

Índice

1. Introducción y Motivación	1
1.1. Contexto y Justificación	1
1.2. Motivación Práctica	1
1.3. Objetivos y Estructura del Trabajo	2
2. Preliminares Probabilísticos	4
2.1. Fundamentos de Probabilidad	4
2.2. Momentos	5
2.3. Convergencia Estocástica y Notación de Orden	7
2.3.1. Modos de Convergencia	7
2.3.2. Notación de Orden Estocástica (o_p y O_p)	9
2.4. Herramientas de Análisis y Aproximación	10
2.4.1. Clases de Suavidad	10
2.4.2. Desarrollo de Taylor	10
2.4.3. Convolución	11
3. El Estimador de Densidad Tipo Núcleo	12
3.1. Motivación: Del Histograma al <i>Naive Estimator</i>	12
3.1.1. Ilustración Gráfica	13
3.2. Definición Formal y Propiedades del Núcleo	14
3.2.1. El Estimador de Rosenblatt-Parzen	14
3.2.2. Clase de Funciones Admisibles	14
3.2.3. Herencia de Propiedades de Suavidad	16
3.2.4. Nota sobre la Generalización Multivariante	16
3.3. Tipos de Funciones Núcleo	16
3.3.1. 1. Núcleo de Epanechnikov	17
3.3.2. 2. Núcleo Gaussiano (Normal)	17
3.3.3. 3. Núcleo Uniforme (Rectangular)	18
3.3.4. 4. Núcleo Triangular	18
3.3.5. Resumen de Constantes del Núcleo	18
3.3.6. Comparación Visual del Efecto del Núcleo	19
3.4. El Papel del Ancho de Banda (h)	21
3.4.1. Formulación Matemática del Compromiso Sesgo-Varianza	21
3.4.2. Ilustración del Fenómeno	22

1. Introducción y Motivación

1.1. Contexto y Justificación

La inferencia estadística constituye el pilar fundamental para la extracción de conocimiento a partir de datos observados. En el contexto del Máster en Estadística Aplicada de la Universidad de Granada, este trabajo aborda uno de los problemas centrales de la disciplina: la estimación de la función de densidad de probabilidad subyacente a una variable aleatoria, sin imponer restricciones fuertes sobre su forma funcional.

Tradicionalmente, la estadística clásica ha abordado este problema desde un enfoque **paramétrico**. Bajo este paradigma, se asume que los datos provienen de una familia conocida de distribuciones $\mathcal{P} = \{f(x; \theta) : \theta \in \Theta\}$, como la distribución Normal, Gamma o Weibull. En tal escenario, el problema de estimación se reduce drásticamente a la inferencia de un vector de parámetros finito-dimensional θ . Si bien este enfoque es potente y eficiente cuando el modelo asumido es correcto, presenta una debilidad crítica: la **rigidez**. La realidad de los fenómenos estocásticos complejos raramente se ajusta con precisión a estas formas ideales. La imposición de una estructura paramétrica incorrecta conlleva un *sesgo de especificación* sistemático que puede invalidar cualquier análisis posterior, ocultando características locales importantes.

Como respuesta a esta limitación, surge la **estadística no paramétrica** y, específicamente, los métodos de suavizado. La filosofía subyacente es permitir que los datos “hablen por sí mismos” (*data-driven approach*), dejando que sea la propia estructura de la muestra la que determine la forma de la densidad estimada \hat{f} .

Históricamente, el **histograma** ha sido la herramienta precursora en esta tarea. A pesar de su ubicuidad, el histograma padece de defectos teóricos y prácticos notables:

1. Es una función discontinua, lo que impide el uso de herramientas de cálculo diferencial.
2. Su forma depende excesivamente de la elección del origen de la partición y del ancho de los intervalos.
3. Carece de propiedades de eficiencia en dimensiones altas.

Para solventar estas deficiencias, Rosenblatt (1956) y Parzen (1962) formalizaron el **Estimador de Densidad Tipo Núcleo** (Kernel Density Estimator o KDE). Este estimador generaliza la idea del histograma colocando una “masa de probabilidad” suave (el núcleo) sobre cada observación, proporcionando una estimación diferenciable e integrable a uno.

1.2. Motivación Práctica

La relevancia de estudiar el estimador KDE trasciende el interés puramente teórico; su aplicación es crítica en escenarios donde detectar desviaciones sutiles respecto a la normalidad es vital para la gestión de riesgos. Este trabajo toma como caso de estudio central el índice **Standard &**

Poor's 500 (S&P 500), considerado el barómetro más representativo del mercado bursátil global.

Una de las “hechos estilizados” más documentados en la econometría financiera es que la distribución de los rendimientos de los activos **no es Normal**: presenta colas más pesadas (*leptocurtosis*) y asimetrías que los modelos gaussianos subestiman sistemáticamente. Ignorar estas características, especialmente durante periodos de crisis financiera, lleva a una infravaloración del riesgo de eventos extremos (cisnes negros).

El uso de la Estimación de Densidad Tipo Núcleo permite capturar la forma real de la distribución de los rendimientos del S&P 500 sin supuestos previos. Esto es fundamental para:

1. **Cálculo de métricas de riesgo:** Obtener estimaciones más precisas del Valor en Riesgo (VaR) y del *Expected Shortfall* en las colas de la distribución.
2. **Detección de regímenes de mercado:** Visualizar cómo la densidad de los rendimientos cambia de forma bimodal o se aplanan en periodos de alta volatilidad, comportamientos que un modelo paramétrico estático no podría revelar.

La aplicación práctica de este trabajo (Capítulo 7) demostrará cómo el KDE ofrece una radiografía superior del comportamiento del mercado en comparación con el ajuste de una curva de campana tradicional.

1.3. Objetivos y Estructura del Trabajo

El objetivo principal de este Trabajo Fin de Máster es presentar una revisión teórica rigurosa, exhaustiva y autocontenida de la estimación de densidad tipo núcleo, analizando sus propiedades asintóticas, los desafíos de la selección del ancho de banda y su aplicación a datos financieros reales.

Para garantizar una exposición clara y detallada, la memoria se estructura en los siguientes capítulos:

- **Capítulo 2:** Se establecen los **preliminares probabilísticos**, fijando la notación y recordando definiciones clave sobre convergencia estocástica y funciones de distribución.
- **Capítulo 3:** Se dedica a la construcción formal del **Estimador Núcleo Univariante**, definiendo los requisitos analíticos de la función núcleo K y presentando las familias de núcleos más usuales.
- **Capítulo 4:** Aborda las **propiedades estadísticas** del estimador. Se descompone el error de estimación (MISE) en sesgo y varianza, demostrando el compromiso (*trade-off*) necesario.
- **Capítulo 5:** Trata el problema central de la práctica del KDE: la **selección del ancho de banda** (h). Se discuten métodos como la Regla de Silverman y la Validación Cruzada (LSCV).
- **Capítulo 6:** Extiende los conceptos al caso **multivariante**, discutiendo brevemente la “maldición de la dimensionalidad”.

