

Maestría en Computo Estadístico
Inferencia Estadística
Tarea 2

20 de septiembre de 2020

Enrique Santibáñez Cortés

Repositorio de Git: Tarea 3, IE.

Escriba de manera concisa y clara sus resultados, justificando los pasos necesarios. Serán descontados puntos de los ejercicios mal escritos y que contenga ecuaciones sin una estructura gramatical adecuada. Las conclusiones deben escribirse en el contexto del problema. Todos los programas y simulaciones tienen que realizarse en R. 1. Resuelva lo siguiente:

- a) Sea $X \sim \text{Exponencial}(\beta)$. Encuentre $\mathbb{P}(|X - \mu_X| \geq k\sigma_X)$ para $k > 1$. Compare esta probabilidad con la que obtiene de la desigualdad de Chebyshev.

RESPUESTA

Cómo $X \sim \text{Exponencial}(\beta)$ entonces $\mu_X = \beta$ y $\sigma_X = \sqrt{\beta^2} = \beta$. Por lo que:

$$\begin{aligned}\mathbb{P}(|X - \mu_X| \geq k\sigma_X) &= \mathbb{P}(|X - \beta| \geq k\beta) = \mathbb{P}(X - \beta \leq -k\beta) + \mathbb{P}(X - \beta \geq k\beta) \\ &= \mathbb{P}(X \leq -k\beta + \beta) + \mathbb{P}(X \geq k\beta + \beta),\end{aligned}$$

ahora como $k > 1$ eso implica que $-k\beta + \beta < 0$, como X es una variable con distribución exponencial podemos concluir que $\mathbb{P}(X \leq -k\beta + \beta) = 0$. Continuando simplificando y sabiendo que X se distribuye exponencial:

$$\begin{aligned}\mathbb{P}(|X - \mu_X| \geq k\sigma_X) &= \mathbb{P}(X \geq k\beta + \beta) \\ &= F(k\beta + \beta) \quad (\text{fda exponencial}) \\ &= e^{-\frac{k\beta + \beta}{\beta}} \\ &= e^{-k-1}.\end{aligned}$$

Ahora, recordemos la desigualdad de Chebyshev.

Teorema: 1 (*Desigualdad de Chebyshev*) Sea X una v.a., $\mu = \mathbb{E}(X)$ y $\sigma^2 = V(X)$. Entonces, si $t > 0$

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2} \tag{1}$$

Entonces, considerando el teorema anterior y haciendo a $t = k\sigma_X$ (note que se sigue cumpliendo que $t > 0$) tenemos:

$$\mathbb{P}(|X - \mu_X| \geq k\sigma_X) \leq \frac{\sigma_X^2}{(k\sigma_X)^2} = \frac{1}{k^2}.$$

Comparando la probabilidad obtenida con la cota de la desigualdad de Chebyshev:

$$e^{-k-1} \leq \frac{1}{k^2} \quad \blacksquare.$$

- b) Sean $X_1, \dots, X_n \sim \text{Bernoulli}(p)$ independientes y $\bar{X} = n^{-1} \sum_{i=1}^n X_i$. Usando las desigualdades de Chebyshev y Hoeffding, acote $\mathbb{P}(|\bar{X} - p| > \epsilon)$. Demuestre que para n grande la cota de Hoeffding es más pequeña que la cota de Chebyshev. ¿En qué beneficia esto?

RESPUESTA

Para calcular la cota de Chebyshev, primero calculemos $E[\bar{X}]$ y $\text{Var}[\bar{X}]$. Como las X_i son independientes y debido a que $X_i \sim \text{Bernoulli}(p)$:

$$E[\bar{X}] = \frac{1}{n} E\left[\sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{np}{n} = p, \quad y$$

$$\text{Var}[\bar{X}] = \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

Utilizando (1) y haciendo $\epsilon = t$ tenemos:

$$\mathbb{P}(|\bar{X} - p| > \epsilon) \leq \mathbb{P}(|\bar{X} - p| \geq \epsilon) \leq \frac{\frac{p(1-p)}{n}}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2},$$

y como $0 \leq p \leq 1$ podemos ver que $p(1-p) \leq \frac{1}{4}$ y por lo tanto la cota de Chebyshev es:

$$\mathbb{P}(|\bar{X} - p| > \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}. \quad (2)$$

Ahora, recordemos la desigualdad de Hoeffding.

Teorema: 2 (Desigualdad de Hoeffding) Sea Y_1, Y_2, \dots, Y_n v.a independientes tales que $E(Y_i) = 0$, $a_i \leq Y_i \leq b_i$. Sea $\epsilon > 0$. Entonces, para cualquier $t > 0$

$$\mathbb{P}\left(\sum_{i=1}^n Y_i \geq \epsilon\right) \leq e^{-t\epsilon} \prod_{i=1}^n e^{t^2(b_i - a_i)^2/8}. \quad (3)$$

Para determinar la cota de la probabilidad solicitada observemos que

$$|\bar{X} - p| > \epsilon = (\bar{X} - p > \epsilon) \cup (p - \bar{X} > \epsilon). \quad (4)$$

Además,

$$\bar{X} - p = \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n p = \sum_{i=1}^n \frac{X_i - p}{n}. \quad (5)$$

Entonces se reduce a encontrar la cota de Hoeffding para:

$$\mathbb{P}\left((\bar{X} - p > \epsilon) \cup (p - \bar{X} > \epsilon)\right) = \mathbb{P}(\bar{X} - p > \epsilon) \cup \mathbb{P}(p - \bar{X} > \epsilon).$$

Primero encontremos la cota para $\mathbb{P}(\bar{X} - p > \epsilon)$. Para ello denotemos a $Y_i = \frac{X_i - p}{n}$, $i = 1, 2, \dots, n$. Ahora, como $X_i \sim \text{Bernoulli}(p)$ esto implica que

$$\mathbb{E}(Y_i) = \mathbb{E}\left(\frac{X_i - p}{n}\right) = \frac{\mathbb{E}(X_i) - p}{n} = \frac{p - p}{n} = 0.$$

Notemos que Y_i tiene una cota inferior haciendo $X = 0$ y una cota superior $X = 1$ las cuales son:

$$-\frac{p}{n} \leq Y_i \leq \frac{1-p}{n}.$$

Cómo ya cumple todos los supuestos, podemos ocupar la desigualdad de Hoeffding en las Y_i . Sea $\epsilon > 0$, y para cualquier $t > 0$

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^n Y_i > \epsilon\right) &= \mathbb{P}\left(\sum_{i=1}^n \frac{X_i - p}{n} > \epsilon\right) \leq \mathbb{P}\left(\sum_{i=1}^n \frac{X_i - p}{n} \geq \epsilon\right) \leq e^{-t\epsilon} \prod_{i=1}^n e^{t^2(\frac{1-p}{n} + \frac{p}{n})^2/8} \\ &= e^{-t\epsilon} \prod_{i=1}^n e^{\frac{t^2}{8n^2}} \\ &= e^{-t\epsilon} e^{\frac{nt^2}{8n^2}} \\ &= e^{-t\epsilon + \frac{t^2}{8n}}. \end{aligned}$$

Ya encontramos una cota utilizando la desigualdad de Hoeffding, pero no es una cota mínima. Para encontrar la mínima encontremos el valor de t que minimiza el exponente de e , es decir, encontremos un mínimo para $f(t) = -t\epsilon + \frac{t^2}{8n}$. Para ello utilicemos el criterio de primera y segunda derivada, derivamos $f(t)$:

$$f'(t) = -\epsilon + \frac{t}{4n}.$$

Igualemos a cero y despejamos t :

$$\begin{aligned} -\epsilon + \frac{t}{4n} &= 0 \\ t &= 4n\epsilon. \end{aligned}$$

Calculemos la segunda derivada de $f(t)$:

$$f''(t) = \frac{1}{4n}.$$

Como $f''(t) > 0$ implica que $t = 4n\epsilon$ sea un mínimo para $f(t)$, y la cota mínima de Hoeffding para este problema es cuanto $t = 4n\epsilon$. Sustituyendo en la desigualdad calculada:

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^n \frac{X_i - p}{n} > \epsilon\right) &\leq e^{-(4n\epsilon)\epsilon + \frac{(4n\epsilon)^2}{8n}} \\ &= e^{-4n\epsilon^2 + \frac{16n^2\epsilon^2}{8n}} \\ &= e^{-2n\epsilon^2}. \end{aligned}$$

Y por lo tanto (5) podemos concluir que:

$$\mathbb{P}(\bar{X} - p > \epsilon) \leq e^{-2n\epsilon^2}.$$

Realizando un razonamiento análogo determinemos la cota para $\mathbb{P}(p - \bar{X} > \epsilon)$. Para ello denotemos $Z_i = \frac{p - X_i}{n}$, $i = 1, 2, \dots, n$. Ahora, como $X_i \sim \text{Bernoulli}(p)$ esto implica que

$$\mathbb{E}(Z_i) = \mathbb{E}\left(\frac{p - X_i}{n}\right) = \frac{p - \mathbb{E}(X_i)}{n} = \frac{p - p}{n} = 0.$$

