Aproximación de la varianza

Vimos que si $X_1, \ldots, X_n$ son iid $f(x/\theta)$ y $\hat{\theta}$ es el MLE de θ y $I_n(\theta) = E_\theta\left(\frac{\partial}{\partial\theta} \log L(\theta/x)\right)^2$ (la información de Fisher) entonces

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n]{} N\left(0, \frac{1}{I(\theta)}\right) \quad (*)$$

donde $I(\theta) = E_\theta\left(\frac{\partial}{\partial\theta} \log f(\theta/x_1)\right)^2$ (la información de una observación) (Este teorema es conocido como eficiencia asintótica del MLE)

El método delta nos decía que si $\sqrt{n}(Y_n - \theta) \to N(0, \sigma^2)$ entonces $\sqrt{n}(h(Y_n) - h(\theta)) \to N(0, \sigma^2 (h'(\theta))^2)$

Uniendo los 2 resultados tenemos que

$$\sqrt{n}(h(\hat{\theta}) - h(\theta)) \to N\left(0, \frac{(h'(\theta))^2}{I(\theta)}\right)$$

de donde $\mathrm{Var}(\sqrt{n}\, h(\hat{\theta})/\theta) \to \frac{(h'(\theta))^2}{I(\theta)}$ o sea

$$\mathrm{Var}\left(\sqrt{n}\, h(\hat{\theta})/\theta\right) \simeq \frac{\left(h'(\theta)\right)^2}{I(\theta)} \Rightarrow \mathrm{Var}\left(h(\hat{\theta})/\theta\right) \simeq \frac{\left(h'(\theta)\right)^2}{n\, I(\theta)}$$

Vimos también que $n\, I(\theta) = I_n(\theta)$

o sea $$\mathrm{Var}\left(h(\hat{\theta})/\theta\right) \simeq \frac{\left(h'(\theta)\right)^2}{I_n(\theta)}$$

Vimos tambien que $E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X/\theta)\right)^2\right) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X/\theta)\right)$

luego $$\mathrm{Var}\left(h(\hat{\theta})\,|\,\theta\right) \simeq \frac{\left(h'(\theta)\right)^2}{E_\theta\left(-\frac{\partial^2}{\partial\theta^2}\log L(\theta/X)\right)} \simeq$$

$$\simeq \frac{\left[h'(\theta)^2\right]_{\theta=\hat{\theta}}}{\underbrace{-\frac{\partial^2}{\partial\theta^2}\log L(\theta/X)\Big|_{\theta=\hat{\theta}}}_{\hat{I}_n(\theta) \text{ información de Fisher observada}}}$$

$\mathrm{Var}_{\hat{\theta}}(h(\hat{\theta})) = \mathrm{Var}\left(h(\hat{\theta})\,|\,\hat{\theta}\right)$ es un estimador consistente de

$\mathrm{Var}_\theta(h(\hat{\theta})) = \mathrm{Var}\left(h(\hat{\theta})\,|\,\theta\right)$

Ejemplo $X_1, \ldots, X_n \sim Ber(p)$ iid

Sabemos que $\hat{p} = \frac{\sum X_i}{n}$ es un MLE de p

$$Var_p(\hat{p}) = \frac{1}{n^2} Var_p\left(\sum X_i\right) = \frac{\cancel{n}}{n^{\cancel{2}}} Var(X_i) = \frac{p(1-p)}{n}$$

Hallemos usando el resultado anterior $\widehat{Var}_p(\hat{p})$, con $h(p) = p$

$$\widehat{Var}_p(\hat{p}) \simeq \frac{\overbrace{[h'(p)]^2}^{=1}\Big|_{p=\hat{p}}}{-\frac{\partial^2}{\partial p^2} \log L(p/x)\Big|_{p=\hat{p}}}$$

$$\log L(p/x) = \log \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = \log p^{\sum x_i}(1-p)^{n-\sum x_i}$$

$$= \sum x_i \log p + \left(n - \sum x_i\right) \log(1-p) =$$

$$= n\hat{p} \log p + n(1-\hat{p}) \log(1-p)$$

luego $$\frac{\partial^2}{\partial p^2} \log L(p/x) = -\frac{n\hat{p}}{p^2} - \frac{n(1-\hat{p})}{(1-p)^2}$$

entonces $$\partial^2 \log L(p/x)\Big| = -n\hat{p} - n(1-\hat{p}) = -n$$

entonces $\left.\dfrac{\partial^2 \log L(p/x)}{\partial p^2}\right|_{p=\hat{p}} = \dfrac{-n\hat{p}}{\hat{p}^2} - \dfrac{n(1-\hat{p})}{(1-\hat{p})^2} = \dfrac{-n}{\hat{p}(1-\hat{p})}$

Recordemos que usamos $-\dfrac{\partial^2 \log L(p/x)}{\partial p^2} = \dfrac{n}{\hat{p}(1-\hat{p})}$

o sea $\hat{Var}_p(\hat{p}) = \dfrac{1}{n/\hat{p}(1-\hat{p})} = \dfrac{\hat{p}(1-\hat{p})}{n}$