- **Capítulo 7:** Presenta la **aplicación práctica** al índice S&P 500. Se realizará un análisis exploratorio y se aplicarán los estimadores desarrollados teóricamente para modelar la distribución de sus rendimientos, comparando los resultados obtenidos mediante diferentes selecciones de ancho de banda.
- Finalmente, se presentan las **conclusiones** y líneas de trabajo futuro.

2. Preliminares Probabilísticos

El objetivo de este capítulo es establecer el marco matemático y la notación que se utilizarán a lo largo del trabajo. Dado que la estimación de densidad tipo núcleo se fundamenta en propiedades de convergencia de sumas de variables aleatorias, es necesario definir con rigor los conceptos de espacio de probabilidad, variable aleatoria y los modos de convergencia estocástica.

2.1. Fundamentos de Probabilidad

Comenzamos definiendo la estructura formal donde se modelan los experimentos aleatorios.

Definición 2.1 (Espacio de Probabilidad).

Un espacio de probabilidad es una terna $(\Omega, \mathcal{F}, \mathbb{P})$, donde:

1. Ω es el **espacio muestral**, un conjunto no vacío que contiene todos los posibles resultados del experimento.
2. \mathcal{F} es una **σ -álgebra** sobre Ω , es decir, una colección de subconjuntos de Ω que cumple:
 - $\Omega \in \mathcal{F}$.
 - Si $A \in \mathcal{F}$, entonces su complemento $A^c \in \mathcal{F}$.
 - Si $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$, entonces $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.
3. $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ es una **medida de probabilidad**, que satisface $\mathbb{P}(\Omega) = 1$ y la propiedad de σ -aditividad para conjuntos disjuntos.

En el contexto de la inferencia estadística, trabajamos con funciones que trasladan la aleatoriedad desde el espacio abstracto Ω a los números reales.

Definición 2.2 (Variable Aleatoria Real).

Dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, una función $X : \Omega \rightarrow \mathbb{R}$ se dice que es una **variable aleatoria** (v.a.) si es \mathcal{F} -medible; es decir, si para todo conjunto de Borel $B \in \mathcal{B}(\mathbb{R})$ (la σ -álgebra generada por los intervalos abiertos de \mathbb{R}), se cumple que la preimagen $X^{-1}(B) \in \mathcal{F}$.

Cada variable aleatoria induce una medida de probabilidad sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, caracterizada unívocamente por su función de distribución acumulada.

Definición 2.3 (Función de Distribución).

La función de distribución (CDF, por sus siglas en inglés) de una variable aleatoria X se define como la función $F : \mathbb{R} \rightarrow [0, 1]$ tal que:

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\}), \quad \forall x \in \mathbb{R}$$

Esta función es no decreciente, continua por la derecha y satisface $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$.

En este trabajo nos centraremos exclusivamente en variables aleatorias **absolutamente continuas**. Formalmente, esto significa que la medida inducida por X es absolutamente continua respecto a la medida de Lebesgue μ en \mathbb{R} . Por el Teorema de Radon-Nikodym, esto garantiza la existencia de una función de densidad.

Definición 2.4 (Función de Densidad de Probabilidad).

Se dice que una variable aleatoria X tiene una función de densidad de probabilidad (PDF) f , si existe una función no negativa e integrable tal que para todo conjunto de Borel $B \in \mathcal{B}(\mathbb{R})$:

$$\mathbb{P}(X \in B) = \int_B f(t) dt$$

Como consecuencia inmediata, la relación entre la función de distribución y la densidad viene dada por:

$$F(x) = \int_{-\infty}^x f(t) dt$$

Además, si f es continua en x , entonces por el Teorema Fundamental del Cálculo se cumple que:

$$f(x) = F'(x) = \frac{d}{dx} F(x)$$

El objetivo central de la estimación de densidad no paramétrica es estimar esta función f desconocida a partir de una muestra aleatoria simple X_1, \dots, X_n distribuida según F .

2.2. Momentos

Una vez definida la densidad de probabilidad, las características numéricas más relevantes de una variable aleatoria son sus momentos. Estos valores resumen la información contenida en la densidad, describiendo aspectos como la localización, la dispersión o la forma de la distribución. En el contexto de la estimación tipo núcleo, el cálculo de esperanzas y varianzas de transformaciones de la variable aleatoria constituirá la herramienta analítica principal para evaluar la calidad del estimador \hat{f} .

Definición 2.5 (Esperanza Matemática).

Sea X una variable aleatoria absolutamente continua con función de densidad f . Se define la **esperanza matemática** (o valor esperado) de X , denotada por $\mathbb{E}[X]$ o μ , como la integral de Lebesgue:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

siempre que la integral sea absolutamente convergente, es decir, $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$. Si esta

condición no se cumple, la esperanza no existe.

La esperanza es un operador lineal. Dadas dos variables aleatorias X, Y definidas sobre el mismo espacio de probabilidad y constantes $a, b \in \mathbb{R}$, se satisface:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

Para el desarrollo de las propiedades del estimador núcleo, raramente trabajaremos directamente con $\mathbb{E}[X]$, sino con la esperanza de funciones de X (específicamente, el núcleo K evaluado en los datos). El siguiente resultado permite calcular dicha esperanza utilizando directamente la densidad original f .

Teorema 2.1 (Esperanza de una función de una v.a.).

Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función medible Borel. Si X es una variable aleatoria con densidad f , entonces:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) dx$$

siempre que $\mathbb{E}[|g(X)|] < \infty$.

Este resultado es crucial en este trabajo. Por ejemplo, al calcular el sesgo del estimador, evaluaremos la esperanza de $g(X_i) = \frac{1}{h}K\left(\frac{x-X_i}{h}\right)$, lo que se traducirá directamente en una integral de convolución entre el núcleo y la densidad verdadera.

Definición 2.6 (Varianza y Momentos Superiores).

La **varianza** de una variable aleatoria X , denotada por $\mathbb{V}(X)$ o σ^2 , es una medida de dispersión cuadrática en torno a la media. Se define como el segundo momento central:

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Desarrollando el cuadrado, se obtiene la expresión operativa habitual:

$$\mathbb{V}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

De forma general, definimos el **k -ésimo momento** respecto al origen como $\mu'_k = \mathbb{E}[X^k]$, y el **k -ésimo momento central** como $\mu_k = \mathbb{E}[(X - \mu)^k]$. La varianza corresponde a μ_2 . La existencia de estos momentos depende del decaimiento de las colas de la densidad f ; distribuciones de cola pesada (como las observadas en finanzas) pueden carecer de momentos de orden superior.

