

Ejercicio de estimación no paramétrica

En el archivo **buffalo.txt** se encuentran los datos que corresponden a la mediciones de cantidad de nieve caída (en pulgadas) en Buffalo en los inviernos de 1910/1911 a 1972/1973.

1. Realizar un histograma para estos datos utilizando los parámetros por default. Repetir eligiendo como puntos de corte las siguientes secuencias:
 - a) De 20 a 130 con un paso de 10
 - b) De 22 a 132 con un paso de 10
 - c) De 24 a 134 con un paso de 10
 - d) De 26 a 136 con un paso de 10
2. Realizar un histograma para estas observaciones utilizando puntos de corte (10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130). Comparar todos los histogramas obtenidos. ¿Tiene algún efecto la elección del punto inicial para estos datos?
3. Sea X la cantidad de nieve caída en un invierno en Buffalo. Implementar una función que dados los valores x , h y un conjunto de datos, permita estimar a $P(X \in [x-h, x+h]) = P(x-h \leq X \leq x+h)$ para cada valor x .
4. Calcular la estimación de la probabilidad definida en el ítem anterior para cada valor x de la muestra usando $h = 10, 20$ y 30 .
5. Implementar una función **densidad.est.parzen** que dados un conjunto de datos, una ventana h y un punto x y devuelva $\hat{f}_h(x)$, el valor de la estimación de la densidad f en el punto x , utilizando el núcleo uniforme (también llamado rectangular).
6. Con la función **densidad.est.parzen** implementada, estimar la densidad f en el intervalo (25,126.4) (mínimo y máximo de las observaciones) sobre una grilla de 200 puntos equiespaciados para $h = 10, 20$ y 30 . Graficar las estimaciones obtenidas sobre un mismo gráfico.
7. La función de R **density** computa un estimador de la densidad utilizando núcleos a partir de un conjunto de datos y la evalúa en un conjunto de puntos intermedios. Mediante la función de R **density** estimar la función de densidad f de la variable X a partir de los datos de Buffalo utilizando el núcleo normal, el rectangular y el de Epanechnikov con ventana $h = 5$. Realizar un gráfico superponiendo las tres estimaciones de f y el histograma.
8. Repetir el ítem anterior para $h = 1, 5, 10, 50$. Estudiar el comportamiento de las diferentes estimaciones.