El error cuadrático medio integrado

Dada una muestra $X_1, \ldots, X_n$ se define el estimador naive, $f_{n,N}$, mediante

$$f_n(x) \equiv f_{n,N}(x, h, X_1, \ldots, X_n) = \frac{1}{2nh} \sum_{i=1}^{n} \mathbb{I}(x - h < Xi < x + h), \; x \in \mathbb{R},$$

donde $h > 0$ es un parámetro que debe elegir el investigador. Una posibilidad para ello es fijar un criterio de error y seleccionar aquel parámetro que haga que ese criterio alcance el mínimo. Uno de los criterios de error más empleados es el error cuadrático medio integrado, MISE, que se define como

$$MISE(h) = \mathbb{E}\left[\int (f_n(x) - f(x))^2 dx\right].$$

En clase vimos que asintóticamente el MISE se podía aproximar por

$$AMISE(h) = \frac{1}{2nh} + \frac{h^4}{36} \int (f''(x))^2 dx.$$

Vamos a comprobar la validez de esta aproximación mediante un estudio de simulación. Probaremos qué tal funciona en el modelo normal. Para aproximar el MISE mediante simulación sigue los siguientes pasos:

1. Genera $M = 1000$ muestras de la normal estándar

2. Evalúa, en cada una de ellas, el estimador Naive para distintos valores de h. Prueba con 60 valores equiespaciados comprendidos entre 0.05 y 3, separados por 0.05. Evalúa, para cada valor de h, el estimador en una rejilla de 601 puntos equiespaciados en el intervalo $[-3, 3]$, separados por una distancia 0.01.

3. Aproxima el error cuadrático integrado, ISE,

$$ISE(h) \equiv ISE(h, X_1, \ldots, X_n) = \int (f_n(x) - f(x))^2 dx.$$

4. Utiliza los valores en las M muestras del ISE para aproximar el $MISE$ en cada uno de los valores de h considerados.

5. ¿En qué valor de h el AMISE alcanza el mínimo?¿Y la curva obtenida por simulación?

6. Dibuja la curva obtenida y compárala con la curva correspondiente del AMISE, para el modelo normal. ¿Es la aproximación buena en todo el rango de valores de h?

7. Calcula teóricamente el

$$\lim_{h \to \infty} MISE(h)$$

y

$$\lim_{h \to \infty} AMISE(h).$$