Notemos que Z_i tiene una cota inferior haciendo $X = 1$ y una cota superior $X = 0$ las cuales son:

$$\frac{p-1}{n} \leq Z_i \leq \frac{p}{n}.$$

Cómo ya cumple todos los supuestos, podemos ocupar la desigualdad de Hoeffding en las Z_i . Sea $\epsilon > 0$, y para cualquier $t > 0$

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^n Z_i > \epsilon\right) &= \mathbb{P}\left(\sum_{i=1}^n \frac{p - X_i}{n} > \epsilon\right) \leq \mathbb{P}\left(\sum_{i=1}^n \frac{p - X_i}{n} \geq \epsilon\right) \leq e^{-t\epsilon} \prod_{i=1}^n e^{t^2(\frac{p}{n} - \frac{p-1}{n})^2/8} \\ &= e^{-t\epsilon} \prod_{i=1}^n e^{\frac{t^2}{8n^2}} \\ &= e^{-t\epsilon} e^{\frac{nt^2}{8n^2}} \\ &= e^{-t\epsilon + \frac{t^2}{8n}}.\end{aligned}$$

Ya encontramos una cota utilizando la desigualdad de Hoeffding, pero no es una cota mínima. Para encontrar la mínima encontremos el valor de t que minimiza el exponente de e , es decir, encontremos un mínimo para $f(t) = -t\epsilon + \frac{t^2}{8n}$. Para ello utilicemos el criterio de primera y segunda derivada, derivamos $f(t)$:

$$f'(t) = -\epsilon + \frac{t}{4n}.$$

Igualemos a cero y despejamos t :

$$\begin{aligned}-\epsilon + \frac{t}{4n} &= 0 \\ t &= 4n\epsilon.\end{aligned}$$

Calculemos la segunda derivada de $f(t)$:

$$f''(t) = \frac{1}{4n}.$$

Como $f''(t) > 0$ implica que $t = 4n\epsilon$ sea un mínimo para $f(t)$, y la cota mínima de Hoeffding para este problema es cuanto $t = 4n\epsilon$. Sustituyendo en la desigualdad calculada:

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^n \frac{p - X_i}{n} > \epsilon\right) &\leq e^{-(4n\epsilon)\epsilon + \frac{(4n\epsilon)^2}{8n}} \\ &= e^{-4n\epsilon^2 + \frac{16n^2\epsilon^2}{8n}} \\ &= e^{-2n\epsilon^2}.\end{aligned}$$

Entonces:

$$\mathbb{P}(p - \bar{X} > \epsilon) \leq e^{-2n\epsilon^2}.$$

Por lo tanto, ocupando (4) podemos concluir que la cota de Hoeffding es:

$$\mathbb{P}(|\bar{X} - p| > \epsilon) \leq 2e^{-2n\epsilon^2}.$$

Entonces para mostrar que cuando n es grande la cota de Hoeffding es más pequeña que la cota de Chebyshev (1), veamos la primera derivada de cada cota con respecto a n . Sea $f(n)$ el denominador de la cota de Hoeffding y $g(n)$ el denominador la cota de Chebyshev:

$$\begin{aligned}f'(n) &= 2\epsilon^2 e^{2n\epsilon^2} / 2 = \epsilon^2 e^{2n\epsilon^2}. \\ g'(n) &= 4\epsilon^2\end{aligned}$$

De aquí podemos observar que el crecimiento de la cota de Chebyshev es lineal no depende de n y en la cota de Hoeffding depende de n . Entonces de lo anterior implica que para que

$$g'(n) < f'(n)$$

se tiene que cumplir que $4 < e^{2n\epsilon^2}$, entonces es sencillo ver que para cualquier $\epsilon > 0$ se puede encontrar un n lo suficientemente grande para que se cumpla la igualdad. Por lo tanto, para algún n grande cumple que el dominador del cota de Hoeffding es más grande que la cota de Chebyshev y por lo tanto que la cota de Hoeffding es más pequeña que la cota de Chebyshev para algún n grande. ■.

2. Sean $X_1, \dots, X_n \sim \text{Bernoulli}(p)$.

a) Sea $\alpha > 0$ fijo y defina

$$\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}.$$

Sea $\hat{p} = n^{-1} \sum_{i=1}^n X_i$. Defina $C_n = (\hat{p}_n - \epsilon_n, \hat{p}_n + \epsilon_n)$. Use la desigualdad de Hoeffding para demostrar que

$$\mathbb{P}(C_n \text{ contiene a } p) \geq 1 - \alpha$$

Diremos que C_n es un $(1 - \alpha)$ -intervalo de confianza para p . En la practica, se trunca el intervalo de tal manera de que no vaya debajo del 0 o arriba del 1.

RESPUESTA

Observemos que el evento de que C_n contiene a p es igual al evento que:

$$\{p \notin C\} = |\hat{p} - p| > \epsilon_n.$$

Entonces, como $X_i \sim \text{Bernoulli}(p)$ independientes y utilizando la desigualdad de Hoeffding encontrada en el ejercicio anterior tenemos que:

$$\mathbb{P}(p \notin C) = \mathbb{P}(|\hat{p} - p| > \epsilon_n) \leq 2e^{-2n\epsilon_n^2},$$

sustituyendo el valor de ϵ_n :

$$\begin{aligned} \mathbb{P}(p \notin C) &\leq 2e^{-2n\left(\sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}\right)^2} \\ &= 2e^{-2n\left(\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)\right)} \\ &= 2e^{-\log\left(\frac{2}{\alpha}\right)} \\ &= \frac{2\alpha}{2} = \alpha. \end{aligned}$$

Es decir, la probabilidad de que el intervalo C_n no contenga a p es menor que α :

$$\mathbb{P}(p \notin C) \leq \alpha.$$

Recordando la propiedad de probabilidad $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$. Entonces, podemos ver que

$$\begin{aligned} \mathbb{P}(p \notin C) &\leq \alpha \\ -\mathbb{P}(p \notin C) &\geq -\alpha \\ 1 - \mathbb{P}(p \notin C) &\geq 1 - \alpha \\ \mathbb{P}(p \in C) &\geq 1 - \alpha. \end{aligned}$$

Es decir, queda probado que

$$\mathbb{P}(C_n \text{ contiene a } p) \geq 1 - \alpha. \quad \blacksquare.$$

- b) Sea $\alpha = 0,05$ y $p = 0,4$. Mediante simulaciones, realice un estudio para ver que tan a menudo el intervalo de confianza contiene a p (la cobertura). Haga esto para $n = 10, 50, 100, 250, 500, 1000, 2500, 5000, 10000$. Grafique la cobertura contra n .

RESPUESTA

Tenemos que:

```
library(tidyverse)
confianza <-function(p, alpha, n){
  epsilon <- sqrt((1/(2*n))*log(2/alpha))
  limite_inferior <- p-epsilon
  limite_superior <- p+epsilon

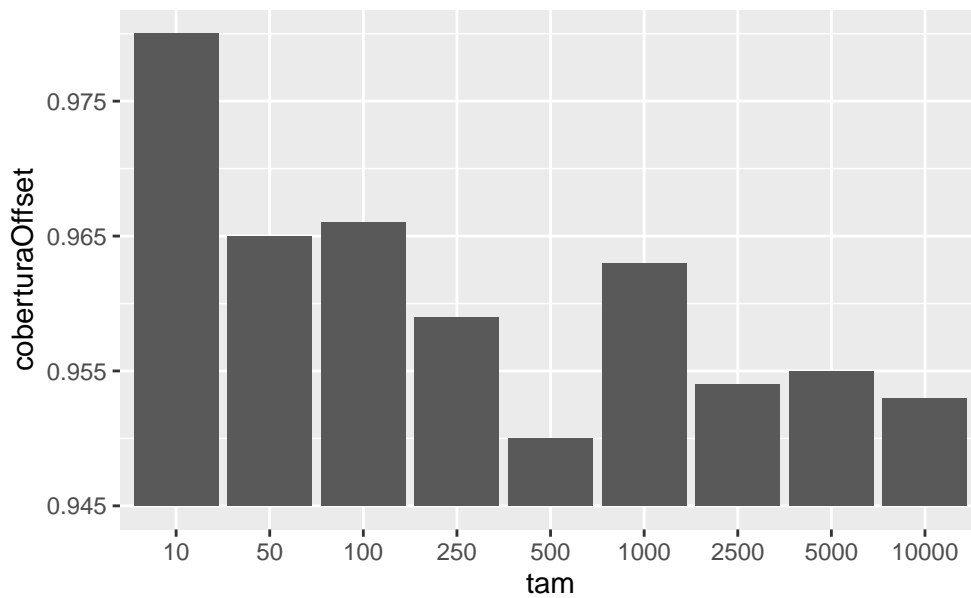
  simulaciones_bernoulli <- data.frame(matrix(rbinom(size = 1, n = 1000*n,p),ncol=1000))%>%
    summarise_all(funs(sum)) %>%
    gather(key = "num_sim", value="conteo") %>%
    mutate(conteo= conteo/n,
           contenido=between(conteo, limite_inferior, limite_superior))
  data.frame(tam=n,cobertura=sum(simulaciones_bernoulli$contenido)/1000,
             lim_sup=limite_superior,
             lim_inf=limite_inferior)
}

n <- c(10, 50, 100, 250, 500, 1000, 2500, 5000, 10000)
resultados_simulacion <- c()

for (i in 1:length(n)){
  set.seed(08081997)
  resultados_simulacion<- rbind(resultados_simulacion,confianza(0.4, 0.25, n=n[i]))
}

resultados_simulacion <- resultados_simulacion %>%
  mutate(alpha=1-cobertura,
         tam=factor(tam),
         coberturaOffset = cobertura - 0.945)

ggplot(resultados_simulacion, aes(x=tam, y=coberturaOffset))+
  geom_col() +
  scale_y_continuous(labels = function(x) x + 0.945)
```

