MAT-466: Quasi-verosimilitud

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Quasi-verosimilitud

Suponga Y variable aleatoria con

$$\mathsf{E}(Y) = \mu, \qquad \mathsf{var}(Y) = \sigma^2 V(\mu),$$

y $\mu = \mu(\boldsymbol{\beta})$ con $V(\mu)$ una función conocida. Bajo estas condiciones, la función

$$\Psi(\mu; Y) = \frac{Y - \mu}{\sigma^2 V(\mu)},$$

disfruta de las siguientes propiedades:

$$\mathsf{E}(\Psi) = 0, \qquad \mathsf{var}(\Psi) = \frac{1}{\sigma^2 V(\mu)}, \qquad \mathsf{E}\left(-\frac{\partial \Psi}{\partial \mu}\right) = \frac{1}{\sigma^2 V(\mu)}.$$

Esto permite definir la función de quasi-verosimilitud, como:

$$Q(\mu; y) = \frac{1}{\sigma^2} \int_y^\mu \frac{y - t}{V(t)}\, \mathsf{d}t$$

Evidentemente,

$$\Psi(\mu) = \frac{\partial Q(\mu; y)}{\partial \mu} = \frac{y - t}{\sigma^2 V(t)}\Big|_y^\mu = \frac{y - \mu}{\sigma^2 V(\mu)}.$$

Resultado 1 (Wedderburn, 1974)[1]:

La función quasi-score, satisface:

(i) $\mathsf{E}\{\Psi(\mu)\} = 0$.

(ii) $\mathsf{E}\left\{\dfrac{\partial \Psi(\mu)}{\partial \beta_j}\right\} = 0$, para $j = 1, \dots, p$.

(iii) $\mathsf{E}\{\Psi^2(\mu)\} = \mathsf{E}(-\partial \Psi(\mu)/\partial \mu) = \sigma^{-2}/V(\mu)$.

(iv) $\mathsf{E}\left\{\dfrac{\partial \Psi(\mu)}{\partial \beta_r}\dfrac{\partial \Psi(\mu)}{\partial \beta_s}\right\} = \mathsf{E}\left(-\dfrac{\partial^2 \Psi(\mu)}{\partial \beta_r \partial \beta_s}\right) = \dfrac{1}{\sigma^2 V(\mu)}\dfrac{\partial \mu}{\partial \beta_r}\dfrac{\partial \mu}{\partial \beta_s}$,
para $r, s = 1, \dots, p$.

[1] Biometrika **61**, 439-447.

Demostración:

(i) es directo desde la definición de Ψ, mientras que (ii) sigue de notar

$$\frac{\partial \Psi}{\partial \beta_j} = \frac{\partial \Psi}{\partial \mu}\frac{\partial \mu}{\partial \beta_j}.$$

Para mostrar (iii), note que

$$\mathsf{E}\{\Psi^2(\mu)\} = \frac{1}{\sigma^4}\,\mathsf{E}\left\{\frac{(Y-\mu)^2}{V^2(\mu)}\right\} = \frac{1}{\sigma^4}\frac{\sigma^2 V(\mu)}{V^2(\mu)} = \frac{1}{\sigma^2 V(\mu)}.$$

Por otro lado,

$$\begin{aligned} -\,\mathsf{E}\left(\frac{\partial \Psi}{\partial \mu}\right) &= -\,\mathsf{E}\left\{\frac{\partial}{\partial \mu}\left(\frac{Y-\mu}{\sigma^2 V(\mu)}\right)\right\} = -\frac{1}{\sigma^2}\,\mathsf{E}\left\{(Y-\mu)\frac{\partial}{\partial \mu}\frac{1}{V(\mu)} - \frac{1}{V(\mu)}\right\} \\ &= \frac{1}{\sigma^2 V(\mu)} \end{aligned}$$

