Distribución asintótica para regresión

Teorema 100: Sean $\{(X_i, Y_i)\ i \geq 1\}$ variables aleatorias independientes idénticamente distribuidas con $X_i \in R^d$, $Y_i \in R$

Sea $F(y/X_1 = v)$ la distribución condicional de Y_1 dado $X_1 = v$

Notemos por: $g(v) = E(Y_1/X_1 = v)$ y $\sigma^2(v) = E\left((Y_1 - g(v))^2/X_1 = v\right)$

Supongamos que:

(i) g es Lipchitz ; $\lim_{\varepsilon \to 0} \frac{g(x+\varepsilon v) - g(x)}{\varepsilon} = g'(x, v)$

es decir g tiene derivada en la dirección de v

Es claro que si g es diferenciable se cumplen las 2 cosas

(ii) $\sigma^2(\cdot)$ es continua en x

(iii) $X_i \sim f$ continua en x y acotada

(iv) $K : R^d \to R$ acotado, no negativo, $\int K = 1$

$$|t|^d K(t) \xrightarrow[|t| \to \infty]{} 0 \qquad t^2 K^2(t) \text{ integrable}$$

v) $\exists\ 0 < \beta < \infty$ tal que $n h^{d+2} \to \beta$, $h \to 0$

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

entonces $(n h^d)^{1/2} (g_n(x) - g(x)) \xrightarrow{w} N\left(b, \frac{\sigma^2(x)}{f(x)} \int K^2(u)\, du\right)$

con $b = \beta^{1/2} \int K(t)\, g'(x, t)\, dt$

con $\quad b = \beta^{1/2} \int K(t)\, g'(x,t)\, dt$

Obs: Como $n h^{d+2} \to \beta$ entonces $n h^d h^2 \to \beta$ y como $h \to 0$ debe

ser que $n h^d \to \infty$

Dem:

$$(n h^d)^{1/2} \left[\frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)} - g(x) \right] =$$

$$= (n h^d)^{1/2} \left[\frac{\frac{1}{n h^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Y_i - g(x) \frac{1}{n h^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}{\frac{1}{n h^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)} \right]$$

Por iii), (iv) y (v) (ver obs $n h^d \to \infty$) se tiene que

$$\frac{1}{n h^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \xrightarrow{P} f(x)$$

Veamos el numerador.

$$(n h^d)^{1/2} \left[\frac{1}{n h^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Y_i - g(x) \frac{1}{n h^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \right]$$

$$= \frac{1}{(n h^d)^{1/2}} \left[\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \left(Y_i - g(x)\right) \right] =$$

$$= \frac{1}{(n h^d)^{1/2}} \left[\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \left(Y_i - g(X_i) + g(X_i) - g(x)\right) \right] =$$

$(nh^d)^{1/2} \Big[\sum_{i=1} \ldots \Big]$

$$= \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^{n} \overbrace{K\left(\frac{x-X_i}{h}\right)(Y_i - g(X_i))}^{V_{ni}} + \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)(g(X_i)-g(x))$$

Las variables V_{ni} son independientes y $E(V_{ni}) = 0$

Tenemos un sistema triangular. Si se verifican las condiciones de Lindeberg se puede aplicar el teorema central del límite.

Recordemoslas: Si $X_1, \ldots, X_n$ son v.a independientes con distribuciones $F_1, F_2, \ldots, F_n$ tales que $E(X_k) = 0$, $Var(X_k) = \sigma_k^2$ y sea

$$S_n^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$$

Supongamos que para cada $t > 0$

(i) $$\frac{1}{S_n^2} \sum_{k=1}^{n} \int_{|y| \geq t S_n} y^2 \, dF_k(y) \longrightarrow 0$$

(ii) $$\frac{1}{S_n^2} \sum_{k=1}^{n} \int_{|y| < t S_n} y^2 \, dF_k(y) \longrightarrow 1$$

entonces $$\frac{\sum_{i=1}^{n} X_i}{S_n} \xrightarrow{\mathcal{D}} N(0,1) \qquad n \to \infty$$

$$S_n^2 = n V(V_{ni}) = n E(V_{n1}^2) = n E\left(K^2\left(\frac{x-X_1}{h}\right)(Y_1 - g(X_1))^2\right) =$$

$$= n E\left(K^2\left(\frac{x-X_1}{h}\right) E\left((Y_1 - g(X_1))^2 / X_1\right)\right) =$$

$$= n \int K^2\left(\frac{x-t}{h}\right) \sigma^2(t) f(t)\, dt$$

Veamos que por Bochner $\frac{1}{h^d}\int K^2\left(\frac{x-t}{h}\right) \sigma^2(t) f(t)\, dt \to \sigma^2(x) f(x) \int K^2 \quad h \to 0$

luego $\quad s_n^2 = n h^d \frac{1}{h^d} \int K^2\left(\frac{x-t}{h}\right)\sigma^2(t) f(t)\, dt \to \infty$ si $h \to 0$, $n h^d \to \infty$

luego (ii)

$$\frac{1}{s_n^2}\, n \int_{|y| < t s_n} K^2\left(\frac{x-y}{h}\right) \sigma^2(y) f(y)\, dy =$$

$$= \frac{\cancel{n h^d}\ \frac{1}{h^d} \int_{|y|<t s_n} K^2\left(\frac{x-y}{h}\right) \sigma^2(y) f(y)\, dy}{\cancel{n h^d}\ \frac{1}{h^d} \int K^2\left(\frac{x-y}{h}\right) \sigma^2(y) f(y)\, dy}, \text{ como } s_n \to \infty$$

es claro que el cociente tiende a 1

La propiedad 1 sale análogamente.

