1 Consistencia de M-estimadores

Supongamos que se tiene una familia de densidades $p(x,\theta)$ discreta o continua donde $\theta \in \Theta \subset R$. Tomemos una función $\psi(x,\theta) : R^2 \to R$ y llamemos $\eta(\theta,\theta^*) = E_\theta(\psi(x,\theta^*))$. Supondremos que para todo $\theta \in \Theta$ se cumple

$$\eta(\theta,\theta) = 0$$

Esta condición será llamada "consistencia de Fisher ".

Dada una muestra aleatoria $X_1,...,X_n$ de $p(x,\theta)$, el correspondiente M-estimador $\widehat{\theta}_n$ de θ se define como la solución de

$$\sum_{i=1}^{n} \psi(X_i,\widehat{\theta}_n) = 0 \tag{1}$$

Teorema. Supongamos que $\psi(x,\theta)$ es estrictamente monótona (creciente o decreciente) y continua respecto de θ. Sea $X_1,...,X_n$ una muestra aleatoria con densidad $p(x,\theta_0)$ Luego $\widehat{\theta}_n$ converge a θ casi seguramente.

Demostración

Vamos a suponer que $\psi(x,\theta)$ es estrictamente creciente como función de θ. La demostración para el caso decreciente es similar. Llamemos

$$\eta(\theta,\theta^*) = E_\theta(\psi(x,\theta^*))$$

Es inmediato que $\eta(\theta,\theta^*)$ es estrictamente monótona como función de θ^*

$\widehat{\theta}_n$ está definido por

$$\frac{1}{n}\sum_{i=1}^{n} \psi(X_i,\widehat{\theta}_n)) = 0$$

Sea

$$L_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \psi(X_i,\theta))$$

Tomemos $\varepsilon > 0$, vamos a mostrar que con probabilidad 1 existe n_0 tal que para $n \geq n_0$ se tiene

$$\theta_0 - \varepsilon \leq \widehat{\theta}_n \leq \theta_0 + \varepsilon$$

Por la monotonía de η tenemos

$$\eta(\theta_0,\theta_0-\varepsilon) < \eta(\theta_0,\theta_0) = 0 < \eta(\theta_0,\theta_0+\varepsilon)$$

Definamos

$$\delta = \min(\eta(\theta_0,\theta_0+\varepsilon), -\eta(\theta_0,\theta_0-\varepsilon)).$$

Luego $\delta > 0$ y

$$\eta(\theta_0,\theta_0-\varepsilon) \leq -\delta \leq \delta \leq \eta(\theta_0,\theta_0+\varepsilon)$$

Como $L_n(\theta_0+\varepsilon)$ converge casi seguramente a $\eta(\theta_0,\theta_0+\varepsilon)$ con probabilidad 1 existe n_1 tal que para todo $n\geq n1$

$$L_n(\theta_0+\varepsilon)\geq\delta/2$$

y como $L_n(\theta_0-\varepsilon)$ converge casi seguramente a $\eta(\theta_0,\theta_0-\varepsilon)$ existe n_2 tal que para todo $n\geq n_2$

$$L_n(\theta_0-\varepsilon)\leq-\delta/2$$

Luego como L_n es estrictamente creciente y continua, si $n_0=\max(n_1,n_2)$, para todo $n\geq n_0$ existe un único valor $\widehat{\theta}_n$ tal que $L_n(\widehat{\theta}_n)=0$ y ese valor tiene que satisfacer $\theta_0+\varepsilon\leq\widehat{\theta}_n\leq\theta_0-\varepsilon$.