Quasi-verosimilitud

Para probar (iv), considere

$$
\begin{aligned}
\mathsf{E}\Big(\frac{\partial\Psi}{\partial\beta_r}\frac{\partial\Psi}{\partial\beta_s}\Big) &= \mathsf{E}\Big\{\Big(\frac{\partial\Psi}{\partial\mu}\Big)\frac{\partial\mu}{\partial\beta_r}\Big(\frac{\partial\Psi}{\partial\mu}\Big)\frac{\partial\mu}{\partial\beta_s}\Big\} = \mathsf{E}\Big\{\Big(\frac{Y-\mu}{V(\mu)}\Big)^2\frac{\partial\mu}{\partial\beta_r}\frac{\partial\mu}{\partial\beta_s}\Big\} \\
&= \frac{1}{\sigma^4 V(\mu)}\,\mathsf{E}\{(Y-\mu)^2\}\frac{\partial\mu}{\partial\beta_r}\frac{\partial\mu}{\partial\beta_s} = \frac{1}{\sigma^2 V(\mu)}\frac{\partial\mu}{\partial\beta_r}\frac{\partial\mu}{\partial\beta_s},
\end{aligned}
$$

pues $\mathsf{var}(Y) = \sigma^2 V(\mu)$. Mientras que

$$
\begin{aligned}
-\,\mathsf{E}\Big(\frac{\partial^2\Psi}{\partial\beta_r\partial\beta_s}\Big) &= -\,\mathsf{E}\Big\{\frac{\partial}{\partial\beta_r}\Big(\frac{Y-\mu}{\sigma^2 V(\mu)}\Big)\frac{\partial\mu}{\partial\beta_r}\Big\} \\
&= -\,\mathsf{E}\Big\{\Big[\frac{Y-u}{\sigma^2}\frac{\partial}{\partial\beta_r}\frac{1}{V(\mu)} - \frac{1}{\sigma^2 V(\mu)}\frac{\partial\mu}{\partial\beta_r}\Big]\frac{\partial\mu}{\partial\beta_s}\Big\} \\
&= \frac{1}{\sigma^2 V(\mu)}\frac{\partial\mu}{\partial\beta_r}\frac{\partial\mu}{\partial\beta_s}.
\end{aligned}
$$

Resultado 2:

Si la distribución de Y es especificada en términos de μ, y suponga que tenemos una log-verosimilitud asociada

$$-\operatorname{E}\left(\frac{\partial^2\Psi}{\partial\mu^2}\right) \leq -\operatorname{E}\left(\frac{\partial^2\ell}{\partial\mu^2}\right)$$

Demostración:

En el Resultado 1, basta notar que

$$\operatorname{var}(Y) \geq -\frac{1}{\operatorname{E}(\partial^2\ell/\partial\mu^2)}.$$

Observación:

Evidentemente, conocer sólo la relación entre la media y varianza ofrece menos información que la obtenida desde la densidad de Y.

Ejemplo (modelo normal):

Suponga $V(\mu) = 1$, así la función de quasi-verosimilitud es dada por

$$Q(\mu; y) = \int_y^\mu \frac{y - t}{\sigma^2} \, \mathrm{d}t = -\frac{1}{\sigma^2} \frac{(y - \mu)^2}{2}, \qquad \mu, y \in \mathbb{R}.$$

Ejemplo (modelo Poisson):

Considere $V(\mu) = \mu$, luego

$$\begin{aligned} Q(\mu; y) = \int_y^\mu \frac{y - t}{\sigma^2 t} \, \mathrm{d}t &= \frac{1}{\sigma^2} (y \log \mu - \mu - y \log y + y) \\ &\propto \frac{1}{\sigma^2} (y \log \mu - \mu). \end{aligned}$$

Si suponemos $\sigma^2 = 1$ sigue que $Q(\mu; y)$ es proporcional a la log-verosimilitud de Poi(μ).

Ejemplo (modelo Bernoulli):

En este caso, $V(\mu) = \mu(1-\mu)$ y

$$Q(\mu; y) = \int_y^\mu \frac{y-t}{\sigma^2 t(1-t)}\,\mathrm{d}t = \frac{1}{\sigma^2}\Big\{y \log\Big(\frac{\mu}{1-\mu}\Big) + \log(1-\mu) - \log y\Big\}$$
$$\propto \frac{1}{\sigma^2}\Big\{y \log\Big(\frac{\mu}{1-\mu}\Big) + \log(1-\mu)\Big\}$$

Ejemplo:

Considere la función de varianza $V(\mu) = \mu^2(1-\mu)^2$

$$Q(\mu; y) = \int_y^\mu \frac{y-t}{\sigma^2 t^2(1-t)^2}\,\mathrm{d}t \propto \frac{1}{\sigma^2}\Big\{(2y-1)\log\Big(\frac{\mu}{1-\mu}\Big) - \frac{y}{\mu} - \frac{1-y}{1-\mu}\Big\}.$$

Este tipo de función se recomienda para $0 < \mu < 1$ y $0 \leq y \leq 1$. Se debe notar que la función $Q(\mu; y)$ no corresponde a ninguna verosimilitud conocida.

