Convergencia Completa

jueves, 8 de noviembre de 2018 11:57

Lema 97 (Desigualdad de Bernstein)

Sean $Y_1, Y_2, \dots, Y_n$ variables aleatorias independientes con

$E(Y_i) = 0$, $|Y_i| \leq c$ ctp $(P(|Y_i| \leq c) = 1)$

$E(Y_i^2) = \sigma_i^2$, $s_n^2 = \sum_{i=1}^{n} \sigma_i^2$ $\quad S_n = \sum_{i=1}^{n} Y_i$

entonces $P(S_n \geq \varepsilon) \leq \exp\left( - \dfrac{\varepsilon^2}{2(s_n^2 + c\varepsilon)} \right)$

En particular se obtiene que:

$$P\left( \frac{1}{n} \left| \sum_{i=1}^{n} Y_i \right| > \varepsilon \right) \leq 2 \exp\left( \frac{-n^2 \varepsilon^2}{2(s_n^2 + nc\varepsilon)} \right)$$

Teorema 98 (Convergencia completa de los estimadores basados en núcleos)

Sean $f_n(x) = \dfrac{1}{nh} \sum_{i=1}^{n} K\left( \dfrac{x - X_i}{h} \right)$ con $X_1, \dots, X_n \sim f$ independientes

El núcleo $K$ cumple las hipótesis del lema de Bochner e

$\int K(u)\,du = 1$ $\qquad h \to 0$, $nh \to \infty$ y $\dfrac{nh}{\log n} \to \infty$

entonces $f_n(x) \xrightarrow{c} f(x)$ en $x$ de continuidad de $f$

Dem: Debemos probar que $\sum^{+\infty} P(|f_n(x) - f(x)| > \varepsilon) < \infty$

Dem: Debemos probar que $\displaystyle\sum_{n=1}^{+\infty} P(|f_n(x)-f(x)|>\varepsilon)<\infty$

$$\{|f_n(x)-f(x)|>\varepsilon\} = \left\{\left|\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)-f(x)\right|>\varepsilon\right\} =$$

$$= \left\{\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{h}K\left(\frac{x-X_i}{h}\right)-f(x)\right)\right|>\varepsilon\right\}$$ entonces

$$P(|f_n(x)-f(x)|>\varepsilon) = P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{h}K\left(\frac{x-X_i}{h}\right)-f(x)\right)\right|>\varepsilon\right) \leq$$

$$\leq P\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{h}K\left(\frac{x-X_i}{h}\right)-f(x)\right|>\varepsilon\right)$$

Llamemos $Y_i = \frac{1}{h}K\left(\frac{x-X_i}{h}\right)-f(x)$

Como los $X_i$ son independientes los $Y_i$ lo son

$$E(Y_i) = E\left(\frac{1}{h}K\left(\frac{x-X_i}{h}\right)\right)-f(x) = \underbrace{\frac{1}{h}\int K\left(\frac{x-t}{h}\right)f(t)\,dt}_{\to f(x)} - f(x) \to 0$$

$$|Y_i| = \left|\frac{1}{h}K\left(\frac{x-X_i}{h}\right)-f(x)\right| \leq c \quad \text{si } K \text{ es acotada}$$

$$E(Y_i^2) = E\left(\frac{1}{h^2}K^2\left(\frac{x-X_i}{h}\right)\right)-2f(x)E\left(\frac{1}{h}K\left(\frac{x-X_i}{h}\right)\right)+f^2(x) =$$

$$E(Y_i^2) = \frac{1}{h}\,\underbrace{\frac{1}{h}\int K^2\left(\frac{x-t}{h}\right)f(t)\,dt}\,-2f(x)\,\underbrace{\frac{1}{h}\int K\left(\frac{x-t}{h}\right)f(t)\,dt}+f^2(x)$$

