

Departamento de Estadística e Investigación Operativa. Universidad de Sevilla
Estadística Computacional I. GE, DGME
Trabajo sobre el Tema 3: Estimación de la función de densidad
Curso 2022-2023

El objetivo de este trabajo es realizar una comparación empírica de dos métodos para calcular el parámetro ancho de ventana en el método de estimación núcleo de la función de densidad. Para ello se debe implementar en R el siguiente algoritmo base.

Algoritmo base:

Sea Z una variable aleatoria conocida.

Fijar M (número de muestras) y n (tamaño muestral).

Construir una rejilla de valores representativos de la v.a. Z con una longitud de 512 puntos.

Calcular la función de densidad real de la v.a. Z en cada punto de la rejilla.

Repetir M veces {

Generar una muestra $\mathbf{X}=(X_1, \dots, X_n)$ de tamaño n de Z .

Estimar (y guardar) la función de densidad en cada punto de la rejilla mediante el método núcleo (Gaussiano) a partir de la muestra \mathbf{X} , eligiendo el parámetro amplitud de ventana mediante el método I.

Estimar (y guardar) la función de densidad en cada punto de la rejilla mediante el método núcleo (Gaussiano) a partir de la muestra \mathbf{X} , eligiendo el parámetro amplitud de ventana mediante el método II.

}

Para la variable aleatoria Z se considerará una mixtura de dos leyes normales, con medias poblacionales respectivas 1.5 y -1.5, ambas leyes con varianza 1, y ambas componentes con el mismo peso.

Los métodos I y II serán elegidos libremente por el alumno.

M debe ser 1000, para n se deben considerar dos valores, 50 y 100.

Por tanto se consideran 4 configuraciones: 2 tamaños muestrales \times 2 métodos de elección del parámetro amplitud de ventana.

Se trata de responder, al menos, a las siguientes preguntas:

- i) Para cada tamaño muestral n , ¿existe diferencia significativa entre los ECMI para los métodos I y II?
- ii) Fijado el método I, ¿existe diferencia significativa entre los ECMI para los dos tamaños muestrales?

Las respuestas se basarán en el estudio descriptivo e inferencial de los datos resultantes de las simulaciones.

La estimación de los ECMI se realiza del siguiente modo, para cada tamaño n :

- A partir de la matriz de M muestras por 512 puntos conteniendo las M estimaciones de la función de densidad en cada uno de los 512 puntos, estimar el ECM en cada uno de esos puntos, distinguiendo las componentes sesgo al cuadrado y varianza, a partir de los M valores.
- El ECMI se estima mediante la media de los 512 estimadores del ECM calculados según el punto anterior.

Un ejemplo para ilustrar cómo calcular estimaciones núcleo en una rejilla de puntos:

```
Minx= -3.5
```

```
Maxx = 3.5
```

```
Rejilla= seq(Minx, Maxx, length=512)
```

```
n=50
```

```
x=rnorm(n)
```

```
densidad=density(x, bw="SJ", from=Minx,to=Maxx)
```

Se debe entregar, hasta el 10 de Abril inclusive, mediante la herramienta de Mensajes de Curso:

Fichero R y una presentación que incluya los resultados principales y su interpretación.

O bien un fichero R Markdown y la salida generada.