

Máster en Estadística Aplicada

Departamento de Estadística e Investigación Operativa

Universidad de Granada



Trabajo fin de máster

Estimación KDE: fundamentos y aplicaciones

Juan Rubio Cobeta

Granada, 16 de diciembre de 2025

Máster en Estadística Aplicada

Departamento de Estadística e Investigación Operativa

Universidad de Granada



Trabajo de investigación presentado por Don Juan Rubio y dirigido por la profesora Dña. Maria Dolores Martínez Miranda.

VºBº

Maria Dolores Martínez Miranda

Juan Rubio Cobeta

# Índice

<b>1. Introducción y Motivación</b>	<b>1</b>
1.1. Contexto y Justificación . . . . .	1
1.2. Motivación Práctica . . . . .	2
1.3. Objetivos y Estructura del Trabajo . . . . .	2
<b>2. Conceptos Básicos y Notación</b>	<b>4</b>
<b>3. El Estimador de Densidad Tipo Núcleo</b>	<b>4</b>

# 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  $\theta$ . 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 aplana 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 distribu-

ción.

- **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. Conceptos Básicos y Notación**

Aquí pegaremos las definiciones de probabilidad, densidad, esperanza, etc.

## **3. El Estimador de Densidad Tipo Núcleo**

Aquí definiremos el estimador y pondremos los gráficos de los núcleos en R.