1.1 Estimadores basados en momentos.

El estimador de los momentos se define por

$$\frac{1}{n}\sum_{i=1}^{n}g(x_i)=E_\theta(g(x_i))$$

Luego si definimos

$$\psi(x_i,\theta)=g(x_i)-E_\theta(g(x_i))$$

tenemos

$$\frac{1}{n}\sum_{i=1}^{n}(g(x_i)-E_\theta(g(x_i)))=0$$

$$\frac{1}{n}\sum_{i=1}^{n}\psi(x_i,\theta)=0$$

Observemos que se tiene

$$\eta(\theta,\theta)=E_\theta(\psi(x,\theta))=E_\theta(g(x))-E_\theta(g(x))=0$$

y por lo tanto se cumple la condición de consistencia de Fisher. Para que se cumpla la consistencia habrá que pedir que $E_\theta(g(x))$ sea una función continua y estrictamente creciente.

1.2 Estimador de máxima verosimilitud.

El estimador de máxima verosimilitud satisface

$$\frac{1}{n}\sum\psi_0(x_i,\theta)=0$$

con

$$\psi_0(x,\theta)=\frac{\partial\log(p(x,\theta))}{\partial\theta}\tag{2}$$

Hemos mostrado que bajo las condiciones de la desigualdad de Rao-Cramer $\eta(\theta,\theta)=E_\theta(\psi_0(x,\theta))=0$. Luego si ψ_0 es estrictamente monótona y continua, tambien es fuertemente consistente. La hipótesis de continuidad y monotonicidad pueden relajarse. También la consistencia fuerte es valida para el caso que se estimen varios parámetros simultaneamente.

2 Asintótica normalidad de los M-estimadores.

Consideremos un M-estimador $\widehat{\theta}_n$ definido por (1)

Teorema. Supongamos que

(1) $\psi(x,\theta)$ tiene dos derivadas continuas respecto de θ.

(2) $\widehat{\theta} \xrightarrow{p} \theta_0$

(3)

$$E_{\theta_0}(\psi^2(x,\theta_0))<\infty$$

(4)

$$E\left(\frac{\partial\psi(x_i,\theta_0)}{\partial\theta}\right)\neq 0$$

(5) Existe $\varepsilon>0$ y $\zeta(x)$ tal que

$$\sup_{|\theta-\theta_0|\leq\varepsilon}\left|\frac{\partial^2\psi(x_i,\theta_0)}{\partial\theta^2}\right|\leq\zeta(x)$$

y

$$E(\zeta(x))<\infty$$

Luego se tiene que

$$n^{1/2}(\widehat{\theta}_n-\theta_0)\xrightarrow{\mathcal{D}} N\left(0,\frac{A(\psi,\theta_0)}{B^2(\psi,\theta_0)}\right)$$

donde

$$A(\psi,\theta)=E_\theta(\psi^2(x,\theta))$$

y

$$B(\psi,\theta)=E\left(\frac{\partial\psi(x,\theta)}{\partial\theta}\right) \tag{3}$$

Demostración

Tenemos

$$\sum_{i=1}^n\psi(x_i,\widehat{\theta}_n)=0$$

y luego desarrollando en serie de Taylor en el punto θ_0 se tiene

$$0=\sum_{i=1}^n\psi(x_i,\widehat{\theta}_n)=\left(\sum_{i=1}^n\psi(x_i,\theta_0)\right)+\left(\sum_{i=1}^n\frac{\partial\psi(x_i,\theta_0)}{\partial\theta}\right)(\widehat{\theta}_n-\theta_0)$$
$$+\frac{1}{2}\left(\sum_{i=1}^n\frac{\partial^2\psi(x_i,\theta_n^*)}{\partial\theta^2}\right)(\widehat{\theta}_n-\theta_0)^2$$