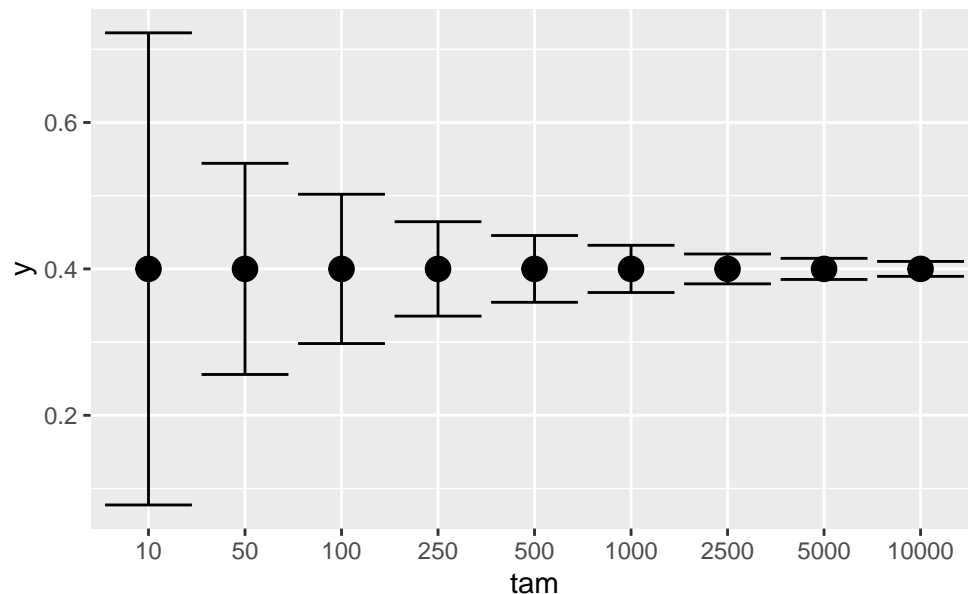


- c) Grafique la longitud del intervalo contra n . Suponga que deseamos que la longitud del intervalo sea menor que 0.05. ¿Qué tan grande debe ser n ?

RESPUESTA

Lo que:

```
ggplot(resultados_simulacion, aes(x=tam, y=0.4))+
  geom_point(size=4)+
  geom_errorbar(aes(ymax=lim_sup, ymin=lim_inf))
```



3. Una partícula se encuentra inicialmente en el origen de la recta real y se mueve en saltos de una unidad. Para cada salto, la probabilidad de que la partícula salte una unidad a la izquierda es p y la probabilidad de que salte una unidad a la derecha es $1 - p$. Denotemos por X_n a la posición de la partícula después de n saltos. Encuentre $\mathbb{E}(X_n)$ y $\text{Var}(X_n)$. Esto se conoce como una caminata aleatoria en una dimensión.

RESPUESTA

Este proceso cambia de un estado a otro en dos tiempos consecutivos de acuerdo con probabilidad de

transición descritas en el problemas. Estas probabilidades se pueden escribir de la forma siguiente:

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \begin{cases} p & \text{si } j = i + 1, \\ 1 - p & \text{si } j = i - 1, \\ 0 & \text{otro caso.} \end{cases}$$

Esa caminata aleatoria puede también definirse de la siguiente forma: ξ_1, ξ_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas. Por la idéntica distribución denotemos a cualquiera de ellas mediante la letra ξ sin subíndice. Ahora, si suponemos que $\mathbb{P}(\xi = +1) = p$ y $\mathbb{P}(\xi = -1) = 1 - p$. Entonces para $n \geq 1$ se define

$$X_n := X_0 + \xi_1 + \xi_2 + \dots + \xi_n.$$

donde X_0 en este caso suponemos que empieza en 0, es decir, el estado inicial de la partícula es cero. Entonces, a partir de la expresión anterior implica que la esperanza es:

$$\mathbb{E}(X_n) = \mathbb{E}(X_0) + \sum_{i=1}^n \mathbb{E}(\xi_i) = n\mathbb{E}(\xi) = n(p - 1 + p) = n(2p - 1).$$

Ahora, como $\mathbb{E}(\xi^2) = p(1)^2 + (1 - p)(-1)^2 = 1$ y $\mathbb{E}(\xi) = 2p - 1$, se tiene que $Var(\xi) = 1 - (2p - 1)^2 = 1 - 4p^2 + 4p - 1 = 4p(1 - p)$. Y por lo tanto la varianza de X_n es:

$$Var(X_n) = \sum_{i=1}^n Var(\xi_i) = nVar(\xi) = 4np(1 - p). \quad \blacksquare.$$

4. El siguiente conjuntos de datos contiene mediciones del diámetro de un agave, medido en decímetros, en distintas localizaciones no cercanas.

23,37	21,87	24,41	21,27	23,33	15,20	24,21	27,52	15,48	27,19
25,05	20,40	21,05	28,83	22,90	18,00	17,55	25,92	23,64	28,96
23,02	17,32	30,74	26,73	17,22	22,81	20,78	23,17	21,60	22,37

- a) Escriba una función en R que calcule la función de distribución empírica para un conjunto de datos dado D . La función debe tomar como parámetros al valor x donde se evalúa y al conjunto de datos D . Utilizando esta función grafique la función de distribución empírica asociada al conjunto de datos de agave. Ponga atención a los puntos de discontinuidad. ¿Qué observa? **Nota:** Escriba la función mediante el algoritmo descrito en las notas de la clase; para este ejercicio no vale usar la funciones implementadas en R que hacen lo pedido.

RESPUESTA

Sea $X_1, \dots, X_n \sim F$, la función de densidad empírica \hat{F} es la función de distribución acumulada que asigna masa $1/n$ en cada punto X_i . Formalmente,

$$\hat{F}(x) = \frac{\sum_{i=1}^n 1_{X_i \leq x}}{n},$$

donde

$$1_{X_i \leq x} = \begin{cases} 1 & \text{si } X_i \leq x \\ 0 & \text{si } X_i > x \end{cases}$$

Utilizando la definición anterior construimos la función en R:

```
fde <- function(x, D){
  n <- length(D)
  if(length(x)>1){
    fde_x <- c()
```



```

    for(i in 1:length(x)){
      fde_x[i] <- sum(D<x[i])/n
    }
  }
  else{fde_x <- sum(D<x)/n}
  fde_x
}

```

Procedemos a calcular con la función anterior la fde en los puntos de discontinuidad, para ello primero ingresemos los datos:

```

df_agave <- data.frame(diametro=c(23.37, 21.87, 24.41, 21.27, 23.33, 15.20, 24.21, 27.52, 15.48,
25.05, 20.40, 21.05, 28.83, 22.90, 18.00, 17.55, 25.92, 23.64, 28.96, 23.02, 17.32,
30.74, 26.73, 17.22, 22.81, 20.78, 23.17, 21.60, 22.37))

df_agave <- df_agave %>% arrange(diametro)

```

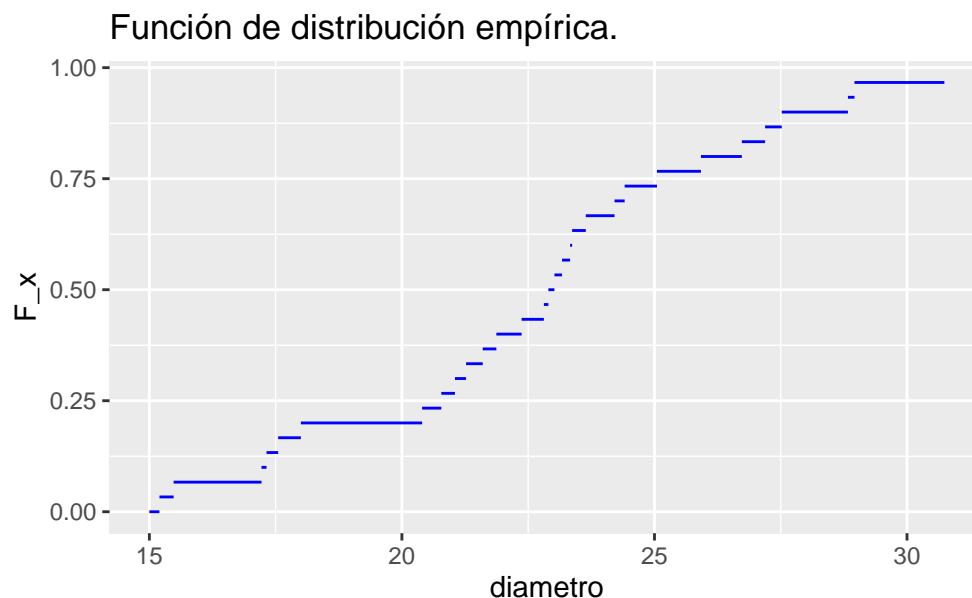
Los puntos de discontinuidad serían cuando cambia fde, es decir, cuando se evalúa en un $x + \epsilon$ tal que $\hat{F}(x) \neq \hat{F}(x + \epsilon)$, con los datos podemos decir que los puntos de discontinuidad son los estadísticos de orden $x_{(i)}$.

```

df_agave_fde <- df_agave %>%
  mutate(F_x=fde(x=diametro, D=df_agave$diametro),
         xend=lag(x=diametro, n = 1, default = 15))

ggplot(df_agave_fde)+
  geom_segment(aes(x=diametro, xend=xend, y=F_x, yend=F_x), col="blue")+
  labs(title = "Función de distribución empírica.")

