

Propuesta de Contenidos del TFM

Estimador de Densidad Kernel (KDE)

Juan Rubio

November 26, 2025

Contents

1 Introducción y Motivación del TFM	2
1.1 Planteamiento del Problema	2
1.2 Objetivos del TFM	2
2 Marco Teórico	2
2.1 Fundamentos básicos de Estadística y Probabilidad	2
2.2 Estimador de Densidad Kernel (KDE) Univariante	2
2.3 Familias de Funciones Kernel y sus Propiedades	2
2.4 El Problema del Ancho de Banda (h): Optimización del Estimador	3
2.5 Generalización a Múltiples Dimensiones (KDE Multivariante)	3
3 Aplicación Financiera	3
4 Conclusiones y Futuras Líneas de Investigación	3

1 Introducción y Motivación del TFM

Esta sección establecerá el contexto y la motivación del trabajo. Explicaremos la necesidad de métodos no paramétricos, como el KDE, para el análisis de distribuciones en el ámbito financiero, donde la asunción de normalidad es a menudo inadecuada (colas pesadas, asimetría). Hacemos hincapié en la limitación del enfoque tradicional para la gestión del riesgo.

1.1 Planteamiento del Problema

Se identificará la limitación de los modelos paramétricos tradicionales (como el uso de la Distribución Normal) en la modelización del riesgo y la volatilidad financiera. Esto justifica el estudio de una técnica más flexible y robusta ante distribuciones no gaussianas.

1.2 Objetivos del TFM

- Objetivo Principal: Demostrar la superioridad del KDE frente a histogramas y distribuciones paramétricas en la estimación de la función de densidad de rendimientos financieros.
- Exponer la teoría matemática del KDE y los métodos de selección del ancho de banda (h).
- Aplicar el KDE a una muestra real de datos financieros.
- Implementar y comparar la eficiencia de diferentes funciones Kernel.

2 Marco Teórico

La idea es intentar realizar un estudio formal y riguroso del KDE.

2.1 Fundamentos básicos de Estadística y Probabilidad

Para empezar, definir los conceptos básicos de estadística y probabilidad necesarios para entender el KDE. Cosas como variables aleatorias, la esperanza, varianza, funciones de distribución acumulada, etc etc (una variable).

Nuestro foco principal será la densidad $f(x)$ como el objeto que buscamos estimar. Justificaremos la necesidad de la estimación no paramétrica mencionando y desarrollando las limitaciones de la estimación paramétrica (por ejemplo asumiendo una distribución Normal) en la modelización de los rendimientos financieros, que a menudo presentan asimetría y curtosis elevadas.

2.2 Estimador de Densidad Kernel (KDE) Univariante

Procederemos a la formalización del KDE para una única variable. Presentaremos la fórmula del estimador $\hat{f}_h(x)$ y definiremos sus componentes clave: el número de observaciones (n), la función Kernel (K), y el ancho de banda (h). Este será el núcleo conceptual de la teoría. Introducir los conceptos de Sesgo (Bias) y Varianza (Var) en este apartado. Explicaremos por qué $\hat{f}_h(x)$ es un estimador, y definiremos las cuatro propiedades matemáticas esenciales que debe cumplir cualquier función Kernel $K(u)$ (simetría, integración a uno, etc.). En estas secciones teóricas me podría alargar tanto como nos pareciera a los dos, dependiendo también un poco de la complejidad del apartado práctico (si con los datos que elegimos no podemos hacer algo demasiado complejo, pues profundizo más en esta parte, y viceversa).

2.3 Familias de Funciones Kernel y sus Propiedades

Presentamos los Kernels más utilizados. Definiremos matemáticamente el Kernel Gaussiano (o Normal), el Epanechnikov, el Rectangular, el Triangular y el Coseno.

Discutiremos cómo la elección del Kernel afecta la suavidad y el soporte del estimador, aunque enfatizaremos que su impacto en la precisión del estimador es generalmente menor que el del ancho de banda (h).

2.4 El Problema del Ancho de Banda (h): Optimización del Estimador

Desarrollo teórico del ancho de banda, su definición, generalización, y su importancia en el equilibrio entre sesgo y varianza. Describimos métodos para elegir diferentes anchos de banda, intentando optimizar, etc,etc.

2.5 Generalización a Múltiples Dimensiones (KDE Multivariante)

Desarrollaremos la generalización del KDE para k dimensiones ($\mathbf{x} \in \mathbb{R}^k$). Generalizar un la teoría expuesta hasta ahora univariante, para el caso multivariante. Me puedo alargar tanto como nos parezca, aunque soy consciente que esta es la parte mas compleja y teórica.

3 Aplicación Financiera

En esta sección, la idea es conseguir un buen conjunto de datos sobre una acción o índice, por ejemplo el SP500.

Al ser el texto dinámico, iremos comentando los resultados de los chunks conforme avanzamos con el desarrollo.

Tengo varios conjuntos de datos mirados, pero es difícil saber si serán lo suficientemente complejos como para que den suficiente juego en la parte práctica. Supongo que será ir probando. La idea es realizar un análisis tal que se aprecie las ventajas de usar KDE frente a histogramas o distribuciones normales. Podemos también jugar con diferentes kernel y anchos de banda para ver posibles resultados alternativos.

4 Conclusiones y Futuras Líneas de Investigación

Típica sección de final de trabajo.