Por lo tanto

$$-\sum_{i=1}^{n}\psi(x_i,\theta_0)=\left[\left(\sum_{i=1}^{n}\frac{\partial\psi(x_i,\theta_0)}{\partial\theta}\right)+\frac{1}{2}\left(\left(\sum_{i=1}^{n}\frac{\partial^2\psi(x_i,\theta_n^*)}{\partial\theta^2}\right)(\widehat{\theta}_n-\theta_0)\right)\right](\widehat{\theta}_n-\theta_0)$$

y por lo tanto

$$(\widehat{\theta}_n-\theta_0)=\frac{-\sum_{i=1}^{n}\psi(x_i,\theta_0)}{\left(\sum_{i=1}^{n}\frac{\partial\psi(x_i,\theta_0)}{\partial\theta}\right)+\frac{\widehat{\theta}_n-\theta_0}{2}\left(\sum_{i=1}^{n}\frac{\partial^2\psi(x_i,\theta_n^*)}{\partial\theta^2}\right)}$$

y

$$\begin{aligned}n^{1/2}(\widehat{\theta}_n-\theta_0)&=\frac{n^{-1/2}\sum_{i=1}^{n}-\psi(x_i,\theta_0)}{\frac{1}{n}\left(\sum_{i=1}^{n}\frac{\partial\psi(x_i,\theta_0)}{\partial\theta}\right)+\frac{1}{2n}\left(\left(\sum_{i=1}^{n}\frac{\partial^2\psi(x_i,\theta_n^*)}{\partial\theta^2}\right)(\widehat{\theta}_n-\theta_0)\right)}\\&=\frac{A_n}{B_n+C_n}\end{aligned}\tag{4}$$

donde

$$A_n=n^{-1/2}\sum_{i=1}^{n}-\psi(x_i,\theta_0)$$

$$B_n=\frac{1}{n}\left(\sum_{i=1}^{n}\frac{\partial\psi(x_i,\theta_0)}{\partial\theta}\right)$$

y

$$C_n=\frac{\widehat{\theta}_n-\theta_0}{2n}\left(\sum_{i=1}^{n}\frac{\partial^2\psi(x_i,\theta_n^*)}{\partial\theta^2}\right)$$

Vamos a mostrar que

$$A\overset{\mathcal{D}}{\to}N(0,A(\psi,\theta_0)),\tag{5}$$

$$B_n\overset{c.s.}{\to}B(\psi,\theta_0),\tag{6}$$

y

$$C_n\overset{p}{\to}0\tag{7}$$

Comencemos mostrando (5). Esto se deduce inmediatamente del Teorema Central del Límite observando que

$$E_{\theta_0}(-\psi(x_i,\theta_0))=0\ \text{ y } \mathrm{var}_{\theta_0}(-\psi(x_i,\theta_0))=A(\psi,\theta_0)$$

Por otro lado (6) se deduce inmediatamente de la Ley Fuerte de los Grandes Números.

Llamemos

$$D_n=\frac{1}{2n}\left|\sum_{i=1}^{n}\frac{\partial^2\psi(x_i,\theta_n^*)}{\partial\theta^2}\right|$$

Sea $U_n = \{|\widehat{\theta}_n - \theta| \leq \varepsilon\}$ y sea $h = \frac{1}{2}E(\zeta(x))$. Tenemos que

$$\begin{aligned} D_n I_{U_n} &\leq \frac{1}{2n}\left(\sum_{i=1}^{n}\left|\frac{\partial^2\psi(x_i,\theta_n^*)}{\partial\theta^2}\right| I_{U_n}\right) \\ &\leq \frac{1}{2n}\left(\sum_{i=1}^{n}\sup_{|\theta-\theta_0|\leq\varepsilon}\left|\frac{\partial^2\psi(x_i,\theta)}{\partial\theta^2}\right|\right) \\ &\leq \frac{1}{2n}\left(\sum_{i=1}^{n}\zeta(x_i)\right) = F_n \end{aligned}$$

Llamemos U_n^c al complemento de U_n. Luego

$$\begin{aligned} P(D_n \geq h+1) &= P(\{D_n I_{U_n} \geq h+1\}\} + P(D_n I_{U_n^c} \geq h+1\}) \\ &\leq P(F_n \geq h+1) + P(U_n^c) \end{aligned}$$