```



- b) Usando la desigualdad de Dvoretzky-Kiefer-Wolfowitz, escriba una función en R que calcule y grafique una región de confianza para la función de distribución empírica. La función debe tomar como parámetros al conjunto de datos que se usan para contruir la función de distribución empírica.

RESPUESTA

Enunciemos la desigualdad de Dvoretzky-Kiefer-Wolfowitz:

Teorema: 3 (Desigualdad de Dvoretzky-Kiefer-Wolfowitz) Sean $X_1, \dots, X_n \sim F$ independientes. Entonces, para todo $\epsilon > 0$,

$$\mathbb{P}\left(\sup_x |\hat{F}_n(x) - F(x)| > \epsilon\right) \leq 2e^{-2n\epsilon^2}.$$

A partir de lo anterior, podemos construir una región de confianza para la distribución empírica. Sea:

$$L(x) = \max\{\hat{F}_n - \epsilon_n, 0\},$$

$$U(x) = \min\{\hat{F}_n + \epsilon_n, 1\}$$

donde

$$\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}.$$

Puede verse que $F(x)$

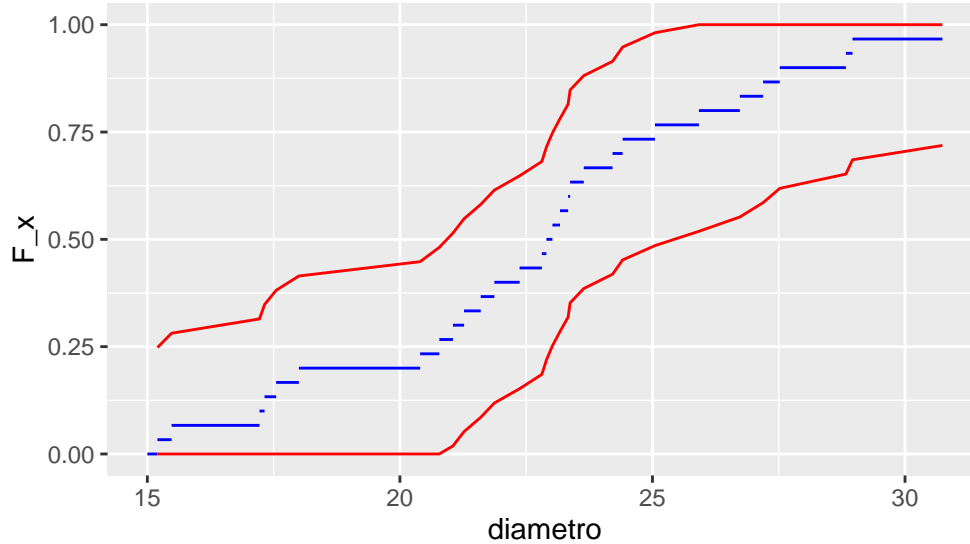
$$\mathbb{P}(L(x) \leq F(x) \leq U(x) \text{ para todo } x) \geq 1 - \alpha.$$

Entonces, una región de confianza con $\alpha = 0,05$ para la distribución empírica se puede programar de la siguiente forma:

```
graphisc_fde <- function(df){
  alpha <- 0.05
  df <- df %>%
    mutate(F_x=fde(x=diametro, D=df_agave$diametro),
           xend=lag(x=diametro, n=1, default = 15),
           L_x=F_x-sqrt((1/(2*30))*log(2/alpha)),
           L_x=if_else(L_x<0,0, L_x),
           U_x=F_x+sqrt((1/(2*30))*log(2/alpha)),
           U_x=if_else(U_x>1,1, U_x))%>%
    gather(key=".", value="y", c(4,5))
  df
}
df_agave_fde_bandas <- graphisc_fde(df_agave)

ggplot(df_agave_fde)+
  geom_segment(aes(x=diametro, xend=xend, y=F_x, yend=F_x), col="blue")+
  labs(title = "Región de confianza de FDE.")+
  geom_line(data=df_agave_fde_bandas, aes(x=diametro, y=y,group=.), col="red")
```

Región de confianza de FDE.



- c) Escriba una función en R que determine la gráfica Q-Q normal de un conjunto de datos. La función debe tomar como parámetro al conjunto de datos y deberá graficar contra el percentil estandarizado de la normal. Para poder comparar el ajuste más claramente, la función además deberá ajustar en rojo a la recta $sx + \bar{x}$ (s =desviación estándar muestral y x =media muestral). Usando esta función, determine la gráfica Q-Q normal. ¿Qué observa? **Nota:** La misma del inciso a).

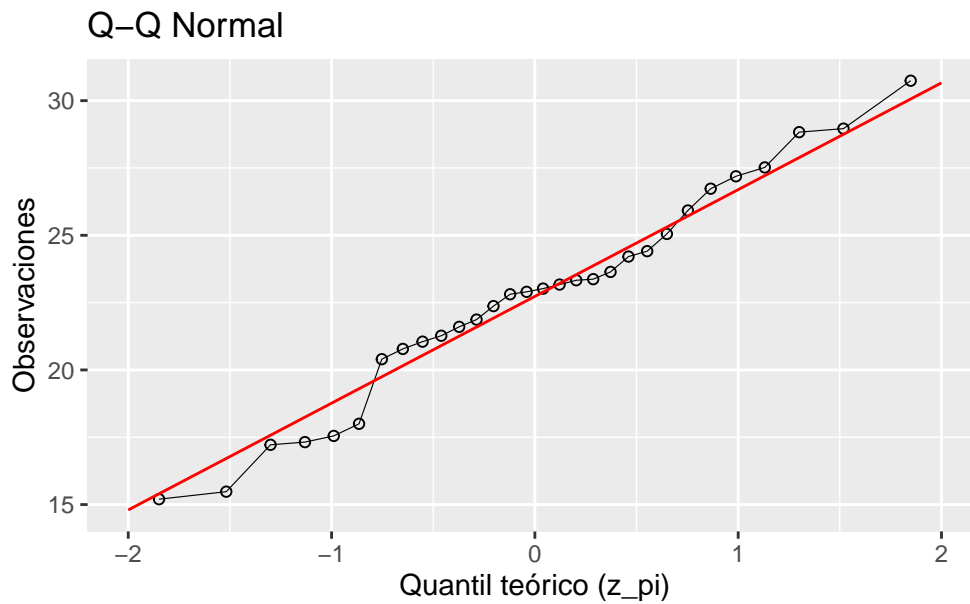
RESPUESTA

```
sd <- sqrt(var(df_agave$diametro))
mu <- mean(df_agave$diametro)
recta <- data.frame(x=c(-2,2), y=c(-sd*2+mu, sd*2+mu))

df_agave$i <- 1:nrow(df_agave)

df_agave <- df_agave %>%
  mutate(pi= i/31,
         z_pi=qnorm(pi))

ggplot(df_agave, aes(x=z_pi, y=diametro))+
  geom_point(shape=1)+
  geom_line(size=0.2)+
  geom_line(data=recta, aes(x=x, y=y), col="red")+
  labs(title = "Q-Q Normal", x="Quantil teórico (z_pi)", y="Observaciones")
```



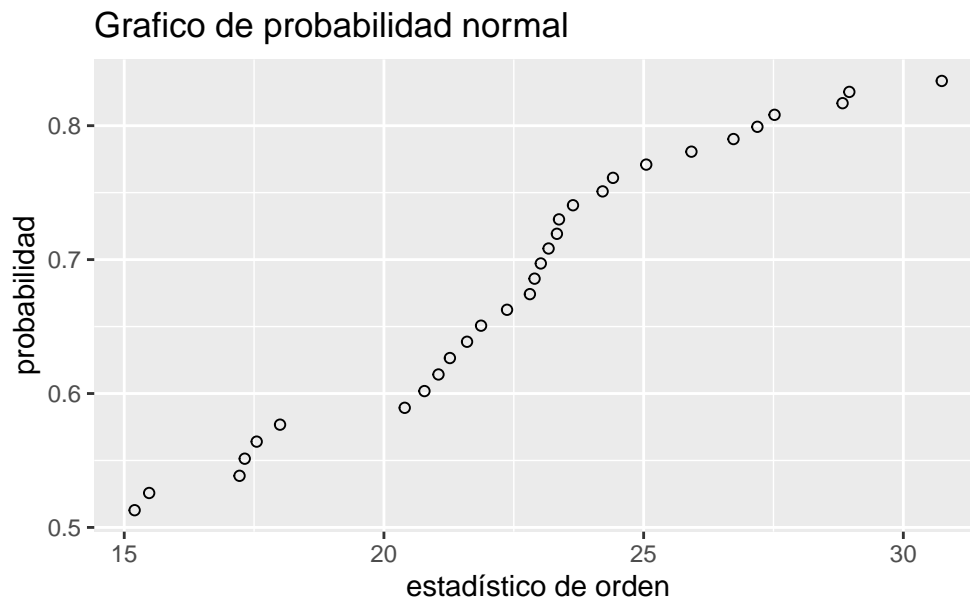
- d) Escriba una función en R que determine el gráfico de probabilidad normal. La función debe tomar como parámetro al conjunto de datos. ¿Qué observa? Nota: La misma del inciso a).

