

En esta guía exploraremos los comandos `ecdf()`, `plot(ecdf())`, `approxfun`, `density()`.

Análisis de datos de Buffalo.

Parte 1: Estimación

Los datos que se muestran a continuación se hallan en el archivo `buffalo.txt` y corresponden a la mediciones de cantidad de nieve caída (en pulgadas) en Buffalo en los inviernos de 1910/1911 a 1972/1973.

| | | | | | | | | | | | |
|-------|-------|-------|-------|-------|------|-------|------|-------|------|-------|-------|
| 126.4 | 82.4 | 78.1 | 51.1 | 90.9 | 76.2 | 104.5 | 87.4 | 110.5 | 25.0 | 69.3 | 53.5 |
| 39.8 | 63.6 | 46.7 | 72.9 | 79.6 | 83.6 | 80.7 | 60.3 | 79.0 | 74.4 | 49.6 | 54.7 |
| 71.8 | 49.1 | 103.9 | 51.6 | 82.4 | 83.6 | 77.8 | 79.3 | 89.6 | 85.5 | 58.0 | 120.7 |
| 110.5 | 65.4 | 39.9 | 40.1 | 88.7 | 71.4 | 83.0 | 55.9 | 89.9 | 84.8 | 105.2 | 113.7 |
| 124.7 | 114.5 | 115.6 | 102.4 | 101.4 | 89.8 | 71.5 | 70.9 | 98.3 | 55.5 | 66.1 | 78.4 |
| 120.5 | 97.0 | 110.0 | | | | | | | | | |

1. Realice un histograma para estos datos utilizando los parámetros por default. Repetir eligiendo como puntos de corte las siguientes secuencias: i) de 20 a 140 con paso 10 y ii) de 20 a 140 con paso 5. Comparar los tres histogramas obtenidos. ¿Tiene algún efecto el refinamiento de los bins?
2. Realice un histograma para estas observaciones utilizando puntos de corte (10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130). Repita corriendo el punto de inicio de los bins en 2 unidades 2 veces consecutivas. Compare los tres histogramas obtenidos. ¿Tiene algún efecto la elección del punto inicial en este ejemplo?
3. Sea X la cantidad de nieve caída en un invierno en Buffalo. Implemente una función, `probab.est(x,dat,h)`, que permita estimar a

$$P(X \in [x - h, x + h]) = P(x - h \leq X \leq x + h)$$

para cada valor x a partir de la cantidad de observaciones entre los datos disponibles $\mathbf{dat} = (x_1, \dots, x_n)$ que están a distancia menor o igual que h de x , siendo h el tamaño de ventana elegida.

Utilizando dicha función estime la probabilidad deseada para los valores $x = 40$ y $h = 10$. Compare con la estimación obtenida si cambia a $h = 20$.

4. Calcule la estimación de la probabilidad definida en el ítem anterior para cada valor x de la grilla de datos observados $\mathbf{dat} = (x_1, \dots, x_n)$, usando $h = 10, 20$ y 30 . Compare.
5. Implemente una función **densidad.est.parzen** que tenga por argumento un conjunto de datos $\mathbf{dat} = (x_1, \dots, x_n)$, 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, estime la densidad f en el intervalo $(25, 126.4)$ (mínimo y máximo de las observaciones) sobre una grilla de 200 puntos equiespaciados para $h = 10$. Grafique el estimador $\hat{f}_h(x)$ obtenido.
7. Estime la función de densidad f de la variable *pulgadas de nieve caída en invierno* a partir de los datos de Buffalo a través de la función **densidad.est.parzen** implementada en el ítem anterior usando $h = 10$. Realice un histograma para los datos de Buffalo y superponga la densidad estimada en los datos mediante la función **densidad.est.parzen** utilizando $h = 10, 20$ y 30 . Observe cómo varía la rugosidad de los estimadores de f obtenidos.
8. La función de R **density** computa un estimador de la densidad a partir de un conjunto de datos $\mathbf{x} = (x_1, \dots, x_n)$ y la evalúa en un conjunto de puntos intermedios. Mediante la función de R **density** estime la función de densidad f de la variable *pulgadas de nieve caída en invierno* a partir de los datos de Buffalo y utilizando el núcleo normal, el rectangular y el de Epanechnikov con ventana $h = 5$. Realice un gráfico en el que superpone las tres estimaciones de f y compare los resultados.