Bochner $\rightarrow f(x)\int K^2(t)\,dt$ $\qquad\qquad \rightarrow f(x)$

entonces

$$E(Y_i^2) \rightarrow \frac{1}{h} f(x) \int K^2(t)\,dt - f^2(x)$$

Aplicando la desigualdad de Bernstein se tiene

$$P\left(\frac{1}{n}\sum_{i=1}^{n}|Y_i| > \varepsilon\right) \leq 2\exp\left(\frac{-n^2\varepsilon^2}{2\left(n\left[\frac{1}{h} f(x)\int K^2(t)\,dt - f^2(x)\right] + cn\varepsilon\right)}\right)$$

ahora

$$\frac{-\cancel{n^2}\varepsilon^2}{2\left(\cancel{n}\left(\frac{f(x)}{h}\int K^2(t)\,dt - f(x) + c\varepsilon\right)\right)} = \frac{-n\varepsilon^2}{2\left(\frac{c_2}{h} - c_1 + c\varepsilon\right)} =$$

$$= \frac{-n\varepsilon^2}{2\left(\frac{c_2}{h} + c_3\right)} \leq \frac{-n\varepsilon^2}{2\frac{c_2}{h}} = -nhc'$$

Entonces $P\left(\frac{1}{n}\sum_{i=1}^{n}|Y_i| > \varepsilon\right) \leq 2\exp(-nhc')$

luego $$\sum_{n=1}^{\infty} P\left(\frac{1}{n}\sum_{i=1}^{n}|Y_i| > \varepsilon\right) \leq \sum_{n=1}^{\infty} 2e^{-nhc'}$$

Si $Q(n) = \frac{nh}{\log n}$ entonces $nh = Q(n)\log n$

$\log n$

luego $\quad 2\, e^{-c' Q(n) \log n} = 2\, e^{\log n^{-c'Q(n)}} = 2\, n^{-c'Q(n)}$

entonces si $Q(n) \to \infty$ resulta que la serie

$\sum_{n=1}^{\infty} 2\, n^{-c'Q(n)}$ converge o sea

$$\sum_{n=1}^{\infty} 2 e^{-nh c'} \quad \text{converge}$$

A partir de esto $f_n \xrightarrow{c} f$

# Regresión no Paramétrica

Se introduce en 1964 por Nadaraya y Watson

Nadaraya (1964) On Estimating Regression (Th, Prob, App)

Watson (1964) Smooth Regression Analysis (Sankhya)

Supongamos que tenemos $(X_1, Y_1) \ldots (X_n, Y_n)$ que son iid con densidad conjunta $f_{X,Y}(x,y)$

Consideremos el modelo $Y = g(X) + e$ con $g$ una función cualquiera que no depende necesariamente de algún parámetro

Entonces sea $g \in \mathcal{G}$, y $X$ y $e$ independientes con $E(e) = 0$

Nuestro objetivo es encontrar $E(Y/X) = g(X)$

Como se trata de una esperanza condicional hallamos la estimación de la densidad condicional

$$f_{Y/X}(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X,Y}(x,y)}{\int f_{X,Y}(x,y)\,dy}$$

así pues

$$g(x) = E(Y/X) = \frac{\int y\, f_{X,Y}(x,y)\,dy}{\int f_{X,Y}(x,y)\,dy}$$

Si podemos estimar $f_{X,Y}(x,y)$ entonces obtendríamos un estimador de $g(x)$.

Osea

$$\hat{g}(x) = \int y \frac{\hat{f}_{X,Y}(x,y)}{\int \hat{f}_{X,Y}(x,y)\,dy}\,dy$$

Si usamos el estimador por núcleo

$$\hat{f}_{X,Y}(x,y) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^2} K\left(\frac{x-X_i}{h}, \frac{y-Y_i}{h}\right)$$

Llamemos

$$K_h(x-X_i, y-Y_i) = \frac{1}{h^2} K\left(\frac{x-X_i}{h}, \frac{y-Y_i}{h}\right)$$

entonces

$$f_{X,Y}(x,y) = \frac{1}{n}\sum_{i=1} \delta_h(x-X_i, y-Y_i)$$

Como $\delta_h$ es un núcleo para estimar la densidad conjunta, esta sucesión de núcleos $\{\delta_h\}$ converge a la masa puntual en $(x,y)$