Propiedades de la Varianza.

1. **No negatividad:** $\mathbb{V}(X) \geq 0$.
2. **Invarianza por traslación y escala:** Para constantes $a, b \in \mathbb{R}$, $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$.

3. **Aditividad:** Si X_1, \dots, X_n son variables aleatorias independientes dos a dos, entonces la varianza de la suma es la suma de las varianzas:

$$\mathbb{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}(X_i)$$

Esta última propiedad es fundamental para derivar la varianza del estimador KDE, ya que este se construye como una suma lineal de variables aleatorias independientes.

Teorema 2.2 (Cambio de Variable en Integración).

Dado que el análisis asintótico del ancho de banda h implica reescalar la variable aleatoria, recordamos la fórmula de cambio de variable, esencial para las demostraciones del Capítulo 4. Sea $\phi : \mathbb{R} \rightarrow \mathbb{R}$ un difeomorfismo. Entonces:

$$\int_{\mathbb{R}} f(\phi(u)) |\phi'(u)| du = \int_{\mathbb{R}} f(x) dx$$

En particular, para la transformación lineal $x = \mu + hu$ (con $h > 0$), tenemos $dx = h du$. Esto nos permite reescribir integrales de convolución de la forma:

$$\int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-y}{h}\right) f(y) dy = \int_{-\infty}^{\infty} K(u) f(x-hu) du$$

Esta identidad convierte el parámetro de escala h en un argumento de la función f , permitiendo posteriormente aplicar desarrollos de Taylor para analizar el comportamiento local del estimador cuando $h \rightarrow 0$.

2.3. Convergencia Estocástica y Notación de Orden

El comportamiento del estimador de densidad tipo núcleo depende intrínsecamente del tamaño de la muestra n . Para analizar sus propiedades asintóticas, es decir, cuando $n \rightarrow \infty$, es imprescindible formalizar en qué sentido las aproximaciones estocásticas se acercan a los valores teóricos poblacionales.

Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias y X una variable aleatoria, todas ellas definidas sobre un mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$.

2.3.1. Modos de Convergencia

Distinguimos tres modos principales de convergencia, ordenados de mayor a menor exigencia sobre la estructura probabilística subyacente.

Definición 2.7 (Convergencia Casi Segura).

Se dice que la sucesión $\{X_n\}$ converge **casi seguramente** (o con probabilidad 1) a X , denotado

por $X_n \xrightarrow{c.s.} X$, si el conjunto de eventos donde la sucesión no converge tiene medida nula:

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

Esta es la forma más fuerte de convergencia clásica. En el contexto de estimación, si un estimador $\hat{\theta}_n$ satisface $\hat{\theta}_n \xrightarrow{c.s.} \theta$, se dice que es **fuertemente consistente**.

Definición 2.8 (Convergencia en Probabilidad).

La sucesión $\{X_n\}$ converge **en probabilidad** a X , denotado por $X_n \xrightarrow{p} X$, si para todo $\epsilon > 0$, la probabilidad de que la distancia entre X_n y X exceda ϵ tiende a cero:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$$

Esta convergencia implica la **consistencia débil** de un estimador. Es una condición más relajada que la casi segura, permitiendo que existan eventos raros donde la discrepancia sea grande, siempre que su probabilidad se desvanezca asintóticamente.

Definición 2.9 (Convergencia en Distribución).

Sean F_n y F las funciones de distribución acumuladas de X_n y X respectivamente. Se dice que X_n converge **en distribución** (o en ley) a X , denotado por $X_n \xrightarrow{d} X$, si:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

para todo punto $x \in \mathbb{R}$ donde F sea continua.

A diferencia de las anteriores, la convergencia en distribución no requiere que las variables estén definidas en el mismo espacio de probabilidad, ya que se refiere únicamente a sus leyes. Es fundamental para la construcción de intervalos de confianza asintóticos mediante el Teorema Central del Límite.

Teorema 2.3 (Jerarquía y Relaciones).

Las implicaciones entre los modos de convergencia son las siguientes:

$$X_n \xrightarrow{c.s.} X \implies X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X$$

Las implicaciones inversas no son ciertas en general, salvo en el caso particular en que X es una constante degenerada ($X = c$), donde $X_n \xrightarrow{d} c \implies X_n \xrightarrow{p} c$.

Para operar algebraicamente con límites estocásticos, enunciamos dos resultados clásicos que utilizaremos para simplificar las expresiones del error asintótico.

Teorema 2.4 (Teorema de la Aplicación Continua y Slutsky).

Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función continua.

1. Si $X_n \xrightarrow{p} X$, entonces $g(X_n) \xrightarrow{p} g(X)$.
2. Si $X_n \xrightarrow{d} X$, entonces $g(X_n) \xrightarrow{d} g(X)$.
3. **(Lema de Slutsky):** Si $X_n \xrightarrow{d} X$ e $Y_n \xrightarrow{p} c$, donde c es una constante, entonces:

$$X_n + Y_n \xrightarrow{d} X + c, \quad X_n Y_n \xrightarrow{d} cX, \quad \text{y} \quad X_n / Y_n \xrightarrow{d} X/c \quad (\text{si } c \neq 0)$$

2.3.2. Notación de Orden Estocástica (o_p y O_p)

En el análisis asintótico del KDE, descompondremos el error en términos principales y términos residuales que se desvanecen rápidamente. Para formalizar la “velocidad” a la que estos términos convergen, generalizamos la notación de Landau (o y O) al contexto probabilístico.

Sea $\{X_n\}$ una sucesión de variables aleatorias y $\{a_n\}$ una sucesión de constantes positivas reales.

Definición 2.10 (Pequeña “o” en probabilidad).

Escribimos $X_n = o_p(a_n)$ si la razón X_n/a_n converge a cero en probabilidad:

$$X_n = o_p(a_n) \iff \frac{X_n}{a_n} \xrightarrow{p} 0$$

Intuitivamente, esto significa que X_n es de un orden asintótico estrictamente menor que a_n . Un caso particular importante es $X_n = o_p(1)$, que es equivalente a decir que $X_n \xrightarrow{p} 0$.

Definición 2.11 (Grande “O” en probabilidad).

