Instrucciones: **A lo largo de esta lista de tareas sugerimos explorar los comandos `ecdf()`, `plot(ecdf())`, `approxfun`, `density()`.**

Ejercicio Análisis de datos Gamma-ray bursts.

Los Gamma-ray bursts (GRBs) son uno de los fenómenos más exóticos de los que estudia la astronomía moderna. Fueron descubiertos por casualidad durante los 60's por los satélites americanos y rusos y aparecieron como explosiones con emisiones de rayos gamma con una duración de 0.1-100 segundos en lugares aleatorios del cielo. A partir de mediados de los 90's se empezó a ver a los GRBs como un efecto secundario del nacimiento de un agujero negro. Los GRBs de larga duración se asocian la emisión de un chorro de material energético a velocidades cercanas a la de la luz (relativistas) de una estrella supergigante antes de su colapso final y explosión de supernova. Los GRBs de corta duración se asocian a dos estrellas neutrónicas que se espiralan y emergen como un agujero negro, otra vez eyectando un chorro de material energético a velocidades relativistas.

Como el chorro se expande y se enfría, según cálculos astrofísicos, un resplandor de radiación en longitudes de onda larga, tales como rayos-X, bandas visibles y de radio, deberían ser detectados y deberían decaer en tiempos de escala horas-días-semanas.

Los datos: las observaciones corresponden al decaimiento de los rayos-X de GRB 050525a obtenidos po el Telescopio de rayos-X (XRT) a bordo del satélite Swift (A. J. Blustin y 64 coautores, Astrophys. J. 637, 901-913 2006. Disponible en

`http://arxiv.org/abs/astro-ph/0507515`

El conjunto de datos corresponden a 63 mediciones del brillo en la escala espectral 0.4-4.5 keV a tiempos que van entre 2 minutos y 5 días después de la explosión. Las columnas registran las siguientes variables: 1) el tiempo de observación en segundos, 2) X-ray flux (en unidades de 10-11 ergcm2s, 2-10 keV) y 3) mediciones del error del flux basadas en la relación señal-ruido. En el análisis que sigue ignoraremos la variable **tiempo**.
Para acceder a los datos puede realizar el comando:

```
read.table("http://astrostatistics.psu.edu/datasets/GRB_afterglow.dat",
header=T, skip=1)
```

y la variable de nuestro interés X corresponde a **flux**, que se halla en la segunda columna.

1. Estimar $P(X \leq 40)$.

2. Graficar la empírica asociada a los datos **flux**.

3. Realizar un histograma para los datos de **flux**. ¿Identifica alguna distribución conocida?

Ejercicio Bonus de La empírica (Para hacer con R)

El propósito de este bonus es apreciar la variabilidad de la empírica de acuerdo a los datos que se tengan.

B1. Fijar una semilla y generar n datos correspondientes a una muestra $X_1, \dots, X_n$, i.i.d., $n = 25$, de variables aleatorias distribuidas como X con una distribución $\mathcal{U}(0,1)$. Graficar la empírica asociada a los datos generados. Superponga la distribución acumulada de una $\mathcal{U}(0,1)$. Comparar.

B2. Repetir el ítem anterior con 3 semillas distintas y disponer los 4 gráficos en un mismo plot. ¿Qué sugieren estos gráficos?

B3. Repetir los ítems anteriores con $n = 1000$ y comparar con los gráficos anteriores.

B4. Repetir los ítems previos generando datos correspondientes ahora a $X_1, \dots, X_n$, i.i.d. de variables aleatorias distribuidas como X con una distribución $\mathcal{E}(\lambda)$ con $\lambda = 1$. Comparar los resultados obtenidos con la Uniforme y la Exponencial.

Ejercicio Análisis de datos de Buffalo.

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
```

4. 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?

5. 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?

6. Sea X la cantidad de nieve caída en un invierno en Buffalo. Implemente una función que permita estimar a $P\Big(X \in [x-h, x+h]\Big) = P\Big(x-h \leq X \leq x+h)\Big)$ 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. Es decir, defina la función $\textit{probab.est}(x, \mathbf{dat}, h)$. Cada grupo debe completar en el documento linkeado las estimaciones para los valores que le solicitamos acá .

7. 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.

8. 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 $\widehat{f}_h(x)$, el valor de la estimación de la densidad f en el punto x, utilizando el núcleo uniforme (también llamado rectangular).

9. 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 $\widehat{f}_h(x)$ obtenido.

10. 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 item 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.

11. La función de R **density** computa un estimador de la densidad a partir de un conjunto de datos $\mathbf{x} = (x_1, \ldots, x_n)$ 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.

12. Para los datos de Buffalo halle la ventana óptima h_S por la regla de Silverman para el núcleo normal, calcule la estimación de f basada en el núcleo normal con dicha ventana y grafique la estimación de la densidad obtenida.

13. Se puede realizar la estimación de la densidad f a partir de un conjunto de datos a través de la función **density** en un punto deseado x (f.hat.x) de la siguiente forma:

    ```
    f.hat.x<- density(datos, from=x, to=x, n=1)
    ```

    También mediante la función **approxfun** de R se puede aproximar la densidad estimada por la función de R **density** en el punto deseado x a partir de un conjunto de datos de la siguiente manera:

    ```
    df <- approxfun(density(datos))
    f.hat.x<-df(x)
    ```

    Calcule la estimación de la densidad usando **density** a partir de los datos de Buffalo usando la ventana de Silverman y con esta estimación obtenga estimaciones de la densidad en los valores observados de nieve caída en Buffalo. Grafique la densidad estimada y superponga en el mismo plot los puntos correspondientes a las observaciones y el valor estimado de la densidad correspondiente en rojo.

14. Llamemos $\widehat{f}_h^{(-i)}$ a la densidad estimada sin utilizar al punto x_i y usando la ventana h. Calcule las estimaciones $\widehat{f}_h^{(-i)}$ para $i = 17, 20, 51$ de la densidad f usando la función **density** a partir de los datos de Buffalo usando el núcleo de Epanechnikov y la ventana $h = 5$. Grafique en un mismo plot la estimación de f basada en todos los datos usando el núcleo de Epanechnikov y la ventana $h = 5$ y las estimaciones $\widehat{f}_h^{(-i)}$ para $i = 17, 20, 51$. Compare las estimaciones obtenidas.

15. Halle la ventana de convalidación cruzada para la estimación de la densidad basada en el núcleo normal utilizando los datos de Buffalo. Explore el comando **bw.ucv**.

16. Grafique en un mismo plot las estimaciones de la densidad obtenidas con el núcleo de normal usando la ventana óptima obtenida en el item anterior y usando la ventana h_S hallada previamente. Compare los resultados.