^\top.$$



Supuesto A4:

Suponga que existen funciones integrables $F_1(x)$, $F_2(x)$ y $H(x)$ tal que

$$\int H(\mathbf{x})f(\mathbf{x};\boldsymbol{\theta})\,d\mathbf{x} < M,$$

para $M \in \mathbb{R}$ un valor apropiado y que se satisface

$$\begin{aligned} \left| \frac{\partial \log f(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i} \right| &< F_1(\mathbf{x}), & \left| \frac{\partial^2 \log f(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right| &< F_2(\mathbf{x}), \\ \left| \frac{\partial^3 \log f(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| &< H(\mathbf{x}), & i, j, k &= 1, \dots, p. \end{aligned}$$

Esta condición implica que podemos intercambiar las operaciones de integración y diferenciación. Por ejemplo,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_A f(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x} = \int_A \frac{\partial}{\partial \boldsymbol{\theta}} f(\mathbf{x}; \boldsymbol{\theta}) \, d\mathbf{x}.$$



Resultado 1:

Bajo las condiciones A1 a A4, tenemos

$$E_{\theta}\{U(\theta; \mathbf{X})\} = \mathbf{0}, \quad \forall \theta \in \Theta$$

Demostración:

Tenemos

$$\begin{aligned} E_{\theta}\{U(\theta; \mathbf{X})\} &= \int_A \frac{\partial}{\partial \theta} \log f(\mathbf{x}; \theta) f(\mathbf{x}; \theta) d\mathbf{x} \\ &= \int_A \frac{1}{f(\mathbf{x}; \theta)} \frac{\partial}{\partial \theta} f(\mathbf{x}; \theta) f(\mathbf{x}; \theta) d\mathbf{x} \\ &= \int_A \frac{\partial}{\partial \theta} f(\mathbf{x}; \theta) d\mathbf{x} = \frac{\partial}{\partial \theta} \int_A f(\mathbf{x}; \theta) d\mathbf{x} \\ &= \mathbf{0} \end{aligned}$$



Ejemplo:

Considere $X \sim N(\theta, 1)$. Entonces,

$$U(\theta; X) = \frac{\partial}{\partial \theta} \left[-\frac{1}{2} \log 2\pi - \frac{1}{2}(X - \theta)^2 \right] = X - \theta.$$

De este modo, $U(\theta; X) \sim N(0, 1)$. Así, es directo

$$E\{U(\theta; X)\} = E(X - \theta) = 0.$$



Definición 3 (Matriz de información de Fisher):

Suponga las condiciones A1 a A3. Entonces la **matriz de información de Fisher** se define como:

$$\mathcal{F}(\boldsymbol{\theta}) = \text{Cov}_{\boldsymbol{\theta}}\{U(\boldsymbol{\theta}; \mathbf{X})\} = E_{\boldsymbol{\theta}}\{U(\boldsymbol{\theta}; \mathbf{X})U^{\top}(\boldsymbol{\theta}; \mathbf{X})\}.$$

Es decir, $\mathcal{F}(\boldsymbol{\theta})$ tiene elementos

$$\mathcal{F}_{ij}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \left\{ \frac{\partial}{\partial \theta_i} \ell(\boldsymbol{\theta}; \mathbf{X}) \frac{\partial}{\partial \theta_j} \ell(\boldsymbol{\theta}; \mathbf{X}) \right\}.$$

Observación:

En ocasiones escribimos $\mathcal{F}_{\mathbf{X}}(\boldsymbol{\theta})$ pero la información **no** es aleatoria.



Ejemplo:

Sean X_1, \dots, X_n variables aleatorias $N(\mu, \sigma^2)$ con σ^2 conocido. Entonces,

$$L(\mu) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \right\},$$

así

$$\ell(\mu) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 + c,$$

con c una constante. Además,

$$U(\mu; \mathbf{X}) = \dot{\ell}(\mu) = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu).$$

De este modo,²

$$\mathcal{F}_n(\mu) = \text{var}\{U(\mu; \mathbf{X})\} = \frac{1}{\sigma^4} \sum_{i=1}^n \text{var}(X_i - \mu) = \frac{n}{\sigma^2}.$$

²Hay más información, conforme $\sigma^2 \rightarrow 0$ o $n \rightarrow \infty$.