Escribimos $X_n = O_p(a_n)$ si la razón X_n/a_n está **acotada en probabilidad**. Formalmente, esto significa que para todo $\epsilon > 0$, existe una constante finita $M_\epsilon > 0$ y un entero N_ϵ tales que:

$$\mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M_\epsilon\right) < \epsilon, \quad \forall n \geq N_\epsilon$$

Si $X_n = O_p(1)$, decimos que la sucesión es **estocásticamente acotada**. Cabe destacar que toda sucesión convergente en distribución es $O_p(1)$ (Teorema de Prohorov), lo que implica que $X_n \xrightarrow{d} X \implies X_n = O_p(1)$.

Álgebra de órdenes estocásticos.

Estas notaciones simplifican enormemente las demostraciones al permitirnos absorber términos despreciables. Algunas reglas aritméticas que utilizaremos (análogas al caso determinista) son:

1. $o_p(1) + o_p(1) = o_p(1)$
2. $O_p(1) + o_p(1) = O_p(1)$
3. $O_p(1) \cdot o_p(1) = o_p(1)$
4. Si $X_n = O_p(n^{-\alpha})$ y $Y_n = O_p(n^{-\beta})$, entonces $X_n Y_n = O_p(n^{-(\alpha+\beta)})$.

Gracias a esta notación, en el Capítulo 4 podremos expresar, por ejemplo, que la varianza del estimador decae a una velocidad específica, escribiendo $\mathbb{V}[\hat{f}(x)] = \frac{1}{nh}f(x)R(K) + o_p\left(\frac{1}{nh}\right)$.

2.4. Herramientas de Análisis y Aproximación

El análisis de las propiedades del estimador núcleo se basa fundamentalmente en aproximaciones locales de la función de densidad desconocida f . Para garantizar la validez de estas aproximaciones, es necesario restringir el espacio de funciones admisibles y disponer de herramientas de cálculo para cuantificar el error de aproximación.

2.4.1. Clases de Suavidad

La velocidad de convergencia del estimador KDE depende directamente de la “suavidad” de la densidad subyacente. Formalizamos este concepto mediante las clases de diferenciabilidad.

Definición 2.12 (Espacio \mathcal{C}^k).

Sea $I \subseteq \mathbb{R}$ un intervalo abierto. Definimos $\mathcal{C}^k(I)$ como el conjunto de funciones $g : I \rightarrow \mathbb{R}$ que tienen k derivadas continuas en I . En particular, asumiremos habitualmente que la densidad $f \in \mathcal{C}^2(\mathbb{R})$ y que su segunda derivada f'' es acotada y continuo-cuadrada integrable. Esta asunción de regularidad es necesaria para asegurar que la curvatura de la densidad está bien definida en todo punto.

2.4.2. Desarrollo de Taylor

La herramienta principal para evaluar el sesgo del estimador será la expansión de Taylor. Dado que el estimador núcleo evalúa la densidad en puntos vecinos $x - hu$, necesitaremos aproximar $f(x - hu)$ en términos de $f(x)$ cuando el ancho de banda h tiende a cero.

Teorema 2.5 (Teorema de Taylor con resto de Peano).

Sea $g \in \mathcal{C}^k(\mathbb{R})$ definida en un entorno del punto x . Para un incremento t suficientemente pequeño, se cumple:

$$g(x + t) = g(x) + g'(x)t + \frac{g''(x)}{2!}t^2 + \cdots + \frac{g^{(k)}(x)}{k!}t^k + o(|t|^k)$$

donde el término residual $o(|t|^k)$ satisface $\lim_{t \rightarrow 0} \frac{o(|t|^k)}{|t|^k} = 0$.

En el contexto del KDE, aplicaremos este teorema con $t = -hu$. Si $f \in \mathcal{C}^2$, la expansión hasta segundo orden resulta:

$$f(x - hu) = f(x) - huf'(x) + \frac{1}{2}h^2u^2f''(x) + o(h^2)$$

Esta expresión es la llave maestra que nos permitirá demostrar en el Capítulo 4 que el sesgo del estimador es de orden cuadrático ($O(h^2)$) respecto al ancho de banda.

2.4.3. Convolución

Finalmente, observamos que el estimador de densidad puede interpretarse matemáticamente como una operación de convolución.

Definición 2.13 (Convolución).

Dadas dos funciones integrables $f, g : \mathbb{R} \rightarrow \mathbb{R}$, se define su **convolución**, denotada por $(f * g)(x)$, como:

$$(f * g)(x) = \int_{-\infty}^{\infty} f(y)g(x - y) dy$$

Como veremos en el siguiente capítulo, el estimador de densidad es esencialmente la convolución de la función de densidad empírica con la función núcleo reescalada. Esta perspectiva permite utilizar propiedades del análisis armónico (como la Transformada de Fourier) para estudiar la eficiencia del estimador, aunque en este trabajo nos centraremos en el enfoque algebraico directo.

3. El Estimador de Densidad Tipo Núcleo

3.1. Motivación: Del Histograma al *Naive Estimator*

La construcción del estimador de densidad tipo núcleo surge como una evolución natural necesaria para superar las deficiencias analíticas del histograma clásico. Si bien el histograma es la herramienta exploratoria más universal, su formulación matemática presenta limitaciones que lo inhabilitan para la inferencia avanzada.

Consideremos una muestra aleatoria X_1, \dots, X_n de una variable con densidad f . El histograma se construye definiendo una partición del soporte en intervalos (*bins*) B_j de longitud h , y contando la frecuencia de observaciones en cada uno. Formalmente, para un $x \in B_j$, el estimador histograma es:

$$\hat{f}_H(x) = \frac{1}{nh} \sum_{i=1}^n \mathbb{I}(X_i \in B_j)$$

Este estimador presenta dos problemas fundamentales:

1. **Discontinuidad:** \hat{f}_H es una función escalonada, con discontinuidades en los bordes de los intervalos. Esto hace imposible obtener derivadas de la densidad estimada, impidiendo el análisis de puntos de inflexión o la curvatura.
2. **Dependencia de la Rejilla:** La forma del histograma cambia drásticamente dependiendo de la elección del punto de origen de la partición x_0 . Dos estadísticos con los mismos datos podrían llegar a conclusiones visuales distintas simplemente por empezar los intervalos en puntos diferentes.

Para eliminar la arbitrariedad de la rejilla, Rosenblatt (1956) propuso liberar a los intervalos de su posición fija. La idea intuitiva es: en lugar de forzar a x a caer en una caja predefinida, construyamos una caja de ancho $2h$ centrada exactamente en cada x .