RESPUESTA

Utilizando:

```
df_agave <- df_agave %>%
  mutate(z_pi_in = pnorm(pi))

ggplot(df_agave, aes(y=z_pi_in, x=diametro))+
  geom_point(shape=1)+
  labs(title = "Grafico de probabilidad normal", y="probabilidad", x="estadístico de orden")
```



- e) ¿Los datos anteriores se distribuyen normalmente? Argumente.
5. En este ejercicio repasaré la estimación de densidades.
- a) Escriba una función en R que estime una densidad por el método de kernels. La función deberá

recibir al punto x donde se evalúa al estimador, al parámetro de suavidad h , al kernel que se utilizará en la estimación y al conjunto de datos.

RESPUESTA

Para el caso univariado, el KDE está dado por

$$\hat{f} = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right), \quad x \in \mathbb{R}, h > 0.$$

K es la función Kernel, y h es el ancho de banda que determina la suavidad de la estimación. Procedemos a definir las funciones Kernel's, solo consideraremos las siguientes: gaussian, epanechnikov, rectangular, triangular, consine.

```
# Definimos las funciones kernel:
kernel_gaussian <- function(x,x_i,h){
  dnorm((x-x_i)/h)
}

kernel_rectangular <- function(x,x_i,h){
  dunif((x_i - x)/h, min = -1, max = 1)
}

kernel_triangular <- function(x, x_i, h){
  u = (x_i-x)/h
  (1-abs(u))*(abs(u) <= 1)
}

kernel_epanechnikov <- function(x, x_i, h){
  u = (x_i - x)/h
  3/4*(1-u^2)*(abs(u) <= 1)
}

kernel_cosine <- function(x, x_i, h){
  u = (x_i - x)/h
  (1+cos(pi*u))/2*(abs(u) <= 1)
}
```

Una vez definido lo anterior procedemos a definir la función para estimar la densidad por el método de kerneles (https://github.com/JonasMoss/kdensity/blob/master/R/builtin_kernels.R):

```
kde <- function(x, h, kernel, D){
  f_hat <- 0
  n <- length(D)
  for (i in 1:n){
    if (kernel=="gaussian"){
      f_hat <- f_hat + kernel_gaussian(x, D[i], h)
    }
    else if(kernel=="epanechnikov"){
      f_hat <- f_hat + kernel_epanechnikov(x, D[i], h)
    }
    else if(kernel=="rectangular"){
      f_hat <- f_hat + kernel_rectangular(x, D[i], h)
    }
  }
}
```

```

}
else if(kernel=="triangular"){
  f_hat <- f_hat + kernel_triangular(x, D[i], h)
}
else if(kernel=="consine"){
  f_hat <- f_hat + kernel_cosine(x, D[i], h)
}
else{
  print("Kernel incorrecto")
  break()
}
}
f_hat/(n*h)
}

```

- b) Cargue en R al archivo “Tratamiento.csv”, el cual contiene la duración de los períodos de tratamiento (en días) de los pacientes de control en un estudio de suicidio. Utilice la función del inciso anterior para estimar la densidad del conjunto de datos para $h = 20, 30, 60$. Grafique las densidades estimadas. ¿Cuál es el mejor valor para h ? Argumente.

RESPUESTA

Cargamos los datos:

```
tratamiento <- read.csv("Tratamiento.csv")
```

Graficamos las densidad estimada con el método de Kernel, con diferentes anchos de bandas.

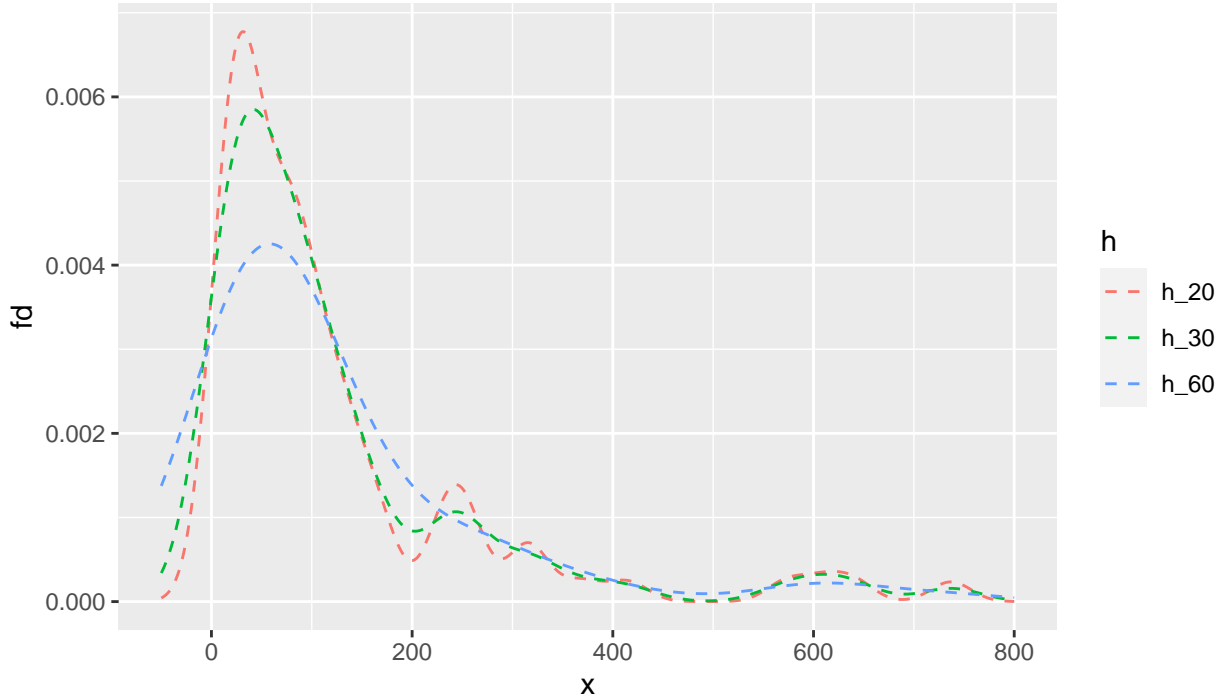
```

x <- seq(-50,800, 1)
fd_tratamiento <- data.frame(x=rep(x,3),
                             fd=c(kde(x, 20, "gaussian", tratamiento$X1),
                                   kde(x, 30, "gaussian", tratamiento$X1),
                                   kde(x, 60, "gaussian", tratamiento$X1)),
                             h=c(rep("h_20", 851), rep("h_30", 851), rep("h_60", 851)))

ggplot(fd_tratamiento, aes(x=x, y=fd)) +
  geom_line(aes(colour=h), linetype=2) +
  labs(title = "Estimación de densidad usando un Kernel.")

```

Estimación de densidad usando un Kernel.



Desde mi perspectiva el mejor valor de ancho de banda h es igual a 30, comparado con $h = 20$ sobreestima a los datos eso se observa en las pequeñas montañas que se aprecian cuando $x \approx 230, 250$; y comparandolo con $h = 60$ observamos que se subestima a los datos.

- c) En el contexto de la estimación de densidades, escriba una función en R que determine el ancho de banda que optimiza al ISE. Grafique la densidad con ancho de banda óptimo para el conjunto de datos de “Tratamiento.csv”.