Resultado 2:

Suponga las condiciones A1 a A4, Entonces

$$\mathcal{F}(\theta) = E_{\theta}\{-\ddot{\ell}(\theta; \mathbf{X})\} = E_{\theta}\left\{-\frac{\partial^2 \ell(\theta; \mathbf{X})}{\partial \theta \partial \theta^{\top}}\right\}.$$

Demostración:

Revisar [página 26](#) de las notas de clase.³

Observación:

Este resultado permite obtener la matriz de información de Fisher de dos maneras equivalentes en [modelos regulares](#) (esto es, bajo los Supuestos A1 a A4). Es decir,

$$\mathcal{F}(\theta) = E_{\theta}\{U(\theta; \mathbf{X})U^{\top}(\theta; \mathbf{X})\} = E_{\theta}\left\{-\frac{\partial^2 \ell(\theta; \mathbf{X})}{\partial \theta \partial \theta^{\top}}\right\}.$$

³Disponible en: <http://fosorios.mat.utfsm.cl/files/lectures/Inferencia.pdf>



Definición 4 (Información observada):

La matriz

$$\mathbf{J}(\boldsymbol{\theta}; \mathbf{X}) = -\frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{X})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},$$

se denomina **información observada**.

Observación:

En efecto,

$$\mathcal{F}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}\{\mathbf{J}(\boldsymbol{\theta}; \mathbf{X})\},$$

que también es llamada **matriz de información esperada**.



Ejemplo:

Considere $X \sim \text{Bin}(n, \theta)$ con $\theta \in (0, 1)$. De este modo,

$$\ell(\theta; x) = \log \binom{n}{x} + (n - x) \log(1 - \theta) + x \log \theta,$$

así,

$$U(\theta; x) = -\frac{n - x}{1 - \theta} + \frac{x}{\theta},$$

obteniendo la derivada de $U(\theta; x)$,

$$U'(\theta; x) = -\frac{n - x}{(1 - \theta)^2} - \frac{x}{\theta^2}.$$

Además, como $E(X) = n\theta$, sigue que

$$\begin{aligned} \mathcal{F}(\theta) &= E_{\theta}\{-U'(\theta; X)\} = \frac{n - E(X)}{(1 - \theta)^2} + \frac{E(X)}{\theta^2} = \frac{n - n\theta}{(1 - \theta)^2} + \frac{n\theta}{\theta} \\ &= n\left(\frac{1}{1 - \theta} + \frac{1}{\theta}\right) = \frac{n}{\theta(1 - \theta)}. \end{aligned}$$



Ejemplo:

Sea $X \sim \text{FE 1-paramétrica}$. En este caso, tenemos

$$\ell(\eta; x) = \log f(x; \eta) = \eta T(x) - b(\eta) + \log h(x),$$

De ahí que,

$$U(\eta; x) = T(x) - b'(\eta),$$

Usando el **Resultado 2** visto en la Sesión 3⁴,

$$\mathbb{E}\{U(\eta; X)\} = \mathbb{E}\{T(X)\} - b'(\eta) = 0.$$

Además,

$$U'(\eta; x) = -\frac{d^2 b(\eta)}{d\eta^2} = -b''(\eta),$$

obtenemos la información de Fisher:

$$\mathcal{F}(\eta) = \mathbb{E}\{-U'(\eta)\} = b''(\eta) = \text{var}(T).$$

⁴URL: http://fosorios.mat.utfsm.cl/files/notes/MAT206_slides-03.pdf



Resultado 3:

Sean \mathbf{X} e \mathbf{Y} vectores aleatorios independientes con informaciones de Fisher $\mathcal{F}_{\mathbf{X}}(\boldsymbol{\theta})$ y $\mathcal{F}_{\mathbf{Y}}(\boldsymbol{\theta})$, respectivamente. La información de Fisher contenida en $\mathbf{Z} = (\mathbf{X}^\top, \mathbf{Y}^\top)^\top$ es dada por

$$\mathcal{F}_{\mathbf{Z}}(\boldsymbol{\theta}) = \mathcal{F}_{\mathbf{X}}(\boldsymbol{\theta}) + \mathcal{F}_{\mathbf{Y}}(\boldsymbol{\theta}).$$

Corolario:

Sea $\mathbf{X} = (X_1, \dots, X_n)^\top$ copias IID de una variable aleatoria X con información de Fisher (común) $\mathcal{F}_1(\boldsymbol{\theta})$. Entonces, la información contenida en la muestra es:

$$\mathcal{F}_n(\boldsymbol{\theta}) = n\mathcal{F}_1(\boldsymbol{\theta}).$$

Observación:

Suponga $\mathbf{X}_1, \dots, \mathbf{X}_n$ vectores aleatorios independientes⁵. Entonces,

$$\ell_n(\boldsymbol{\theta}; \mathbf{X}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}; \mathbf{X}_i), \quad U_n(\boldsymbol{\theta}; \mathbf{X}) = \sum_{i=1}^n U_i(\boldsymbol{\theta}; \mathbf{X}_i), \quad \mathcal{F}_n(\boldsymbol{\theta}) = \sum_{i=1}^n \mathcal{F}_i(\boldsymbol{\theta}),$$

con $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top)^\top$.