Quasi-verosimilitud

Suponga $Y_1, \dots, Y_n$ variables aleatorias independientes con quasi-verosimilitud conjunta:

$$Q(\boldsymbol{\mu}; \boldsymbol{y}) = \sum_{i=1}^{n} Q(\mu_i; y_i),$$

con $g(\mu_i) = \eta_i$ y $\eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, para $i = 1, \dots, n$.

Por analogía, la función quasi-desvío es dada por

$$\begin{aligned} D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}) &= 2\sigma^2 \{Q(\boldsymbol{y}; \boldsymbol{y}) - Q(\widehat{\boldsymbol{\mu}}; \boldsymbol{y})\} = -2\sigma^2 Q(\widehat{\boldsymbol{\mu}}; \boldsymbol{y}) \\ &= -2\sigma^2 \sum_{i=1}^{n} Q(\widehat{\mu}_i; y_i) = 2 \sum_{i=1}^{n} \int_{\widehat{\mu}_i}^{y_i} \frac{y_i - t}{V(t)} \, \mathrm{d}t \end{aligned}$$

Quasi-verosimilitud

Las ecuaciones de quasi-verosimilitud, para estimar $\boldsymbol{\beta}$ se obtienen diferenciando $Q(\boldsymbol{\mu};\boldsymbol{y})$ y pueden ser escritas como $\boldsymbol{\Psi}(\widehat{\boldsymbol{\beta}}) = \mathbf{0}$, con

$$\boldsymbol{\Psi}(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\boldsymbol{D}^\top\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}),$$

donde

$$\boldsymbol{D} = \partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}^\top = \boldsymbol{W}^{1/2}\boldsymbol{V}^{1/2}\boldsymbol{X},$$

con $\boldsymbol{V} = \operatorname{diag}(V_1,\dots,V_n)$, $\boldsymbol{W} = \operatorname{diag}(\omega_1,\dots,\omega_n)$ y $\omega_i = (\mathsf{d}\,\mu_i/\,\mathsf{d}\,\eta_i)^2/V_i$.

Además,

$$\boldsymbol{K}(\boldsymbol{\beta}) = -\,\mathsf{E}\left\{\frac{\partial\boldsymbol{\Psi}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^\top}\right\} = \frac{1}{\sigma^2}\boldsymbol{D}^\top\boldsymbol{V}^{-1}\boldsymbol{D}.$$

Esta matriz sigue el mismo rol que la matriz de información de Fisher. En efecto, la matriz de covarianza asintótica de $\widehat{\boldsymbol{\beta}}$ es

$$\mathsf{Cov}(\widehat{\boldsymbol{\beta}}) \approx \boldsymbol{K}^{-1}(\boldsymbol{\beta}) = \sigma^2(\boldsymbol{D}^\top\boldsymbol{V}^{-1}\boldsymbol{D})^{-1}.$$

Quasi-verosimilitud

El método 'Fisher'-scoring lleva al proceso iterativo

$$\begin{aligned}\boldsymbol{\beta}^{(r+1)} &= \boldsymbol{\beta}^{(r)} + (\boldsymbol{D}^{(r)\top}\boldsymbol{V}^{-(r)}\boldsymbol{D}^{(r)})^{-1}\boldsymbol{D}^{(r)}\boldsymbol{V}^{-(r)}(\boldsymbol{Y} - \boldsymbol{\mu}^{(r)}) \\ &= (\boldsymbol{D}^{(r)\top}\boldsymbol{V}^{-(r)}\boldsymbol{D}^{(r)})^{-1}\boldsymbol{D}^{(r)\top}\boldsymbol{V}^{-(r)}\boldsymbol{Z}_*^{(r)},\end{aligned}$$

con $\boldsymbol{Z}_* = \boldsymbol{D}\boldsymbol{\beta} + \boldsymbol{Y} - \boldsymbol{\mu}$.