RESPUESTA

Ocuparemos el método de Least squares cross validation (LSCV). Tenemos que

$$\begin{aligned} ISE(h) &= \int_{-\infty}^{\infty} (\hat{f}(x) - f(x))^2 dx \\ &= \int_{-\infty}^{\infty} \hat{f}^2(x) dx - 2 \int_{-\infty}^{\infty} \hat{f}(x) f(x) dx + \int_{-\infty}^{\infty} f^2(x) dx \end{aligned}$$

Si observamos, la última integral de la expresión no depende del estimador $\hat{f}(x)$ (es constante), por lo que la elección del ancho de banda (en el sentido de minimizar el ISE) corresponderá a la elección de h que minimiza la función:

$$R(\hat{f}) = \int_{-\infty}^{\infty} \hat{f}^2(x) dx - 2 \int_{-\infty}^{\infty} \hat{f}(x) f(x) dx + \int_{-\infty}^{\infty} f^2(x) dx.$$

La segunda parte usando un estimador de la densidad leave-one-out, $\hat{f}_{-i}(x)$ es:

$$\hat{f}_{-i}(x) = \frac{1}{n-1} \sum_{j \neq i} K(x, x_j)$$

el cual es un estimador de la función de densidad usando todas los datos exepcto x_i . El resultado del criterio LSCV es:

$$LSCV(h) = \int_{-\infty}^{\infty} \hat{f}^2(x) dx - \frac{2}{n} \sum_i \hat{f}_{-i}(x_i).$$

El parametro h optimo es el valor que miniza la función $LSCV(h)$. (Weglarczyk 2018)

Lo anterior igual puede ser visto como (Härdle et al. 1991)

$$LSCV(h) = \frac{R(K)}{nh} + \frac{1}{n(n-1)h} \sum_{i=1}^n \sum_{\substack{j=1, \\ j \neq i}}^n (K * K - 2K) \left(\frac{X_j - X_i}{h} \right).$$

donde para un kernel normal

$$R(K) = \frac{1}{2\sqrt{\pi}}$$

$$K * K(x) = \frac{1}{2\pi} e^{-\frac{1}{2} \left(\frac{x}{\sqrt{2}} \right)^2}.$$

Utilizando lo anterior podemos buscar h que hace optimo $LSCV(h)$. Para hacerlo eficiente creamos una matriz de $n \times n$ donde el elemento (i,j) es igual a $K \left(\frac{X_i - X_j}{h} \right)$ para $i \neq j$, y 0 si $i = j$.

```
completa <- function(x, kernel="gaussian"){
  n <- length(x)
  upper=2*70
  lower=0.1*70
  tol=0.1 * lower

  LSCV_h <- function(h){
    D <- dnorm(outer(x,x,"-")/h)
    diag(D) <- 0
    D <- (1 / ((n-1)*h)) * colSums(D)
    D1 <- mean(D)
    D2 <- dnorm(outer(x,x,"-")/h,mean=0,sd=sqrt(2))
    diag(D2) <- 0
    D3 <- (1/((n-1)*h)) * colSums(D2)
    D4 <- mean(D3)
    (1/(n*h)) * (1/(2*sqrt(pi))) + D4 - 2*D1
  }

  obj <- optimize(LSCV_h , c(lower, upper), tol=tol)
  obj
}

completa(tratamiento$X1)
```

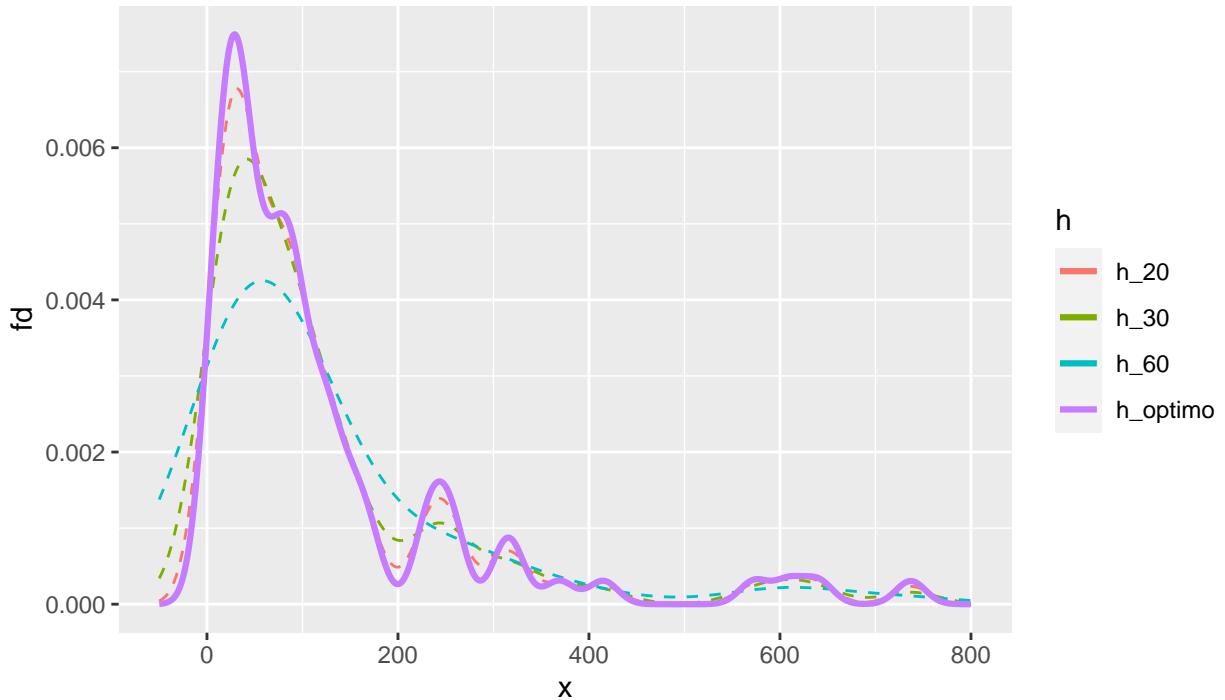
```
## $minimum
## [1] 15.40501
##
## $objective
## [1] -0.004076096
```

Es decir, con $h = 15,40$ se minimiza el ISE. La gráfica de la densidad estimada con esa h optima es:


```
x <- seq(-50,800, 1)
fd_optimo <-data.frame(x=x,
                        fd=kde(x, 15.40, "gaussian", tratamiento$X1),
                        h=c(rep("h_optimo", 851)))

ggplot(fd_tratamiento, aes(x=x, y=fd)) +
  geom_line(aes(colour=h), linetype=2) +
  geom_line(data=fd_optimo, aes(x=x,y=fd, colour=h),linetype=1, size=1.1)+
  labs(title = "Estimación de densidad usando un Kernel.")
```

Estimación de densidad usando un Kernel.



8. En este ejercicio se comprobará que tan buena es la aproximación dada por las reglas empíricas para algunas de las distribuciones estudiadas en la clase. Considerese las distribuciones $Unif(a = -3, b = 3)$, $Normal(0, 1)$, $Exponencial(2)$, $Gamma(\alpha = 2, \beta = 1)$, $Gamma(\alpha = 3, \beta = 1)$, $Beta(\alpha = 2, \beta = 2)$, $Weibull(\alpha = 4, \beta = 1)$ y $Lognormal(\mu = 3, \sigma = 2)$.
- a) Para cada una de las distribuciones anteriores, haga una tabla que muestre las probabilidades contenidas en los intervalos $(\mu - k\sigma, \mu + k\sigma)$, para $k = 1, 2, 3$. Utilice las fórmulas de las medias y varianzas contenidas en las notas para determinar μ y σ en cada caso. Puede usar R para determinar las probabilidades pedidas.

RESPUESTA

Determinemos para cada distribución su media y varianza las cuales se pueden calcular con la table

Distribución de X	$\mathbb{E}[X]$	$\text{Var}(X)$
$Unif(a, b)$	$\frac{b+a}{2}$	$\frac{(a-b)^2}{12}$
$Normal(\mu, \sigma)$	μ	σ^2
$Exponencial(\theta)$	θ	θ^2
$Gamma(\alpha, \beta)$	$\alpha\beta$	$\alpha\beta^2$
$Beta(\alpha, \beta)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta^2)(\alpha+\beta+1)}$
$Weibull(\alpha, \beta)$	$\alpha^{-\frac{1}{\beta}} \Gamma\left(\frac{1}{\beta}\right)$	$\alpha^{-\frac{1}{\beta}} \left(\Gamma\left(\frac{2}{\beta} + 1\right) - \Gamma^2\left(\frac{1}{\beta} + 1\right) \right)$
$LogNormal(\mu, \sigma)$	$e^{\mu + \frac{\sigma^2}{2}}$	$e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$

Para este problema sería:

Distribución de X	$\mathbb{E}[X]$	$\text{Var}(X)$	σ
$Unif(-3, 3)$	$\frac{3-3}{2} = 0$	$\frac{(a-b)^2}{12} = \frac{(-3-3)^2}{12} = 3$	1.73
$Normal(0, 1)$	0	1	1
$Exponencial(2)$	2	4	2
$Gamma(2, 1)$	2	2	1.41
$Gamma(3, 1)$	3	3	1.73
$Beta(2, 2)$	$\frac{2}{2+2} = \frac{1}{2}$	$\frac{2(2)}{(2+2^2)(2+2+1)} = \frac{4}{30}$	3
$Weibull(\alpha, \beta)$	$\alpha^{-\frac{1}{\beta}} \Gamma\left(\frac{1}{\beta}\right)$	$\alpha^{-\frac{1}{\beta}} \left(\Gamma\left(\frac{2}{\beta} + 1\right) - \Gamma^2\left(\frac{1}{\beta} + 1\right) \right)$	
$LogNormal(3, 2)$	$e^{3 + \frac{2^2}{2}} = 148,41$	$e^{2(3)+4} (e^3 - 1)$	1086.544