Matemáticamente, esto se justifica recordando la definición de densidad como la derivada de la función de distribución F :

$$f(x) = F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$

Si sustituimos la distribución teórica F por la **Función de Distribución Empírica** $F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x)$, obtenemos el llamado *Naive Estimator*:

$$\hat{f}_{Naive}(x) = \frac{F_n(x+h) - F_n(x-h)}{2h} = \frac{1}{2nh} \#\{X_i : x-h < X_i \leq x+h\}$$

Podemos reescribir este estimador introduciendo una función de peso o “función núcleo” auxiliar

$w(u)$:

$$\hat{f}_{Naive}(x) = \frac{1}{nh} \sum_{i=1}^n w\left(\frac{x - X_i}{h}\right), \quad \text{donde } w(u) = \frac{1}{2} \mathbb{I}(|u| \leq 1)$$

El Naive Estimator resuelve el problema de la dependencia del origen, pero **no resuelve el problema de la suavidad**. Dado que la función $w(u)$ es rectangular (vale 1/2 dentro del intervalo y 0 fuera), el estimador resultante \hat{f}_{Naive} sigue siendo una función escalonada, presentando saltos discontinuos cada vez que una observación entra o sale de la ventana deslizante $(x - h, x + h)$.

Para obtener una estimación que sea suave y diferenciable (clase \mathcal{C}^∞), es necesario reemplazar la función rectangular $w(\cdot)$ por una función suave $K(\cdot)$ que decaiga gradualmente. Esta generalización da lugar al Estimador de Densidad Tipo Núcleo.

3.1.1. Ilustración Gráfica

A continuación, ilustramos visualmente esta transición. La Figura 3.1 muestra cómo el histograma depende de los intervalos fijos, mientras que el naive estimator centra las cajas en los datos pero mantiene la rugosidad. Finalmente, el estimador núcleo (con núcleo Gaussiano) recupera la suavidad de la densidad subyacente.

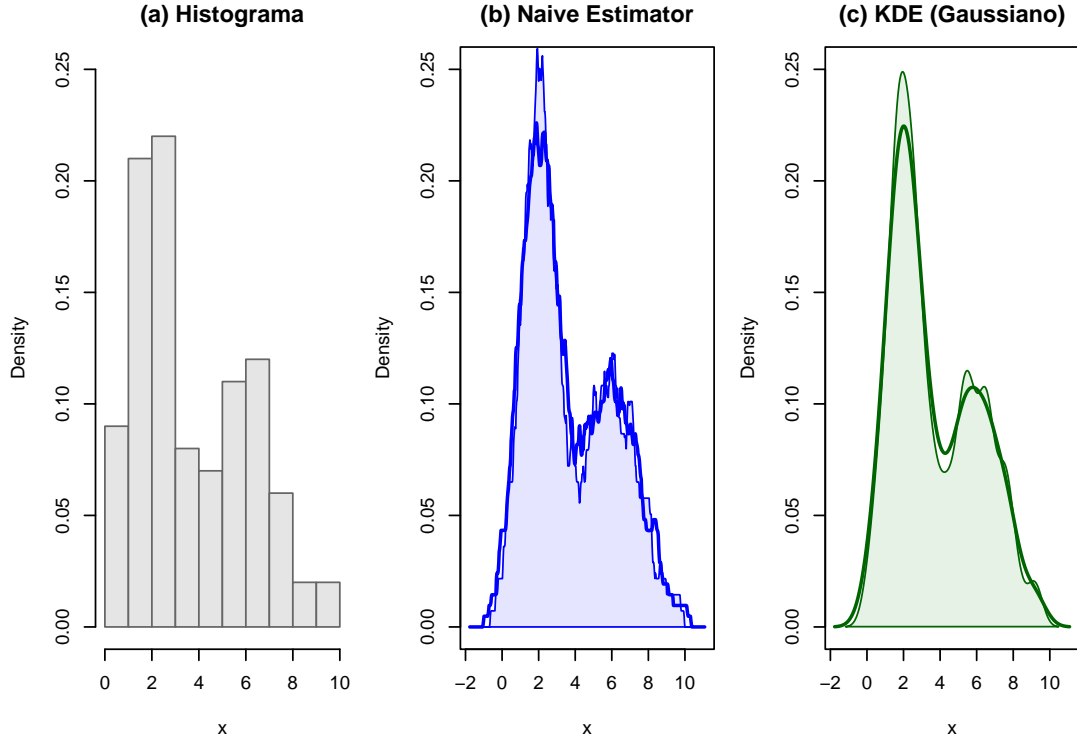


Figura 1: (a) Histograma, (b) Naive Estimator (Rectangular), (c) Estimador Núcleo Suave (Gaussiano)

3.2. Definición Formal y Propiedades del Núcleo

Tras establecer la motivación intuitiva, procedemos a la construcción rigurosa del estimador. La generalización del histograma mediante el uso de funciones de ponderación suaves fue formalizada independientemente por Rosenblatt (1956) y Parzen (1962), dando lugar al objeto central de este estudio.

3.2.1. El Estimador de Rosenblatt-Parzen

Definición 3.2 (Estimador de Densidad Tipo Núcleo Univariante). Sea X_1, X_2, \dots, X_n una muestra aleatoria simple (i.i.d.) proveniente de una población con función de densidad de probabilidad desconocida f . Se define el estimador de densidad tipo núcleo (KDE), denotado por $\hat{f}_h(x)$, como:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad \forall x \in \mathbb{R}$$

donde:

- $K : \mathbb{R} \rightarrow \mathbb{R}$ es la función **núcleo** (*kernel*).
- $h > 0$ es el **ancho de banda** (*bandwidth*) o parámetro de suavizado.

Matemáticamente, este estimador puede interpretarse como la **convolución** de la función de densidad empírica con el núcleo reescalado. Si definimos la densidad empírica como una suma de deltas de Dirac, $f_n(x) = \frac{1}{n} \sum \delta(x - X_i)$, y el núcleo reescalado como $K_h(u) = \frac{1}{h} K(u/h)$, entonces:

$$\hat{f}_h(x) = (f_n * K_h)(x) = \int_{-\infty}^{\infty} K_h(x - t) dF_n(t)$$

Esta interpretación es fundamental en análisis funcional, pues transfiere las propiedades de regularidad de K a la estimación \hat{f}_h .

3.2.2. Clase de Funciones Admisibles

Para que $\hat{f}_h(x)$ sea un estimador válido de una densidad de probabilidad, la función núcleo K debe satisfacer ciertas restricciones analíticas. Definimos la clase de funciones admisibles \mathcal{K} mediante las siguientes condiciones de regularidad.

Condición 1: Normalización e Integrabilidad.