Tenemos $\boldsymbol{D} = \boldsymbol{W}^{1/2}\boldsymbol{V}^{1/2}\boldsymbol{X}$, de este modo

$$\begin{aligned}\boldsymbol{D}^\top\boldsymbol{V}^{-1}\boldsymbol{D} &= \boldsymbol{X}^\top\boldsymbol{V}^{1/2}\boldsymbol{W}^{1/2}\boldsymbol{V}^{-1}\boldsymbol{W}^{1/2}\boldsymbol{V}^{1/2}\boldsymbol{X} = \boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X} \\ \boldsymbol{D}^\top\boldsymbol{V}^{-1} &= \boldsymbol{X}^\top\boldsymbol{W}^{1/2}\boldsymbol{V}^{-1/2} = \boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{W}^{-1/2}\boldsymbol{V}^{-1/2}.\end{aligned}$$

De ahí que

$$\boldsymbol{\beta}^{(r+1)} = (\boldsymbol{X}^\top\boldsymbol{W}^{(r)}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{W}^{(r)}\boldsymbol{Z}^{(r)},$$

con $\boldsymbol{Z} = \boldsymbol{\eta} + \boldsymbol{W}^{-1/2}\boldsymbol{V}^{-1/2}(\boldsymbol{Y} - \boldsymbol{\mu})$.

El estimador de momentos de σ^2 es dado por

$$\widehat{\sigma}^2_{\mathsf{MM}} = \frac{1}{n-p}\sum_{i=1}^n \frac{(y_i - \widehat{\mu}_i)^2}{V(\widehat{\mu}_i)}$$

Nelder y Pregibon (1987)[2] propusieron la función de quasi-verosimilitud extendida

$$Q^{+}(\mu; y) = -\frac{1}{2\sigma^2}\, D(y;\mu) - \frac{1}{2}\log\{2\pi\sigma^2 V(y)\},$$

donde

$$D(y;\mu) = 2\int_{\mu}^{y} \frac{y-t}{V(t)}\,\mathrm{d}t.$$

De este modo los estimadore de $\boldsymbol{\beta}$ y σ^2 son obtenidos mediante maximizar $Q^{+}(\boldsymbol{\mu}; \boldsymbol{y})$.

El estimador de quasi-verosimilitud extendidad $\widehat{\boldsymbol{\beta}}$, coincide con el estimador basado en $Q(\boldsymbol{\mu}; \boldsymbol{y})$. Mientras que el estimador para σ^2 es dado por

$$\widehat{\sigma}^2_{\mathsf{EQL}} = \left\{\frac{D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}})}{n}\right\}^{-1}.$$

[2] Biometrika **74**, 221-232.

Quasi-verosimilitud

Considere $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$ vectores aleatorios independientes, con $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ y suponga

$$f(y; \theta_{ij}, \phi) = \exp[\phi\{y\theta_{ij} - b(\theta_{ij})\} + c(y; \phi)],$$

donde $\mathsf{E}(Y_{ij}) = \mu_{ij} = b'(\theta_{ij})$, $\mathsf{var}(Y_{ij}) = \phi^{-1} V_{ij}$, $V_{ij} = \mathsf{d}\,\mu_{ij} / \mathsf{d}\,\theta_{ij}$. Además,

$$g(\mu_{ij}) = \eta_{ij}, \qquad \eta_{ij} = \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}.$$

De este modo,

$$\boldsymbol{\Psi}(\boldsymbol{\beta}) = \phi \sum_{i=1}^n \boldsymbol{D}_i^\top \boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i), \qquad \boldsymbol{K}(\boldsymbol{\beta}) = \phi \sum_{i=1}^n \boldsymbol{D}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{D}_i,$$

donde $\boldsymbol{D}_i = \boldsymbol{W}_i^{1/2} \boldsymbol{V}_i^{1/2} \boldsymbol{X}_i$.