Tenemos además el núcleo marginal $\delta_h^{(m)}(x) = \int \delta_h(x,y)\, dy$

Pedimos que $\delta_h$ sea simétrica como función de $y$ o sea

$$\int y\, \delta_h(x,y)\, dy = 0$$

Además $\delta_h(x,y) \geq 0$ y $\iint \delta_h(x,y)\, dx\, dy = 1$

Ahora

$$\int \hat{f}_{X,Y}(x,y)\, dy = \int \frac{1}{n}\sum_{i=1}^{n} \delta_h(x-X_i, y-Y_i)\, dy =$$

$$\overset{\text{cv } u=y-Y_i}{=} \frac{1}{n}\sum_{i=1}^{n} \underbrace{\int \delta_h(x-X_i, u)\, du}_{\delta_h^{(m)}(x-X_i)} = \frac{1}{n}\sum_{i=1}^{n} \delta_h^{(m)}(x-X_i)$$

También

$$\frac{1}{n}\sum_{i=1}^{n} \int y\, \delta_h(x-X_i, y-Y_i)\, dy \overset{u=y-Y_i}{=} \frac{1}{n}\sum_{i=1}^{n} \int (u+Y_i)\, \delta_h(x-X_i, u)\, du$$

$$= \frac{1}{n}\sum^{n} \overbrace{\int u\, \delta_h(x-X_i, u)\, du}^{0 \text{ por hipótesis}} + \frac{1}{n}\sum^{n} \int Y_i\, \delta_h(x-X_i, u)\, du =$$

$$= \frac{1}{n}\sum_{i=1}^{n} \int u\, \delta_h(x-X_i, u)\,du + \frac{1}{n}\sum_{i=1}^{n} \int Y_i\, \delta_h(x-X_i, u)\,du =$$

$$= \frac{1}{n}\sum_{i=1}^{n} Y_i\, \delta_h^{(m)}(x-X_i)$$

Así pues

$$\hat{g}(x) = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i\, \delta_h^{(m)}(x-X_i)}{\frac{1}{n}\sum_{i=1}^{n} \delta_h^{(m)}(x-X_i)}$$

Como $\delta_h^{(m)}$ es un núcleo podemos entonces proponer como estimador

$$\hat{g}(x) = \frac{\sum_{i=1}^{n} Y_i\, K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)} = \sum_{i=1}^{n} w_i(x)\, Y_i$$

con $w_i(x) = \dfrac{K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)}$

Es claro que $\sum_{i=1}^{n} w_i(x) = 1$

Ver un trabajo en Annals Statistics (1977) de Stone.

Para que exista una esperanza condicional no tiene por qué haber densidades pero se pensó así.

Hay otra manera de ver esto

Supongamos que estamos en el caso discreto

Tengo $(Y_1, X_1) \ldots (Y_n, X_n)$.

Para calcular $F_{Y/X=x}$ veríamos cuales de los $X_i$ son $x$ y luego cuento los $Y_i$ es decir considero la empírica.
El problema puede ser que existen muchos $X_i$
Luego tomo esperanza y tengo una estimación de la esperanza condicional.

Como tengo el problema que pueden existir muchos $X_i$ que están cerca de $x$ y así obtengo la distribución condicional y los pesos $w_i(x)$

O sea tendría $\sum_{i=1}^{n} w_i(x) \mathbb{1}_{\{|X_i - x| > \varepsilon\}} \xrightarrow{P} 0$ es decir se concentra el peso alrededor de $x$.

Consistencia:

Veremos primero la demostración de Nadaraya y luego una más general.

Nadaraya supone que los $X \sim f_X$ es decir que tienen densidad.