Calculemos las probabilidades:

```

parametros_distribuciones <- data.frame(distribucion=c("Unif", "Normal", "Exp", "Gamma", "Gamma_2",
resultados_uniforme <- parametros_distribuciones %>% filter(distribucion=="Unif")%>%
  mutate(prob_between_x_s=punif(mu+std, min=-3, max=3)-punif(mu-std, min=-3, max=3),
    prob_between_x_2s=punif(mu+2*std, min=-3, max=3)-punif(mu-2*std, min=-3, max=3),
    prob_between_x_3s=punif(mu+3*std, min=-3, max=3)-punif(mu-3*std, min=-3, max=3))

resultados_normal <- parametros_distribuciones %>% filter(distribucion=="Normal")%>%
  mutate(prob_between_x_s=pnorm(mu+std)-pnorm(mu-std),
    prob_between_x_2s=pnorm(mu+2*std)-pnorm(mu-2*std),

```

```

        prob_between_x_3s=pnorm(mu+3*std)-pnorm(mu-3*std))

resultados_exponencial <- parametros_distribuciones %>% filter(distribucion=="Exp")%>%
  mutate(prob_between_x_s=pexp(mu+std, 2)-pexp(mu-std,2),
         prob_between_x_2s=pexp(mu+2*std, 2)-pexp(mu-2*std, 2),
         prob_between_x_3s=pexp(mu+3*std, 2)-pexp(mu-3*std, 2))

resultados_gamma <- parametros_distribuciones %>% filter(distribucion=="Gamma")%>%
  mutate(prob_between_x_s=pgamma(mu+std, 2, 1)-pgamma(mu-std, 2, 1),
         prob_between_x_2s=pgamma(mu+2*std, 2, 1)-pgamma(mu-2*std, 2, 1),
         prob_between_x_3s=pgamma(mu+3*std, 2, 1)-pgamma(mu-3*std, 2, 1))

resultados_gamma2 <- parametros_distribuciones %>% filter(distribucion=="Gamma")%>%
  mutate(prob_between_x_s=pgamma(mu+std, 3, 1)-pgamma(mu-std, 3, 1),
         prob_between_x_2s=pgamma(mu+2*std, 3, 1)-pgamma(mu-2*std, 3, 1),
         prob_between_x_3s=pgamma(mu+3*std, 3, 1)-pgamma(mu-3*std, 3, 1))

resultados_beta <- parametros_distribuciones %>% filter(distribucion=="Beta")%>%
  mutate(prob_between_x_s=pgamma(mu+std, 2, 2)-pgamma(mu-std, 2, 2),
         prob_between_x_2s=pgamma(mu+2*std, 2, 2)-pgamma(mu-2*std, 2, 2),
         prob_between_x_3s=pgamma(mu+3*std, 2, 2)-pgamma(mu-3*std, 2, 2))

resultados_weibull <- parametros_distribuciones %>% filter(distribucion=="Weibull")%>%
  mutate(prob_between_x_s=pgamma(mu+std, 4, 1)-pgamma(mu-std, 3, 1),
         prob_between_x_2s=pgamma(mu+2*std, 4, 1)-pgamma(mu-2*std, 4, 1),
         prob_between_x_3s=pgamma(mu+3*std, 4, 1)-pgamma(mu-3*std, 4, 1))

resultados_lnorm <- parametros_distribuciones %>% filter(distribucion=="LNormal")%>%
  mutate(prob_between_x_s=pgamma(mu+std, 3, 2)-pgamma(mu-std, 3, 2),
         prob_between_x_2s=pgamma(mu+2*std, 3, 2)-pgamma(mu-2*std, 3, 2),
         prob_between_x_3s=pgamma(mu+3*std, 3, 2)-pgamma(mu-3*std, 3, 2))

resultados_teoricos <- rbind(resultados_beta, resultados_exponencial, resultados_gamma, resultados_gamma2,
                             resultados_lnorm, resultados_normal, resultados_uniforme, resultados_weibull)

resultados_teoricos %>% select(distribucion, prob_between_x_s, prob_between_x_2s, prob_between_x_3s)

##  distribucion prob_between_x_s prob_between_x_2s prob_between_x_3s
## 1      Beta      0.461211399      0.68452660      0.8084851
## 2      Exp      0.999664537      0.99999386      0.9999999
## 3      Gamma    0.735668689      0.95305130      0.9857609
## 4      Gamma    0.640046162      0.85934589      0.9475408
## 5      LNormal  1.000000000      1.00000000      1.0000000
## 6      Normal   0.841344746      0.95449974      0.9973002
## 7      Unif     0.576666667      1.00000000      1.0000000
## 8      Weibull  -0.002236385      0.06389083      0.1006770

```

- b) En R, simule $n = 1000$ muestras de cada una de las distribuciones anteriores y calcule la media muestral \bar{x} y la varianza muestral s^2 como se mencionó en la clase. En cada caso, calcule la proporción de observaciones que quedan en los intervalos $(\bar{x} - k, \bar{x} + ks)$, para $k = 1, 2, 3$. Reporte sus hallazgos

en una tabla como la del inciso anterior. ¿Qué tanto se parecen la tabla de este inciso y la del anterior?

RESPUESTA

Calculemos

```
set.seed(08081997)

muestras_simuladas<- data.frame(x=c(runif(1000, -3, 3), rnorm(1000), rexp(1000, 2),
  rgamma(1000, 2, 1), rgamma(1000, 3, 1), rbeta(1000, 2, 2), rweibull(1000, 4, 1),
  rlnorm(1000,meanlog= 3, sdlog = 2)), distribucion=c(rep("Uniforme", 1000), rep("Normal", 1000),
  rep("Exp", 1000), rep("Gamma_1", 1000), rep("Gamma_2", 1000), rep("Beta", 1000),
  rep("Weib", 1000), rep("LNormal", 1000)))

estadisticos_muestrales <- muestras_simuladas %>% group_by(distribucion) %>%
  summarise(mean_muestral = mean(x),
    std_muestral = sqrt(var(x)))

head(estadisticos_muestrales,8)

## # A tibble: 8 x 3
##   distribucion mean_muestral std_muestral
##   <fct>          <dbl>          <dbl>
## 1 Beta           0.505           0.234
## 2 Exp            0.523           0.505
## 3 Gamma_1        2.04           1.49
## 4 Gamma_2        2.97           1.67
## 5 LNormal       152.           562.
## 6 Normal         0.0111          0.998
## 7 Uniforme      -0.0369           1.74
## 8 Weib           0.907           0.257

resultados <- merge(muestras_simuladas, estadisticos_muestrales)

resultados %>% mutate(
  x_s = ifelse(x>(mean_muestral-std_muestral) & x<(mean_muestral+std_muestral), 1, 0),
  x_2s = ifelse(x>(mean_muestral-2*std_muestral) & x<(mean_muestral+2*std_muestral), 1, 0),
  x_3s = ifelse(x>(mean_muestral-3*std_muestral) & x<(mean_muestral+3*std_muestral), 1, 0)) %>%
  group_by(distribucion) %>% summarise(propor_x_s = sum(x_s)/1000,
    propor_x_2s = sum(x_2s)/1000,
    propor_x_3s = sum(x_3s)/1000)

## # A tibble: 8 x 4
##   distribucion propor_x_s propor_x_2s propor_x_3s
##   <fct>          <dbl>          <dbl>          <dbl>
## 1 Beta           0.619           0.992           1
## 2 Exp            0.829           0.942           0.988
## 3 Gamma_1        0.75           0.955           0.986
## 4 Gamma_2        0.711           0.953           0.988
## 5 LNormal        0.961           0.976           0.987
## 6 Normal         0.672           0.962           0.999
## 7 Uniforme       0.567           1             1
```

```
## 8 Weib          0.697          0.951          0.999
```

```
resultados_teoricos %>% select(distribucion, prob_between_x_s, prob_between_x_2s, prob_between_x_3s)
```

```
##  distribucion prob_between_x_s prob_between_x_2s prob_between_x_3s
## 1      Beta    0.461211399      0.68452660      0.8084851
## 2      Exp     0.999664537      0.99999386      0.9999999
## 3      Gamma   0.735668689      0.95305130      0.9857609
## 4      Gamma   0.640046162      0.85934589      0.9475408
## 5     LNormal  1.000000000      1.00000000      1.0000000
## 6     Normal   0.841344746      0.95449974      0.9973002
## 7      Unif    0.576666667      1.00000000      1.0000000
## 8     Weibull  -0.002236385      0.06389083      0.1006770
```

7. Demuestre que la fórmula de la densidad de la Beta integra 1.

RESPUESTA

Sea $X \sim \text{Beta}(\alpha, \beta)$, entonces su función de densidad esta definida como:

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1.$$

Entonces, la integral de la densidad sería:

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^1 \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx. \quad (6)$$

Ahora recordemos una identidad con la función Gamma, la cuál es:

$$\int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

Por lo anterior tenemos en (6):

$$\int_{-\infty}^{\infty} f(x) dx = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} = 1.$$

Por lo tanto, queda demostrado que la densidad de v.a con distribución Beta integra 1. ■.