Calculemos la varianza límite

$$\operatorname{Var}\left(\frac{1}{(n h^d)^{1/2}} \sum_{i=1}^n V_{ni}\right) = \frac{1}{\cancel{n} h^d} \cancel{n}\, E(V_{n}^2) \to \sigma^2(x) f(x) \int K^2$$

luego

$$\frac{1}{(n h^d)^{1/2}} \sum_{i=1}^n V_{ni} \longrightarrow N\left(0, \sigma^2(x) f(x) \int K^2\right)$$

Ahora

$$\frac{\frac{1}{(nh^d)^{1/2}} \sum_{i=1}^{n} V_{ni}}{\frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)} \longrightarrow N\left(0, \frac{\sigma^2(x) \int K^2}{f(x)}\right)$$

Consideremos ahora el término

$$\frac{1}{(nh^d)^{1/2}} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\left(g(X_i)-g(x)\right)$$

$$E\left(\frac{1}{(nh^d)^{1/2}} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\left(g(X_i)-g(x)\right)\right) =$$

$$= \frac{1}{(nh^d)^{1/2}}\, n\, \frac{h^d}{h^d} \int K\left(\frac{x-t}{h}\right)\left(g(t)-g(x)\right) f(t)\, dt = \quad \frac{x-t}{h}=u$$

$$= (nh^d)^{1/2} \int K(u)\left(g(x-uh)-g(x)\right) f(x-uh)\, du =$$

$$= (nh^d)^{1/2} \int K(u) \left(\frac{g(x-uh)-g(x)}{h}\right) f(x-uh)\, h\, du \quad (*)$$

Ahora $K(u) \frac{g(x-uh)-g(x)}{h} f(x-uh) \xrightarrow[h\to 0]{} K(u)\, g'(x,u)\, f(x)$

Como g es Lipchitz $|g(x-uh)-g(x)| \le C_1 |uh|$

Luego $\left| K(u) \frac{g(x-uh)-g(x)}{h} f(x-uh)\right| \le |K(u)|\, C_1 |u|\, C_2 = C|K(u)||u|$

luego $\left| K(u) \frac{g(x-uh)-g(x)}{h} f(x-uh) \right| \leq |K(u)| C_1 |u| C_2 = C |K(u)| |u|$

y como $u^2 K^2$ es integrable $|K(u)||u|$ es integrable

Podemos aplicar convergencia dominada luego

$$\int K(u) \frac{g(x-uh)-g(x)}{h} f(x-uh)\,du \longrightarrow f(x) \int K(t)\, g'(x,t)\,dt$$

$$\circledast \quad (nh^d)^{1/2} h \int K(u) \frac{g(x-uh)-g(x)}{h} f(x-uh)\,du \longrightarrow \beta^{1/2} f(x) \int K(t)\, g'(x,t)\,dt$$

pues $(nh^d)^{1/2} h = (n h^{d+2})^{1/2} = \beta^{1/2}$

$$\operatorname{Var}\left(\frac{1}{(nh^d)^{1/2}} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\left(g(X_i)-g(x)\right)\right) = \frac{1}{\cancel{n} h^d} \cancel{n} \operatorname{Var}\left(K\left(\frac{x-X_1}{h}\right)\left(g(X_1)-g(x)\right)\right)$$

$$\leq \frac{1}{h^d} E\left(K^2\left(\frac{x-X_1}{h}\right)\left(g(X_1)-g(x)\right)^2\right) =$$

$$= \frac{1}{h^d} \int K^2\left(\frac{x-t}{h}\right)\left(g(t)-g(x)\right)^2 f(t)\,dt =$$

$$= \frac{1}{h^d} \int K^2(u) \left(g(x-uh)-g(x)\right)^2 f(x-uh)\,du =$$

$$= h^2 \int K^2(u) \left(\frac{g(x-uh)-g(x)}{h}\right)^2 f(x-uh)\,du = \circledast\circledast$$

Por los mismos argumentos manejados antes podemos aplicar convergencia dominada y

$$\int K^2(u) \left(\frac{g(x-uh) - g(x)}{h} \right)^2 f(x-uh)\,du \longrightarrow f(x) \int K^2(u) \left(g'(x,u)\right)^2 du$$

luego $\ast\ast \xrightarrow[h \to 0]{} 0$

Entonces $\frac{1}{(nh^d)^{1/2}} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\left(g(X_i) - g(x)\right) \xrightarrow{P} \beta^{1/2} f(x) \int K(t)\, g'(x,t)\,dt$

A partir de los resultados anteriores aplicando Slutsky se obtiene el resultado.