Para garantizar que $\int_{-\infty}^{\infty} \hat{f}_h(x) dx = 1$, el núcleo debe ser una densidad de probabilidad:

$$\int_{-\infty}^{\infty} K(u) du = 1$$

Demostración:

$$\int \hat{f}_h(x) dx = \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K\left(\frac{x - X_i}{h}\right) dx$$

Haciendo el cambio de variable $u = (x - X_i)/h$, tenemos $dx = hdu$, por lo que la integral vale $\int K(u) du = 1$ para cada sumando.

Condición 2: No negatividad.

Para que $\hat{f}_h(x) \geq 0$ para todo x , se requiere:

$$K(u) \geq 0, \quad \forall u \in \mathbb{R}$$

Aunque existen núcleos de orden superior que permiten valores negativos para reducir el sesgo asintótico (kernels de corrección de sesgo), estos generan estimaciones que no son densidades válidas *stricto sensu*. En este trabajo, nos restringiremos a núcleos no negativos (Clase \mathcal{K}_2).

Condición 3: Simetría y Centrado.

Se asume que K es una función par:

$$K(u) = K(-u)$$

Esto implica que el primer momento del núcleo es nulo (siempre que la integral exista):

$$\mu_1(K) = \int_{-\infty}^{\infty} u K(u) du = 0$$

Esta propiedad es crucial para asegurar que la estimación no esté sistemáticamente desplazada hacia la izquierda o derecha de los datos (ausencia de sesgo de fase).

Condición 4: Momentos Finitos.

Se requiere que el núcleo tenga varianza finita y no nula. Definimos el segundo momento del núcleo como:

$$\mu_2(K) = \int_{-\infty}^{\infty} u^2 K(u) du = \sigma_K^2 > 0$$

Este parámetro $\mu_2(K)$ aparecerá directamente en el término principal del sesgo asintótico. Un núcleo más disperso (mayor $\mu_2(K)$) tenderá a suavizar más la muestra.

Condición 5: Rugosidad (Norma L_2).

Para el análisis de la varianza del estimador, se requiere que el núcleo sea cuadrado-integrable. Definimos la **rugosidad** del núcleo como:

$$R(K) = \int_{-\infty}^{\infty} K(u)^2 du = \|K\|_2^2 < \infty$$

Esta cantidad mide la concentración de la masa del núcleo. Cuanto mayor es $R(K)$, más “picuda”

es la función K .

3.2.3. Herencia de Propiedades de Suavidad

Una de las ventajas teóricas más potentes del estimador núcleo frente al histograma es la transmisión de propiedades analíticas.

Teorema 3.1 (Diferenciabilidad). Sea $K \in \mathcal{C}^k(\mathbb{R})$, es decir, K es k -veces diferenciable con continuidad. Entonces, el estimador $\hat{f}_h(x)$ también pertenece a $\mathcal{C}^k(\mathbb{R})$ y su r -ésima derivada viene dada por:

$$\hat{f}_h^{(r)}(x) = \frac{1}{nh^{r+1}} \sum_{i=1}^n K^{(r)}\left(\frac{x - X_i}{h}\right)$$

Este resultado implica que si elegimos un núcleo como el Gaussiano ($K \in \mathcal{C}^\infty$), la densidad estimada será una función infinitamente diferenciable, permitiendo no solo estimar la densidad, sino también sus derivadas (gradientes, curvaturas) sin esfuerzo adicional. Por el contrario, un núcleo como el Triangular ($K \in \mathcal{C}^0$) producirá una estimación continua pero no diferenciable en los puntos de la muestra.

3.2.4. Nota sobre la Generalización Multivariante

Aunque este capítulo se centra en el caso univariante ($d = 1$), la definición 3.2 es fácilmente generalizable a \mathbb{R}^d . En dicho caso, el escalar h se sustituye por una matriz de anchos de banda simétrica definida positiva \mathbf{H} (o un vector), y el núcleo K pasa a ser una función $K : \mathbb{R}^d \rightarrow \mathbb{R}$. Las propiedades de integración y simetría se mantienen análogas en el espacio vectorial.

3.3. Tipos de Funciones Núcleo

La elección de la función núcleo K constituye el segundo grado de libertad en el diseño del estimador, junto con el ancho de banda h . Si bien, como se demostrará en el análisis asintótico del Capítulo 4, la elección del núcleo tiene una influencia menor en el error cuadrático medio (MISE) que la elección del ancho de banda, esta decisión no es trivial y obedece a criterios de **regularidad analítica** y **eficiencia computacional**.

Desde el punto de vista de la eficiencia estadística, se busca un núcleo que minimice el MISE asintótico. Hodgers y Lehmann (1956) y posteriormente Epanechnikov (1969) demostraron que existe una forma funcional óptima que minimiza este error. Sin embargo, la pérdida de eficiencia al utilizar otros núcleos es marginal (apenas un 5-7% de diferencia en el peor de los casos habituales), por lo que en la práctica la elección suele guiarse por las propiedades de suavidad de la estimación resultante \hat{f}_h .

Recordando el **Teorema 3.1**, la estimación \hat{f}_h hereda las propiedades de diferenciabilidad del núcleo K . Por tanto:

- Si el objetivo es obtener una densidad suave e infinitamente diferenciable (útil para estimar derivadas de la densidad o buscar modas), se preferirá un núcleo como el **Gaussiano**.
- Si el objetivo es la eficiencia computacional o minimizar el error cuadrático estricto en un dominio acotado, se preferirá un núcleo de soporte compacto como el de **Epanechnikov**.

A continuación, definimos formalmente los núcleos pertenecientes a la clase \mathcal{K}_2 más relevantes en la literatura, detallando sus momentos y normas cuadráticas, valores que serán constantes necesarias para las fórmulas de selección de ancho de banda.

3.3.1. 1. Núcleo de Epanechnikov

Este núcleo es el **óptimo teórico** en el sentido de minimizar el error asintótico medio integrado cuadrático (AMISE). Corresponde a una parábola invertida restringida a un soporte compacto.

Definición:

$$K_{Epa}(u) = \frac{3}{4}(1 - u^2) \cdot \mathbb{I}(|u| \leq 1)$$

donde $\mathbb{I}(\cdot)$ es la función indicadora.

Propiedades:

- **Soporte:** Compacto en $[-1, 1]$. Esto implica que para estimar $f(x)$, solo se utilizan las observaciones X_i tales que $|x - X_i| \leq h$, lo que reduce el coste computacional.
- **Regularidad:** Es una función continua, pero su derivada es discontinua en $u = \pm 1$. Por tanto, \hat{f}_h será continua pero no diferenciable en todo el dominio.
- **Momentos:**
 - Varianza del núcleo: $\mu_2(K) = \int_{-1}^1 u^2 \frac{3}{4}(1 - u^2) du = \frac{1}{5}$.
 - Rugosidad: $R(K) = \int_{-1}^1 \left[\frac{3}{4}(1 - u^2) \right]^2 du = \frac{3}{5}$.