Aplicando el teorema de eficiencia asintótica del MLE

$$\sqrt{n}(\hat{p}-p) \to N(0, p(1-p))$$

ahora

$$\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}} = \underbrace{\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{p(1-p)}}}_{\to N(0,1)} \underbrace{\frac{\sqrt{p(1-p)}}{\sqrt{\hat{p}(1-\hat{p})}}}_{\to 1} \to N(0,1)$$

Slutsky

Definición 85 Si dos estimadores W_n y V_n satisfacen

$$\sqrt{n}\,[W_n - \tau(\theta)] \to N(0, \sigma_W^2)$$

$$\sqrt{n}\,[V_n - \tau(\theta)] \to N(0, \sigma_V^2)$$

La eficiencia relativa asintótica (ARE) de V_n con respecto a W_n es

$$ARE(V_n, W_n) = \frac{\sigma_W^2}{\sigma_V^2}$$

Estimación no paramétrica de densidades

Supongamos que $X_1, \ldots, X_n$ son v.a independientes idénticamente distribuidas con densidad f.

Sea $\mathcal{F}$ la familia de funciones absolutamente continuas con densidad f

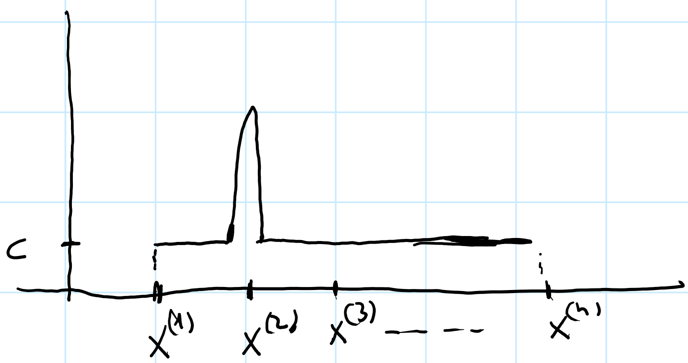
La idea es hallar f a partir de la muestra

Una primera idea es usar el método de máxima verosimilitud o sea maximizando

$$\prod_{i=1}^{n} \hat{f}(X_i)$$

Pero este es un enfoque que no es adecuado pues por ejemplo si consideremos la muestra ordenada $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$

Considero una función como la dibujada.



Tomo $c>0$ tal que el área del rectángulo sea pequeño y construyo una función que en $X^{(2)}$ sea arbitrariamente grande (pareciéndose a una masa puntual) y que su integral sea 1 y tal que la función sea mayor que c siempre

luego

$$\prod_{i=1}^{n} f(X_i) \geq c^{n-1} f(X^{(2)})$$

Como n y c son fijos c^{n-1} es fijo, luego puedo elegir f tal que $c^{n-1} f(X^{(2)}) > M \quad \forall M$ luego $\max_{f \in \mathcal{F}_c} \prod f(X_i) = +\infty$

así pues la verosimilitud no tiene máximo y este método no sirve.

Hay algunos casos que no trataremos donde máxima verosimilitud funciona como sea en la familia de las funciones monótonas o cuando son lipschitz de cte fija.

Van Vapnik (1982) y Thompson (1978)

Otro enfoque sería construir histogramas

Consideremos $I_1, I_2, \dots, I_p$ una serie de intervalos disjuntos. Definimos el histograma como

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n} \mathbb{1}_{I_i}(X_j) \mathbb{1}_{I_i}(x)$$

Si $x \in I_i$ cuento cuantos $X_j \in I_i$ o sea

$$f_n(x) = \frac{\#\{j : X_j \in I_i\}}{\quad} \quad \text{si } x \in I_i$$

$$f_n(x) = \frac{\#\{j : X_j \in I_i\}}{n} \quad \text{si } x \in I_i$$

Observemos el estimador

$$f_n(x) = \sum_{i=1}^{p} \hat{a}_i \mathbb{1}_{I_i}(x) \quad \text{con} \quad \hat{a}_i = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{I_i}(X_j)$$

Por la ley de los grandes números $\hat{a}_i \to P(X \in I_i) = a_i$

luego

$$f_n(x) \xrightarrow{c.s} \sum_{i=1}^{p} a_i \mathbb{1}_{I_i}(x)$$

Vemos que el límite resultante es un histograma pero no toda densidad es el histograma. O sea al considerar estimadores de histograma en realidad restringí la familia de funciones que considero.

Cuando analizamos el estimador vemos que el problema es que p es fijo, entonces podemos considerar $p(n) \to +\infty$ cuando $n \to \infty$.

Tenemos dos convergencias: de $\hat{a}_i \to a_i$ y la del histograma a f.