Por la ley de los grandes números $F_n \overset{c.s.}{\rightarrow} h$, se tiene que

$$P(F_n \geq h+1) \rightarrow 0$$

y por la hipótesis 2

$$P(U_n^c) \rightarrow 0$$

Luego

$$P(D_n \geq h+1) \rightarrow 0 \tag{8}$$

Luego dado $\delta > 0$

$$\{|C_n| > \delta\} \subset \{D_n \geq h+1\} \cup \{|\theta_n - \theta_0| \geq \delta/(h+1)\}$$

y por lo tanto

$$P(|C_n| > \delta) \leq P(D_n \geq h+1) + P(|\theta_n - \theta_0| \geq \delta/(h+1))$$

Usando (8) y la hipótesis (5) se tiene

$$P(C_n > \delta) \rightarrow 0$$

y por lo tanto (7) se cumple

De (4) (5), 6) y (7) por Slutzky resulta que

$$n^{1/2}(\widehat{\theta}_n - \theta_0) \overset{\mathcal{D}}{\rightarrow} N\left(0, \frac{A(\psi,\theta_0)}{B^2(\psi,\theta_0)}\right)$$

2.1 Estimadores basados en momentos

Apliquemos el teorema a los estimadores basados en momentos.

Luego las hipotesis se transforman en
(1) $E_\theta g(x,\theta)$ tiene dos derivadas continuas despecto de θ.
(2) $\widehat{\theta}_n \xrightarrow{p} \theta_0$
(3)

$$E_{\theta_0}(g^2(x)) < \infty$$

(4)

$$E\left(\frac{\partial g(x_i,\theta_0)}{\partial \theta}\right) \neq 0$$

(5) Existe $\varepsilon > 0$ tal que

$$\sup_{|\theta-\theta_0|\leq\varepsilon} \left| \left.\frac{\partial^2 E_\theta(g(x))}{\partial \theta^2}\right|_{\theta=\theta_0} \right| \leq \infty$$

En este caso

$$A(\psi,\theta) = \mathrm{var}_\theta(g(x)) \ \text{ y } B(\psi,\theta) = -E\left(\frac{\partial g(x_i,\theta_0)}{\partial \theta}\right)$$

2.2 Estimador de máxima verosimilitud

En el caso del estimador de máxima verosimilitud se tiene que $\psi = \psi_0$ dada por 2) y luego

$$A(\psi_0,\theta) = I(\theta)$$

El siguiente teorema muestra quien es $B(\psi_0,\theta)$.

Teorema. Si se cumplen las hipótesis de Rao Cramer entonces $B(\psi_0,\theta) = -I(\theta)$.

Demostración. Tenemos que

$$E_\theta\left(\psi_0(x,\theta)\right) = 0$$

es decir

$$\int I_S(x)\frac{\delta \log(p(x,\theta)}{\partial \theta}p(x,\theta)dx = 0$$

donde $S = \{x : p(x,\theta) > 0\}$. Derivando nuevamente dentro de la integral se tiene

$$\int I_S(x)\frac{\delta^2 \log(p(x,\theta)}{\partial \theta}p(x,\theta)dx + \int I_S(x)\frac{\delta \log(p(x,\theta)}{\partial \theta}\frac{\partial p(x,\theta)}{\partial \theta}dx = 0 \qquad (9)$$

Luego

$$\int I_S(x)\frac{\delta \log(p(x,\theta)}{\partial\theta}\frac{\partial p(x,\theta)}{\partial\theta}dx = \int I_S(x)\left(\frac{\delta \log(p(x,\theta)}{\partial\theta}\right)^2 p(x,\theta)dx$$
$$= E_\theta\left(\left(\frac{\delta \log(p(x,\theta)}{\partial\theta}\right)^2\right)$$
$$= I(\theta)$$

y reemplazando en (9) se obtiene

$$\int I_S(x)\frac{\delta^2 \log(p(x,\theta)}{\partial\theta}p(x,\theta)dx = -I(\theta)$$

y luego

$$B(\psi_0,\theta) = E_\theta\left(\frac{\partial\psi(x,\theta_0)}{\partial\theta}\right)$$
$$= \int I_S(x)\frac{\delta^2 \log(p(x,\theta)}{\partial\theta}p(x,\theta)dx$$
$$= -I(\theta)$$

Esto demuestra el Teorema.

