Consistencia del estimador núcleo

Teorema: *Sean $X_1, X_2, \dots$ v.a.i.i.d. con distribución de probabilidad continua de densidad f. Supongamos que*

a) *El núcleo K es una densidad acotada tal que $|x|K(x) \xrightarrow{x\pm\infty} 0$.*

b) *El parámetro de suavizado satisface que $h = h_n \xrightarrow{n\to\infty} 0$ y $nh_n \xrightarrow{n\to\infty} \infty$.*

c) *La densidad f es acotada y continua en el punto t.*

Entonces el estimador kernel converge en probabilidad a la densidad poblacional en el punto t:

$$\hat{f}_n(t) \xrightarrow[n\to\infty]{P} f(t). \tag{1}$$

Para demostrar este teorema utilizamos el lema de Bochner, que enunciamos a continuación. Este lema establece condiciones bajo las que la convolución de una función (g) con una aproximación de la identidad dada por otra función K reescalada converge a la función g original.

Lema de Bochner: *Sea K una función integrable y acotada tal que $\int K = 1$ y $|x||K(x)| \to 0$ cuando $|x| \to \infty$. Sea g una función integrable. Entonces, para $K_h(t) := \frac{1}{h}K\left(\frac{t}{h}\right)$, tenemos que $g * K_h(x) \to g(x)$ en todo punto x de continuidad de g.*

Demostración del teorema: La convergencia (1) se da si y sólo si, para todo $\epsilon > 0$,

$$\mathbb{P}\{|\hat{f}_n(t) - f(t)| > \epsilon\} \xrightarrow{n\to\infty} 0.$$

Utilizando la desigualdad de Markov podemos acotar esta probabilidad de la manera siguiente:

$$\mathbb{P}\{|\hat{f}_n(t) - f(t)| > \epsilon\} = \mathbb{P}\{(\hat{f}_n(t) - f(t))^2 > \epsilon^2\} \leq \frac{\mathbb{E}((\hat{f}_n(t) - f(t))^2)}{\epsilon^2}.$$

El término $\mathbb{E}((\hat{f}_n(t) - f(t))^2)$ es el error cuadrático medio (ECM) del estimador núcleo en el punto t. Cualquier ECM de un estimador se puede descomponer como suma de la varianza del estimador y su sesgo al cuadrado:

$$\begin{aligned}
\mathbb{E}((\hat{f}_n(t) - f(t))^2) &= \mathbb{V}(\hat{f}_n(t)) + \text{Sesgo}^2(\hat{f}_n(t)) \\
&= \mathbb{E}((\hat{f}_n(t) - \mathbb{E}\hat{f}_n(t))^2) + (\mathbb{E}\hat{f}_n(t) - f(t))^2.
\end{aligned}$$

- Para que el término de sesgo tienda a 0 (cuando $n \to \infty$) necesitamos que $h = h_n \xrightarrow{n\to\infty} 0$:

$$\begin{aligned}
\text{Sesgo}(\hat{f}_n(t)) &= \mathbb{E}\hat{f}_n(t) - f(t) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} K_h(t - X_i)\right) - f(t) \\
&= \mathbb{E}(K_h(t - X)) - f(t) = \int_{\mathbb{R}} K_h(t-x)f(x)dx - f(t) \\
&= (K_h * f)(t) - f(t) \xrightarrow{h\to 0} 0 \quad \text{(por el lema de Bochner)}.
\end{aligned}$$

- Para que el término de varianza tienda a 0 (cuando $n \to \infty$) necesitamos que $nh_n \xrightarrow{n\to\infty} \infty$:

$$\begin{aligned}
\mathbb{V}(\hat{f}_n(t)) &= \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} K_h(t - X_i)\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}(K_h(t - X_i)) \\
&= \frac{1}{n}\mathbb{V}(K_h(t - X)) = \frac{1}{n}\{\mathbb{E}(K_h^2(t - X)) - \underbrace{\mathbb{E}^2(K_h(t - X))}_{\substack{\xrightarrow{h\to 0} f^2(t) \\ \text{(por lema de Bochner)}}}\}
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}(K_h^2(t-X)) &= \int K_h^2(t-x)f(x)dx = \frac{1}{h^2}\int K^2\left(\frac{t-x}{h}\right)f(x)dx \\
&= \frac{1}{h}\int \frac{1}{h}K^2\left(\frac{t-x}{h}\right)f(x)dx = \frac{1}{h}\int (K^2)_h(t-x)f(x)dx \\
&= \frac{1}{h}\,(K^2)_h * f(t) = \frac{\|K^2\|_1}{h}\quad \underbrace{\left(\frac{K^2}{\|K^2\|_1}\right)_h * f(t)}_{\substack{\xrightarrow{h\to 0} f(t) \\ \text{(por lema de Bochner)}}}
\end{aligned}
$$

Resumiendo

$$
\mathbb{V}(\hat{f}_n(t)) \quad = \quad \frac{\|K^2\|_1}{nh}\left(\frac{K^2}{\|K^2\|_1}\right)_h * f(t) + \frac{1}{n}\mathbb{E}^2(K_h(t-X)) \xrightarrow{h\to 0,\, nh_n\to 0} 0.
$$

$\square$