Encararemos este problema de otra manera. Un primer antecedente está en M. Rosemblatt (1956) y en Parzen (1962)

Debroye dice que Akaike lo hizo antes también en 1956

Podemos pensarlo como un funcional tal que $F' = f$ o sea

$$f(x) = F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h}$$

entonces podemos considerar un estimador

$$\hat{f}_{n,h}(x) = \lim_{h \to 0} \frac{F_n(x+h) - F_n(x-h)}{2h}$$

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leq x\}$$

Por el teorema fundamental de la estadística $F_n \xrightarrow{c.s} F$ y también tenemos la dependencia de h. Tenemos entonces un problema de límite doble $n \to +\infty$, $h \to 0$.

Hay un compromiso pues si $h \to 0$ muy rápidamente entonces no se logra que $F_n \to F$. Por eso es conveniente considerar h como función de n o sea $h(n)$

Consideremos ahora el estimador

$$f_n(x) = \frac{1}{2h(n)} \left(F_n(x+h(n)) - F_n(x-h(n)) \right) =$$

$$= \frac{1}{n h(n)} \frac{1}{2} \# \{ i : X_i \in (x-h(n), x+h(n)] \} =$$

$$= \frac{1}{n h(n)} \sum_{i=1}^{n} \frac{1}{2} \mathbb{1}_{(x-h(n),\, x+h(n)]}(X_i) =$$

$$= \frac{1}{n h(n)} \sum_{i=1}^{n} \frac{1}{2} \mathbb{1}_{(-1,\, 1]}\left(\frac{x-X_i}{h(n)}\right) = \frac{1}{n h(n)} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h(n)}\right)$$

donde $K(t) = \frac{1}{2} \mathbb{1}_{(-1,1]}(t)$

Olvidemos que K sea el indicador y consideremos

$$\hat{f}_n(x) = \frac{1}{n h(n)} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h(n)}\right) = \int \frac{1}{h(n)} K\left(\frac{x-t}{h(n)}\right) dF_n(t) = K_h(t) * F_n$$

con $K_h(t) = \frac{1}{h(n)} K(t/h)$

Si $K(t) \geq 0$, $\int K(t)\,dt = 1$ y si $K, f \in L^1 \cap L^2$ se obtiene $K_h * F \to f$ casi seguramente si $h \to 0$ en cada punto de continuidad x de f

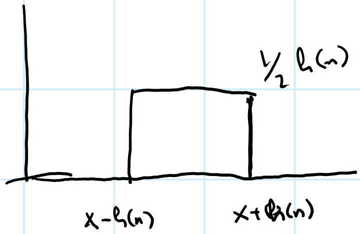
Necesitamos que $F_n \to F$ y ver si $h(n) \to 0$ funciona que $K_h * F_n \underset{n,h}{\to} f$

Cuando se utiliza el núcleo uniforme $K(t) = \frac{1}{2} I_{(-1,1]}(t)$

al considerar $\frac{1}{\ldots} \mathbb{1}_{(x-h(n),\, x+h(n)]}(x)$

$2h(n)$

La función es



Veamos que lo que hacemos fue centrar en x y escalar, el área sigue siendo 1 todavía o sea lo transformamos en algo parecido a la masa puntual

Volvamos ahora al núcleo uniforme y al estimador

$$\hat{f}_{n,h} = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)$$

Veamos que es consistente para ello calculemos la esperanza y la varianza

$$E\left(\hat{f}_{n,h}(x)\right) = \frac{1}{h}E\left(K\left(\frac{x-X_i}{h}\right)\right) = \frac{1}{2h}E\left(\mathbb{1}_{(x-h,\,x+h]}(X_i)\right) =$$

$$= \frac{1}{2h}\int_{x-h}^{x+h} f(t)\,dt \xrightarrow[h\to 0]{} f(x)$$ por el teorema fundamental del cálculo si f es continua en x

o sea $\hat{f}_{n,h}(x)$ es asintóticamente insesgado para x fijo y de continuidad de x.

$$\mathrm{Var}\left(\hat{f}_{n,h}(x)\right) = \frac{1}{\quad}\,\mathrm{Var}\left(I_{(x-h(n),\,x+h(n)]}(X_i)\right) =$$

$$\mathrm{Var}\left(\hat{f}_{n,h}(x)\right) = \frac{1}{n^2h^2 4} \cancel{n}\, \mathrm{Var}\left(I_{(x-h(n),\, x+h(n)]}(X_i)\right) =$$

$$= \frac{1}{4nh^2}\, p(1-p) \circledast$$ con $p = P(X_i \in (x-h, x+h])$ pues la indicatriz tiene distribución bernoulli

$$\circledast = \frac{(\overbrace{1-p}^{\to 1 \text{ si } h\to 0})}{2nh} \underbrace{\frac{1}{2h}\int_{x-h}^{x+h} f(t)\,dt}_{\to f(x) \text{ si } h\to 0 \text{ pues } f \text{ es continua en } x} \longrightarrow \frac{f(x)}{2nh}$$

Ahora si hago que $h \to 0$ y $nh \to \infty$ obtengo que $\mathrm{Var}(\hat{f}_{n,h}(x)) \to 0$ y entonces $\hat{f}_{n,h}(x) \xrightarrow{P} f(x)$ luego es consistente