Teorema 9.9: Si $X \sim f_X$, $f_X(x) > 0$. Sea $x$ un punto de continuidad de $f_X$. Supongamos que $h \to 0$, $nh \to \infty$, $\int K(u)\,du = 1$, $|x K(x)| \to 0$ si $|x| \to \infty$. Además $E(|Y|) < \infty$ y se cumple alguna de las dos condiciones

(I) $E(Y^2) < \infty$ y $nh^2 \to \infty$

(II) $E(Y^2 / X = x)$ es continua en $x$

entonces $\hat{g}(x)$ es un estimador consistente de $E(Y/X) = g(X)$

Dem: $$\hat{g}(x) = \frac{\frac{1}{nh}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h}\right)}{\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)} = \frac{R_{1n}(x)}{R_{0n}(x)}$$

Es claro que $R_{0n}(x) \xrightarrow{P} f(x)$ pues es un estimador de la densidad $f$

Consideremos ahora $R_{1n}(x)$

$$E\left(E\left(\tfrac{1}{h}K\left(\tfrac{x-X_1}{h}\right)Y_1/X_1\right)\right)$$

$$E(R_{1n}(x)) = E\left(\frac{1}{h}K\left(\frac{x-X_1}{h}\right)Y_1\right) = \frac{1}{h}E\left(K\left(\frac{x-X_1}{h}\right)\underbrace{E(Y_1/X_1)}_{g(X_1)}\right) =$$

$$= \frac{1}{h}\int K\left(\frac{x-t}{h}\right)g(t)f_X(t)\,dt \xrightarrow{\text{lema de Bochner extendido}} g(x)f(x)$$ pues $x$ es un punto de continuidad de $g \cdot f$. Hemos aplicado el lema de Bochner extendido a $g \cdot f$

Debemos ver que $g \cdot f \in L^1$

$$\int |g(t)| f_X(t)\,dt = E(|g(X)|) = E(|E(Y/X)|) \leq E(E(|Y|/X)) = E(|Y|)$$

o sea basta pedir que $E(|Y|) < \infty$

Usamos que

$$\left(E(E(Y/X)) = E(Y)\right)$$

$$E(f(X)/X=x) = f(x)$$

K cumple las hipótesis de Bochner extendido

Calculemos ahora la varianza de $R_{1n}(x)$

$$\mathrm{Var}\left(R_{1n}(x)\right) = \frac{1}{nh^2}\mathrm{Var}\left(K\left(\frac{x-X_1}{h}\right)Y_1\right) \leq \frac{1}{nh^2}E\left(Y_1^2 K^2\left(\frac{x-X_1}{h}\right)\right)$$

$$= \frac{1}{nh^2}E\left(K^2\left(\frac{x-X_1}{h}\right)E\left(Y_1^2/X_1\right)\right) = \frac{1}{nh^2}\int K^2\left(\frac{x-t}{h}\right)m_2(t)f(t)\,dt$$

donde hemos llamado $m_2(x) = E\left(Y^2/X\right)$

Ahora $E\left(m_2(X)\right) = E\left(E\left(Y^2/X\right)\right) = E(Y^2)$

Veamos que pasa con

$$\frac{1}{nh^2}\int K^2\left(\frac{x-t}{h}\right)m_2(t)f(t)\,dt \quad (1)$$

Caben 2 posibilidades

a) Si $E(Y^2) < \infty$ y $nh^2 \to \infty$ como $K$ es acotado se tiene que $(1) \Rightarrow 0$

b) Si pedimos

$$(1) = \frac{1}{nh}\,\frac{1}{h}\int K^2\left(\frac{x-t}{h}\right)m_2(t)f_X(t)\,dt \xrightarrow{\text{lema de Bochner}} \frac{1}{nh}m_2(x)f(x)\int K^2$$

Siendo $x$ punto de continuidad de $m_2 f$ o sea

debo pedir que $m_2$ sea continua en $x$ o sea pido que la esperanza condicional sea continua ($E(Y^2/X)$ continua en $x$)

En este caso $(1) \longrightarrow 0$ si $nh \to \infty$

Como $E(R_{1n}(x)) \longrightarrow g(x)\, f(x)$

y $Var(R_{1n}(x)) \longrightarrow 0$

se tiene que $R_{1n}(x) \xrightarrow{P} g(x)\, f(x)$

luego $\hat{g}(x) = \dfrac{R_{1n}(x)}{R_{0n}(x)} \xrightarrow{P} \dfrac{g(x)\cancel{f(x)}}{\cancel{f(x)}} = g(x)$

o sea $\hat{g}(x)$ es un estimador consistente de $g(x) = E(Y/X)$