⁵no necesariamente copias



Resultado 4:

Sea X una variable aleatoria con distribución P_θ . Suponga otra parametrización dada por $\theta = h(\phi)$ con h diferenciable. Entonces, la información de Fisher con relación a ϕ es dada por

$$\mathcal{F}_X^*(\phi) = \mathcal{F}_X(h(\phi))\{h'(\phi)\}^2,$$

donde $\mathcal{F}_X(\theta)$ es la información que $X \sim P_\theta$ contiene respecto de θ y $\mathcal{F}_X^*(\phi)$ es la información que $X \sim P_{h(\phi)}$ contiene sobre ϕ .

Demostración:

La función de log-verosimilitud con relación a $P_{h(\phi)}$ es:

$$\ell^*(\phi; x) = \log f(x; h(\phi)).$$

De este modo, la función score asume la forma:

$$U^*(\phi; X) = \frac{\partial}{\partial \phi} \log f(x; h(\phi)) = \frac{\partial}{\partial \theta} \log f(x; \theta) \Big|_{\theta=h(\phi)} h'(\phi),$$

luego,

$$\begin{aligned} \mathcal{F}_X^*(\phi) &= \text{var}_\phi \{U^*(\phi; X)\} = \text{var}_\phi \{U(h(\phi); X) h'(\phi)\} \\ &= \{h'(\phi)\}^2 \text{var}_\phi \{U(h(\phi); X)\} = \{h'(\phi)\}^2 \mathcal{F}_X(h(\phi)). \end{aligned}$$



Resultado 4:

Sea X una variable aleatoria con distribución P_θ . Suponga otra parametrización dada por $\theta = h(\phi)$ con h diferenciable. Entonces, la información de Fisher con relación a ϕ es dada por

$$\mathcal{F}_X^*(\phi) = \mathcal{F}_X(h(\phi))\{h'(\phi)\}^2,$$

donde $\mathcal{F}_X(\theta)$ es la información que $X \sim P_\theta$ contiene respecto de θ y $\mathcal{F}_X^*(\phi)$ es la información que $X \sim P_{h(\phi)}$ contiene sobre ϕ .

Demostración:

La función de log-verosimilitud con relación a $P_{h(\phi)}$ es:

$$\ell^*(\phi; x) = \log f(x; h(\phi)).$$

De este modo, la función score asume la forma:

$$U^*(\phi; X) = \frac{\partial}{\partial \phi} \log f(x; h(\phi)) = \frac{\partial}{\partial \theta} \log f(x; \theta) \Big|_{\theta=h(\phi)} h'(\phi),$$

luego,

$$\begin{aligned} \mathcal{F}_X^*(\phi) &= \text{var}_\phi \{U^*(\phi; X)\} = \text{var}_\phi \{U(h(\phi); X) h'(\phi)\} \\ &= \{h'(\phi)\}^2 \text{var}_\phi \{U(h(\phi); X)\} = \{h'(\phi)\}^2 \mathcal{F}_X(h(\phi)). \end{aligned}$$



Ejemplo:

Considere la distribución $\text{Poi}(\theta)$. Así,

$$f(x; \theta) = \exp(x \log \theta - \theta) / x!$$

Note que el parámetro natural es $\eta = \log \theta$. Luego,

$$f(x; \eta) = \exp(x\eta - e^\eta) / x!$$

En la parametrización original, tenemos que

$$U(\theta; x) = \frac{\partial}{\partial \theta} \log f(x; \theta) = \frac{\partial}{\partial \theta} (x \log \theta - \theta) = \frac{x}{\theta} - 1.$$

De este modo,

$$\mathcal{F}_X(\theta) = \text{var}_\theta \{U(\theta; X)\} = \text{var}_\theta \left(\frac{X}{\theta} - 1 \right) = \frac{1}{\theta^2} \text{var}_\theta(X) = \frac{1}{\theta}.$$

Por otro lado, $\theta = e^\eta = h(\eta)$. Así,

$$\mathcal{F}_X^*(\eta) = \mathcal{F}_X(h(\eta)) \{h'(\eta)\}^2 = \frac{1}{e^\eta} (e^\eta)^2 = e^\eta.$$



Observación:

Considere $\boldsymbol{\theta} = \mathbf{h}(\phi)$ con $\mathbf{h} : \mathbb{R}^k \rightarrow \mathbb{R}^p$ ($p \leq k$) y sea $\mathbf{H}(\phi) = \partial \mathbf{h}(\phi) / \partial \phi^\top$ matriz de rango completo. Entonces,

$$\mathcal{F}_*(\phi) = \mathbf{H}^\top(\phi) \mathcal{F}(\mathbf{h}(\phi)) \mathbf{H}(\phi).$$