3.3.2. 2. Núcleo Gaussiano (Normal)

Es la elección estándar en la mayoría de software estadístico debido a sus propiedades de suavidad. Corresponde a la función de densidad de la distribución normal estándar.

Definición:

$$K_{Gauss}(u) = \phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right), \quad u \in \mathbb{R}$$

Propiedades:

- **Soporte:** Infinito $(-\infty, \infty)$. Cada observación X_i contribuye a la estimación en cualquier punto x , aunque su influencia decae exponencialmente con la distancia.
- **Regularidad:** Pertenece a la clase \mathcal{C}^∞ (infinitamente diferenciable). Esto garantiza que la estimación \hat{f}_h también sea \mathcal{C}^∞ , produciendo curvas visualmente muy agradables sin “picos”.
- **Momentos:**
 - Varianza del núcleo: $\mu_2(K) = \int_{-\infty}^{\infty} u^2 \phi(u) du = 1$.

- Rugosidad: $R(K) = \int_{-\infty}^{\infty} \phi(u)^2 du = \frac{1}{2\sqrt{\pi}} \approx 0.282$.

3.3.3. 3. Núcleo Uniforme (Rectangular)

Este núcleo convierte al estimador KDE en el método de “ventana móvil” (*moving average*). Aunque ineficiente para la visualización debido a su falta de suavidad, es útil como referencia teórica.

Definición:

$$K_{Unif}(u) = \frac{1}{2} \cdot \mathbb{I}(|u| \leq 1)$$

Propiedades:

- **Regularidad:** Es una función discontinua en los bordes del soporte. La estimación resultante será una función escalonada (similar a un histograma suavizado).
- **Momentos:**
 - Varianza del núcleo: $\mu_2(K) = \frac{1}{3}$.
 - Rugosidad: $R(K) = \frac{1}{2}$.

3.3.4. 4. Núcleo Triangular

Representa un compromiso lineal entre el núcleo rectangular y el parabólico.

Definición:

$$K_{Tri}(u) = (1 - |u|) \cdot \mathbb{I}(|u| \leq 1)$$

Propiedades:

- **Regularidad:** Función continua, pero no diferenciable en el origen ($u = 0$) ni en los bordes.
- **Momentos:**
 - Varianza del núcleo: $\mu_2(K) = \frac{1}{6}$.
 - Rugosidad: $R(K) = \frac{2}{3}$.

3.3.5. Resumen de Constantes del Núcleo

Dado que las expresiones del sesgo y la varianza asintótica del estimador dependen explícitamente de las cantidades $\mu_2(K)$ y $R(K)$, resumimos sus valores en la siguiente tabla. Estas constantes determinan el factor de escala necesario para comparar anchos de banda entre diferentes núcleos (concepto de *ancho de banda canónico*).

Cuadro 1: Propiedades y constantes de los núcleos más usuales. La columna eficiencia se refiere a la eficiencia asintótica relativa respecto al núcleo de Epanechnikov.

Núcleo (K)	Expresión $K(u)$	Varianza $\mu_2(K)$	Rugosidad $R(K)$	Eficiencia
Epanechnikov	$\frac{3}{4}(1 - u^2)\mathbb{I}_{(u \leq 1)}$	1/5	3/5	100 %
Gaussiano	$\frac{1}{\sqrt{2\pi}}e^{-u^2/2}$	1	$1/(2\sqrt{\pi})$	$\approx 95 \%$
Uniforme	$\frac{1}{2}\mathbb{I}_{(u \leq 1)}$	1/3	1/2	$\approx 93 \%$
Triangular	$(1 - u)\mathbb{I}_{(u \leq 1)}$	1/6	2/3	$\approx 99 \%$

3.3.6. Comparación Visual del Efecto del Núcleo

Para ilustrar las diferencias estructurales de cada núcleo, repetimos el experimento reduciendo el tamaño muestral a $n = 40$. Esto permite visualizar cómo la forma individual de cada función núcleo contribuye a la suma total, evitando que el solapamiento excesivo enmascare las discontinuidades.

La Figura 3.3 muestra los resultados con un ancho de banda fijo ($h = 0.8$):

- En el gráfico (c) (**Rectangular**) se aprecia claramente la naturaleza de **función escalonada** del estimador. Cada observación añade un "bloque", y la suma de bloques resulta en una función con saltos discretos y derivada nula en casi todo punto.
- En el gráfico (d) (**Triangular**) se observan los **picos angulosos** en los máximos locales y los cambios bruscos de pendiente, lo que confirma que la estimación no es diferenciable.
- Por el contrario, los núcleos **Gaussiano (a)** y **Epanechnikov (b)** producen curvas suaves, siendo el Gaussiano el único que reparte probabilidad en todo el eje real (nótese cómo las colas no tocan el cero).

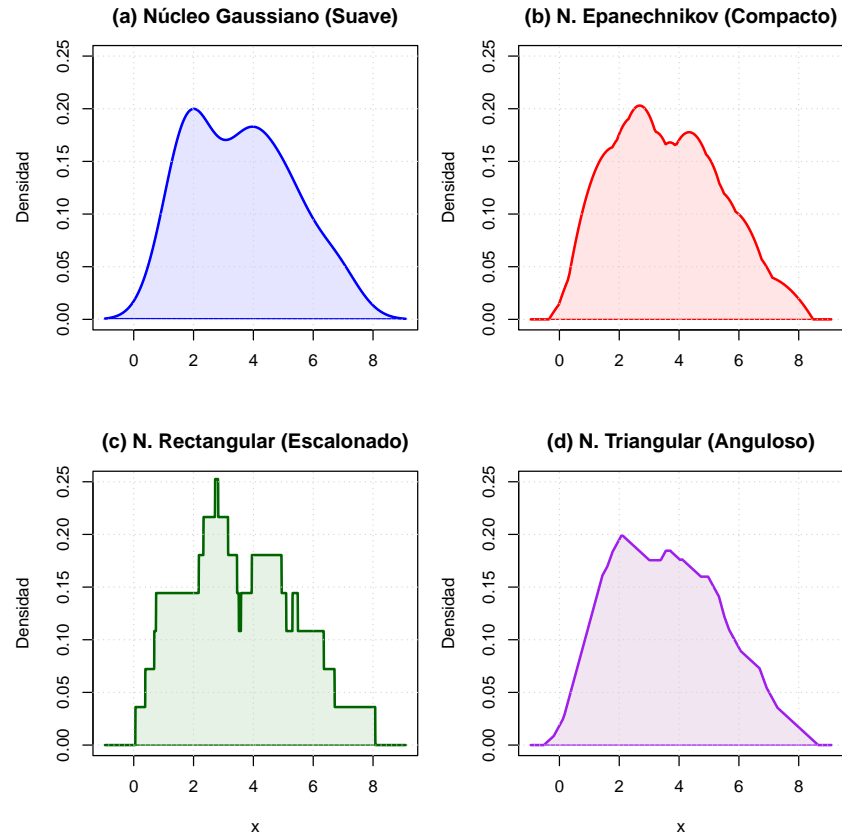


Figura 2: Efecto de la forma del núcleo con muestra reducida ($n=40$, $h=0.8$). Nótese la naturaleza escalonada del núcleo Rectangular (c) y la rugosidad del Triangular (d) frente a la suavidad del Gaussiano (a).

