Sea $\hat{f}_n$ un estimador kernel de la densidad basado en un núcleo K que es una función de densidad con media finita. Comprobar que, en general, $\hat{f}_n(t)$ es un estimador sesgado de $f(t)$ en el sentido de que NO se tiene $\mathbb{E}(\hat{f}_n(t)) = f(t)$, para todo t y para toda densidad f.

Solución: La expresión del estimador núcleo con núcleo K y parámetro de suavizado $h = h_n$ evaluado en el punto t es

$$\hat{f}_n(t) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h}\right).$$

Observemos que $h = h_n$ quiere decir que, cuando el tamaño muestral n cambia, el valor del parámetro de suavizado también, es decir tenemos una *sucesión* de parámetros de suavizado. Para estudiar el sesgo del estimador núcleo, calculamos su valor esperado:

$$\begin{aligned}
\mathbb{E}\hat{f}_n(t) &= \mathbb{E}\left[\frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h}\right)\right] = \frac{1}{h_n}\mathbb{E}\left[K\left(\frac{t - X_i}{h}\right)\right] \\
&= \frac{1}{h_n}\int_{-\infty}^{\infty} K\left(\frac{t - x}{h}\right) f(x)dx = \int_{-\infty}^{\infty} K(z)f(t - hz)dz.
\end{aligned}$$

Ahora escribimos la expresión del sesgo. Utilizamos que, como K es una función de densidad, $\int K = 1$.

$$\begin{aligned}
\text{Sesgo}(\hat{f}_n(t)) &= \mathbb{E}\hat{f}_n(t) - f(t) = \int_{-\infty}^{\infty} K(z)f(t - hz)dz - \int_{-\infty}^{\infty} K(z)f(t)dz \\
&= \int_{-\infty}^{\infty} K(z)(f(t - hz) - f(t))dz
\end{aligned}$$

Observemos que, como K es una densidad, se cumple que $K(z) \geq 0$ para todo z. Por tanto, para que el sesgo fuera 0, en el conjunto donde $K > 0$, tendría que suceder que $f(t - hz) - f(t) = 0$, es decir, que en un entorno del punto t los valores de la densidad fueran iguales a $f(t)$ (que la densidad sea constante en un pequeño entorno de t). Y esta es una hipótesis muy restrictiva, porque además normalmente interesa estimar la densidad $f(t)$ en todo t.
Por si contribuye a aclarar ideas vamos a desarrollar la expresión del sesgo un poco más. Emplearemos el desarrollo en serie de Taylor de la densidad f (suponiendo que f cumple unas mínimas condiciones de diferenciabilidad): $f(t - hz) - f(t) = f'(t)(-hz) + f''(t)(-hz)^2/2 + f'''(t)(-hz)^3/3! + \ldots$

$$\text{Sesgo}(\hat{f}_n(t)) = -hf'(t)\int_{-\infty}^{\infty} zK(z)dz + \frac{1}{2}h^2 f''(t)\int_{-\infty}^{\infty} z^2K(z)dz - \frac{1}{3}h^3 f'''(t)\int_{-\infty}^{\infty} z^3K(z)dz + \ldots$$

Y ahora supongamos que K es una función par, por lo que $\int z^j K(z)dz = 0$ para todo j impar. Entonces

$$\text{Sesgo}(\hat{f}_n(t)) = \frac{1}{2}h^2 f''(t)\int_{-\infty}^{\infty} z^2K(z)dz + \text{términos de orden } h^4 \text{ o más}$$

Lo que vemos es que es imposible que el sesgo sea 0 salvo que todas las derivadas de orden par de f en t sean 0. Pero sí podemos hacer que el sesgo sea tan pequeño como queramos a medida que el tamaño muestral aumenta utilizando un parámetro de suavizado h tal que $h_n \to 0$ cuando $n \to \infty$. Esta suposición sobre el parámetro h es estándar en estadística no paramétrica.