Honors problems

1.

a) Sea X una v.a. discreta con media finita y que toma valores en el conjunto $0, 1, 2 \dots$. Demuestre que

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k).$$

RESPUESTA

Definamos la siguiente notación para hacer más entendible la demostración:

$$p_x = \mathbb{P}(X = x), \quad x = 0, 1, \dots,$$

y

$$q_x = \mathbb{P}(X > x) = \sum_{k=x+1}^{\infty} p_k \quad x = 0, 1, \dots.$$

Entonces debemos probar que:

$$\mathbb{E}(x) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k) = \sum_{x=0}^{\infty} q_x.$$

Como sabemos que se cumple que $\mathbb{P}(X \leq x) + \mathbb{P}(x > x) = 1$, observemos que:

$$\begin{aligned} \sum_{x=1}^N xp_x &= p_1 + 2p_2 + 3p_3 + \cdots + Np_N \\ &= (p_1 + p_2 + p_3 + \cdots + p_N) + (p_2 + p_3 + \cdots + p_N) + \cdots + (p_{N-1} + p_N) + p_N \\ &= \sum_{x=0}^{N-1} q_x - Nq_N. \end{aligned}$$

Es decir,

$$\sum_{x=1}^N xp_x = \sum_{x=0}^{N-1} q_x - Nq_N. \quad (7)$$

Utilizando que X tiene media finita, es decir, como $E(X) = \sum_{x=0}^{\infty} xp_x$ podemos decir que la serie $\sum_{x=0}^{\infty} xp_x$ es convergente. Ahora observemos que se cumple la siguiente desigualdad:

$$0 \leq N \sum_{x=N+1}^{\infty} p_x \leq \sum_{x=N+1}^{\infty} xp_x. \quad (8)$$

La justificación de la desigualdad, se debe a que $n > N$ por como se definieron los límites de la suma y de lado derecho a que N y p_x son positivos.

Ahora, ocupemos una propiedad conocida de series convergentes:

Teorema: 4 (Condición necesaria de convergencia) Si la serie $\sum_{n=1}^{\infty} a_n$ es convergente, entonces

$$\lim_{n \rightarrow \infty} a_n = 0.$$

Ocupando la propiedad anterior en (8):

$$\begin{aligned} \lim_{N \rightarrow \infty} 0 &\leq \lim_{N \rightarrow \infty} N \sum_{x=N+1}^{\infty} p_x \leq \lim_{N \rightarrow \infty} \sum_{x=N+1}^{\infty} xp_x \\ &= \sum_{x=N+1}^{\infty} \lim_{x \rightarrow \infty} xp_x \\ &= 0. \end{aligned}$$

Es decir, podemos decir que

$$\lim_{N \rightarrow \infty} N \sum_{x=N+1}^{\infty} p_x = 0.$$

Entonces haciendo un límite en (7) tenemos que:

$$\begin{aligned}
 \lim_{N \rightarrow \infty} \sum_{x=1}^N x p_x &= \lim_{N \rightarrow \infty} \left(\sum_{x=0}^{N-1} q_x - N q_N \right) \\
 &= \lim_{N \rightarrow \infty} \left(\sum_{x=0}^{N-1} q_x \right) - \lim_{N \rightarrow \infty} (N q_N) \\
 &= \lim_{N \rightarrow \infty} \left(\sum_{x=0}^{N-1} q_x \right) - \lim_{N \rightarrow \infty} \left(N \sum_{x=N+1}^{\infty} p_x \right) \\
 &= \sum_{x=0}^{\infty} q_x = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k).
 \end{aligned}$$

Es decir, queda demostrado que

$$\mathbb{E}(X) = \sum_{x=1}^{\infty} x p_x = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k). \quad \blacksquare.$$

- b) Sea X una v.a. continua no-negativa con media finita, función de densidad f y función de distribución F . Demuestre que

$$\mathbb{E}(X) = \int_0^{\infty} (1 - F(t)) dt.$$

RESPUESTA

Utilizaremos algunas definiciones vistas en clase de confiabilidad y su relación con función de densidad. $R(x)$ se le conoce como la confiabilidad de X , la cual se define como:

$$R(x) = 1 - F(x).$$

Es claro que se cumple que:

$$R(0) = 1 \quad \text{y} \quad \lim_{x \rightarrow \infty} R(x) = 0.$$

Derivando de ambos lados observamos que:

$$R'(x) = -F'(x) = -f(x).$$

Ocupando lo anterior y la definición de esperanza de una variable continua tenemos que:

$$\begin{aligned}
 \mathbb{E}(X) &= \int_{-\infty}^{\infty} xf(x)dx && \text{definición de esperanza} \\
 &= \int_0^{\infty} -xR'(x)dx && \text{relación confiabilidad-densidad} \\
 &= \int_0^{\infty} -xR'(x) - R(x) + R(x)dx && \text{sumamos un cero} \\
 &= \int_0^{\infty} -(xR(x))' + R(x)dx && \text{definición de derivada} \\
 &= -xR(x)|_{x=0}^{\infty} + \int_0^{\infty} R(x)dx \\
 &= \int_0^{\infty} R(x)dx \\
 &= \int_0^{\infty} (1 - Fx(x))dx \quad \blacksquare.
 \end{aligned}$$

c) ¿Cómo cambia la fórmula del caso anterior cuando el soporte de X es todo \mathbb{R} ?

RESPUESTA

Cuando el soporte de X es todo \mathbb{R} no tiene sentido la función de confiabilidad. Por lo que

2. Sea X una v.a. continua con primer momento finito. Demuestre que la función $G(c) = E(|X-c|)$, $c \in \mathbb{R}$, se minimiza en $c = M(X)$ para $M(X)$ la mediana de X .

RESPUESTA

Sea $f(x)$ la función de densidad de X . Por propiedades de la esperanza tenemos que:

$$\begin{aligned}
 G(c) &= E(|X - c|) = \int_{-\infty}^{\infty} |x - c|f(x)dx = \int_{-\infty}^c (c - x)f(x)dx + \int_c^{\infty} (x - c)f(x)dx \\
 &= c \int_{-\infty}^c f(x)dx - \int_{-\infty}^c xf(x)dx + \int_c^{\infty} xf(x)dx - c \int_c^{\infty} f(x)dx.
 \end{aligned}$$

Ahora diferenciamos con respecto a c e igualamos a cero la expresión anterior:

$$\begin{aligned}
 G'(c) &= \frac{d}{dc} \left(c \int_{-\infty}^c f(x)dx - \int_{-\infty}^c xf(x)dx + \int_c^{\infty} xf(x)dx - c \int_c^{\infty} f(x)dx \right) \\
 &= \int_{-\infty}^c f(x)dx + c \frac{d}{dc} \int_{-\infty}^c f(x)dx - \frac{d}{dc} \int_{-\infty}^c xf(x)dx + \frac{d}{dc} \int_c^{\infty} xf(x)dx - \int_c^{\infty} f(x)dx - c \frac{d}{dc} \int_c^{\infty} f(x)dx
 \end{aligned}$$

Ahora por el Teorema Fundamental del Calculo.

$$\begin{aligned}
 G'(c) &= \int_{-\infty}^c f(x)dx + cf(x)|_{x=c} - xf(x)|_{x=c} + xf(x)|_{x=c} - \int_c^{\infty} f(x)dx - cf(x)|_{x=c} \\
 &= \int_{-\infty}^c f(x)dx - \int_c^{\infty} f(x)dx.
 \end{aligned}$$

Ahora igualamos a cero:

$$G'(c) = \int_{-\infty}^c f(x)dx - \int_c^{\infty} f(x)dx = 0$$
$$\int_{-\infty}^c f(x)dx = \int_c^{\infty} f(x)dx.$$

Es decir,

$$\mathbb{P}(X \leq c) = \mathbb{P}(X > c).$$

Por definición de probabilidad sabemos que se cumple que $\mathbb{P}(X \leq c) + \mathbb{P}(X > c) = 1$. Lo que implica que:

$$\mathbb{P}(X \leq c) = \mathbb{P}(X > c) = \frac{1}{2}.$$

Y por lo tanto, cuando c es la mediana (por definición) de X minimiza la función $G(c)$ ■.

Bibliografía

Härdle, W.K., D.R. Brillinger, S.E. Fienberg, J. Gani, J.A. Hartigan, J.C. Kiefer, and K. Krickeberg. 1991. *Smoothing Techniques: With Implementation in S*. Springer Series in Statistics. Springer. https://books.google.ne/books?id=2dzo/_zyRagC.

Weglarczyk, Stanislaw. 2018. "Kernel Density Estimation and Its Application." *ITM Web of Conferences* 23 (January): 00037. <https://doi.org/10.1051/itmconf/20182300037>.