Corolario. Bajo las condiciones 1-5 y las hipotesis de Rao Cramer, el estimador de máxima verosimilitud $\widehat{\theta}_n$ satisface

$$n^{1/2}(\widehat{\theta}_n-\theta_0)\overset{\mathcal{D}}{\to}N\left(0,\frac{1}{I(\theta)}\right)$$

El siguiente Teorema muestra que cualquier otro M-estimador tiene una varianza mayor que el estimador de máxima verosimilitud.

Teorema. Supongamos que se cumplan las condiciones de Rao Cramer. Luego

(a) si $\widehat{\theta}_n$ es un M-estimador entonces

$$\frac{A(\psi,\theta)}{B^2(\psi,\theta)}\geq I(\theta)$$

(b) La igualdad se cumple si $\widehat{\theta}_n$ coincide con el M-estimador correspondiente a $\psi(x,\theta)$ tal que

$$P(\psi(x,\theta)=\psi_0(x,\theta))=1$$

y luego $\widehat{\theta}_n$ coincide con probabilidad 1 con el estimador de máxima verosimilitud

Demostración. Como

$$E_\theta(\psi(x,\theta)) = \int I_S(x)\psi(x,\theta)p(x,\theta)dx = 0$$

derivando se obtiene

$$\int I_S \frac{\partial \psi(x,\theta_0)}{\partial \theta} p(x,\theta)dx + \int I_S(x)\psi(x,\theta)\frac{\partial p(x,\theta}{\partial \theta}dx = 0$$

$$\begin{aligned}
B(\psi,\theta) &= E_\theta\left(\frac{\partial \psi(x,\theta_0)}{\partial \theta}\right) \\
&= \int I_S \frac{\partial \psi(x,\theta_0)}{\partial \theta} p(x,\theta)dx \\
&= -\int I_S(x)\psi(x,\theta)\frac{\partial p(x,\theta)}{\partial \theta}dx \\
&= -\int I_S(x)\psi(x,\theta)\frac{\partial \log p(x,\theta}{\partial \theta}p(x,\theta)dx \\
&- \int I_S(x)\psi(x,\theta)\psi_0(x,\theta)p(x,\theta)dx
\end{aligned}$$

Como $E_\theta\left(\psi(x,\theta)\right) = E_\theta\left(\psi_0(x,\theta)\right) = 0$, se tiene

$$B(\psi,\theta) = \mathrm{cov}_\theta(\psi(x),\psi_0(x))$$

y por la desigualdad de Cauchy-Schwarz se tiene

$$B^2(\psi,\theta) \le \mathrm{var}_\theta(\psi(x))\mathrm{var}_\theta(\psi_0(x))$$

Pero como $\mathrm{var}_\theta(\psi(x)) = A(\psi,\theta)$ y $\mathrm{var}_\theta(\psi_0(x)) = I(\theta)$ se tiene

$$B^2(\psi,\theta) \le A(\psi,\theta)I(\theta)$$

y luego

$$\frac{A(\psi,\theta)}{B^2(\psi,\theta)} \ge I(\theta)$$

Esto prueba (a)

Para probar (b) observemos que la igualdad se cumple si y solo si

$$P_\theta(\psi(x) = A(\theta)\psi_0(x) + B(\theta)) = 1$$

Pero como

$$\begin{aligned}
E_\theta(A(\theta)\psi_0(x) + B(\theta)) &= A(\theta)E_\theta(\psi_0(x)) + B(\theta) \\
&= B(\theta) = 0
\end{aligned}$$

Resulta que $B(\theta) = 0$. Pero es inmediato que $A(\theta)\psi_0(x)$ define el mismo estimador que $\psi_0(x)$. Luego $\widehat{\theta}_n$ coincide con el estimador de máxima verosimilitud con probabilidad 1.