3.4. El Papel del Ancho de Banda (h)

El ancho de banda h es el parámetro libre más crítico en la estimación de densidad tipo núcleo. Mientras que la elección del núcleo K determina propiedades cualitativas como la diferenciabilidad de la curva estimada, el parámetro h gobierna las propiedades cuantitativas y estocásticas del estimador: el sesgo y la varianza.

Matemáticamente, el ancho de banda actúa como un factor de escala que determina el tamaño del entorno local sobre el cual se promedian las observaciones. Para comprender su impacto, es necesario formalizar el error que cometemos al estimar f .

3.4.1. Formulación Matemática del Compromiso Sesgo-Varianza

Consideremos el **Error Cuadrático Medio** (MSE) del estimador en un punto x fijo, definido como la esperanza del cuadrado de la diferencia entre el estimador y el valor real:

$$MSE[\hat{f}_h(x)] = \mathbb{E} \left[(\hat{f}_h(x) - f(x))^2 \right]$$

Mediante la descomposición estándar del error cuadrático, podemos separar este término en dos componentes aditivos: el cuadrado del sesgo y la varianza.

$$MSE[\hat{f}_h(x)] = \underbrace{\left(\mathbb{E}[\hat{f}_h(x)] - f(x) \right)^2}_{\text{Sesgo}^2(x)} + \underbrace{\mathbb{V}[\hat{f}_h(x)]}_{\text{Varianza}(x)}$$

El comportamiento de estos dos componentes respecto a h es antagónico, lo que da lugar al problema fundamental de selección del ancho de banda:

1. **El Sesgo (*Bias*):** Es una función creciente de h . Un ancho de banda grande “aplana” la estimación, alejando sistemáticamente el valor esperado $\mathbb{E}[\hat{f}_h(x)]$ de la verdadera curvatura de la densidad $f(x)$.

$$\lim_{h \rightarrow \infty} \text{Sesgo} \text{ (Alto)}$$

2. **La Varianza (\mathbb{V}):** Es una función decreciente de h . Un ancho de banda pequeño reduce el número efectivo de observaciones que contribuyen a la estimación en x , aumentando la volatilidad muestral.

$$\lim_{h \rightarrow 0} \text{Varianza} \text{ (Alta)}$$

Dado que nuestro objetivo es minimizar el error global sobre todo el dominio, definimos el **Error Cuadrático Integrado Medio (MISE)** como la integral del MSE:

$$MISE(h) = \int_{-\infty}^{\infty} MSE[\hat{f}_h(x)] dx = \int_{-\infty}^{\infty} \text{Sesgo}^2(x, h) dx + \int_{-\infty}^{\infty} \text{Varianza}(x, h) dx$$

El objetivo central de la teoría de selección de ancho de banda, que desarrollaremos analíticamente en los Capítulos 4 y 5, es encontrar el valor h_{opt} que minimice este funcional:

$$h_{opt} = \underset{h>0}{\operatorname{argmin}} MISE(h)$$

3.4.2. Ilustración del Fenómeno

Empíricamente, este equilibrio matemático se manifiesta visualmente en la suavidad de la curva. La elección incorrecta de h conduce a dos patologías opuestas:

- **Infrasuavizado (*Undersmoothing*):** Ocurre cuando $h \ll h_{opt}$. La varianza domina el error. La estimación es ruidosa y presenta falsos picos locales generados por la aleatoriedad de la muestra y no por la estructura de la población.
- **Sobresuavizado (*Oversmoothing*):** Ocurre cuando $h \gg h_{opt}$. El sesgo domina el error. La estimación es excesivamente plana, ocultando características importantes como la multimodalidad o la asimetría.

La Figura 3.4 ilustra estos regímenes utilizando el núcleo Gaussiano.

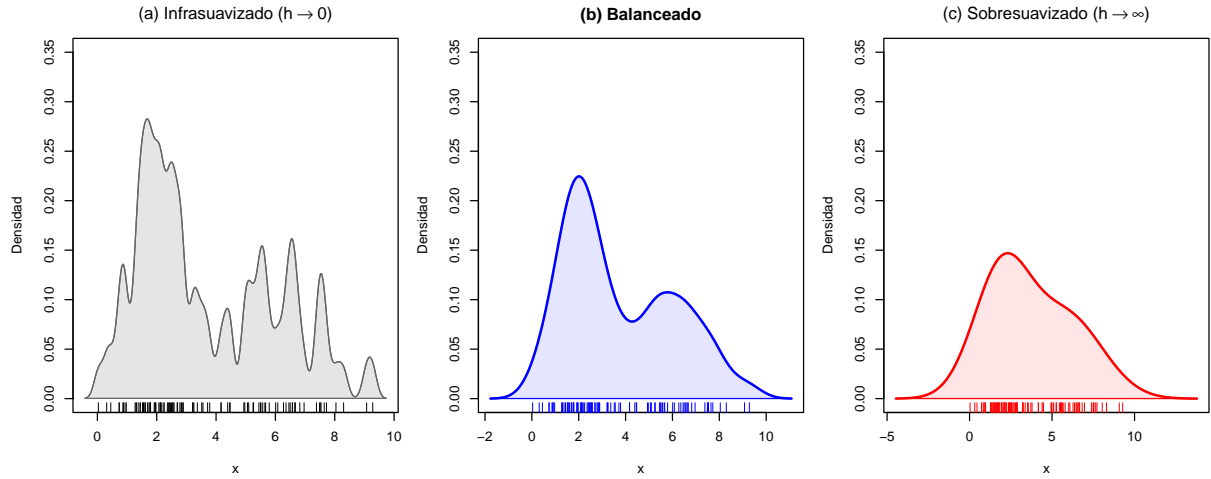


Figura 3: Ilustración del Bias-Variance Tradeoff.