Godinez Bravo Diego
Inferencia Estadística
Tarea 4

1. Resuelva lo siguiente:

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

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

Solución

a) Podemos expresar $\mid X - \mu x \mid > k\sigma x$ como $-k\sigma x < X - \mu x < k\sigma x$, entonces

$$P(\mid X - \mu_x \mid > k\sigma_x) = 1 - P(-k\sigma_x < X - \mu_x < k\sigma_x)$$

Sumando μx en ambos lados de la desigualdad

$$= 1 - P(\mu_x - k\sigma_x < X < \mu_x + k\sigma_x)$$

Lo cual podemos expresar como

$$= 1 - (F(\mu_x + k\sigma_x) - F(\mu_x - k\sigma_x))$$

Sea X $Exponencial(\beta)$

$$F(x) = 1 - e^{\lambda t}$$

Sea $\lambda = \frac{1}{\beta}$, entonces

$$= 1 - (1 - e^{-\frac{\beta + k\beta}{\beta}}) + (1 - e^{-\frac{\beta - k\beta}{\beta}})$$

$$= 1 + e^{-\frac{\beta + k\beta}{\beta}} - e^{-\frac{\beta - k\beta}{\beta}}$$

Finalmente

$$P(\mid X - \mu_x \mid > k\sigma_x) = 1 + e^{-(1+k)} - e^{-(1-k)}$$

Por otro lado, dado que $X \sim Exponencial(\beta)$, y sabiendo que $\mu = \beta$ y $\sigma^2 = \beta^2$; entonces la expresión

$$P(\mid X - \mu_X \mid \geq k\sigma_X)$$

Puede expresarse de la siguiente manera

$$P(\mid X - \beta_X \mid \geq k\beta_X)$$

Sea $t = k\beta_X$, entonces

$$P(\mid X - \beta_X \mid \geq t)$$

De acuerdo con la desigualdad de Chebyshev

$$P(\mid X - \beta_X \mid \geq t) \leq \frac{\sigma^2}{t^2}$$

Reemplazando los valores de σ y t

$$P(\mid X - \beta_X \mid \geq k\beta_X) \leq \frac{\beta_X^2}{k^2\beta_X^2}$$

Por lo tanto

$$P(\mid X - \beta_X \mid \geq k\beta_X) \leq \frac{1}{k^2}$$

La probabilidad $P(\mid X-\mu_X \mid \ \geq \ k\sigma_X \ )$ en una distribución exponencial $Exponencial(\beta)$, será menor o igual que $\frac{1}{k^2}$, de acuerdo con la desigualdad de Chebyshev.

Solución

b) Sean $\mu = p$, $\sigma^2 = V(\bar{X}_n)$ y $t = \epsilon$; entonces

$$P(\mid \bar{X}_n - p \mid > \epsilon) \leq \frac{V(\bar{X}_n)}{\epsilon^2}$$

Sabemos que $V(\bar{X}_n) = \frac{V(X_i)}{n} = \frac{p(1-p)}{n}$, por lo tanto

$$P(\mid \bar{X}_n - p \mid > \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \qquad \epsilon \ > \ 0$$

Dado que

$$\frac{d}{dp}\ p(1-p) = 1-2p \qquad \frac{d}{dp}\ 1-2p = -2$$

Entonces

$$f(\frac{1}{2}) = 1-2(\frac{1}{2}) = 1-1 = 0$$

$\frac{1}{2}$ es un máximo; por lo tanto $p(1-p) \leq \frac{1}{4}$

Finalmente podemos expresar la desigualdad de la siguiente manera

$$P(\mid \bar{X}_n - p \mid > \epsilon) \leq \frac{1}{4n\epsilon^2} \qquad \epsilon \ > \ 0$$

desigualdad de Chebyshev

Por otro lado, tenemos que para un conjunto de variables $X_1, X_2, ..., X_n \sim Bernoulli(p)$ **independientes**

$$P(\mid \bar{X}_n - p \mid > \epsilon) \leq 2e^{-2n\epsilon^2} \qquad \epsilon \ > \ 0$$

desigualdad de Hoeffding

Dado que queremos comprobar que la cota de Hoeffding es más pequeña que la cota de Chebyshev, podemos comparar ambas expresiones de manera directa. En particular queremos demostrar que

$$2e^{-2n\epsilon^2} < \frac{1}{4n\epsilon^2}$$

Simplificando la desigualdad

$$e^{-2n\epsilon^2} < \frac{1}{8n\epsilon^2}$$

Tomando el logaritmo natural de ambos lados, y dado que $log_b\ x^n = nlog_b\ x$; entonces

$$-2n\epsilon^2 < ln\Big(\frac{1}{8n\epsilon^2}\Big)$$

Finalmente

$$-n\epsilon^2 < \frac{1}{2}\ ln\Big(\frac{1}{8n\epsilon^2}\Big)$$

Podemos observar que, a medida que n aumenta, la expresión $-n\epsilon^2$ decrece en mayor medida que la expresión $ln(\frac{C}{n\epsilon^2})$. De esta manera podemos decir que la cota de Hoeffding es más pequeña que la cota de Chebyshev para valores de n grandes.

Esto implica que la cota de Hoefdding brinda menor margen de error y por ende, es más precisa al encontrar la probabilidad de que la media muestral $\bar{X}_n$ sea mayor dado un valor $\epsilon > 0$, cuando el tamaño de la muestra $X_1, ..., X_n \sim Bernoulli(p)$ aumenta.

2. Sean $X_1, ..., X_n \sim Bernoulli(p)$.

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

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

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

$$P(C_n \ contiene \ a \ p) \geq 1 - \alpha$$

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

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.

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?

Solución

a) Como se expuso con anterioridad, sea $X_1, ..., X_n \sim Bernoulli(p)$ una muestra de variables **independientes**, entonces, para cualquier $\epsilon > 0$

$$P(\mid \bar{X}_n - p \mid > \epsilon_n) \leq 2e^{-2n\epsilon_n^2}$$

Sea $\hat{p}_n = n^{-1} \sum_{i=1}^{n} X_i$

$$P(\mid \hat{p}_n - p \mid > \epsilon_n) \leq 2e^{-2n\epsilon_n^2}$$

Reemplazando el valor $\epsilon_n = \sqrt{\frac{1}{2n} log(\frac{2}{\alpha})}$

$$P(\mid \hat{p}_n - p \mid > \epsilon_n) \leq 2e^{-2n\left(\sqrt{\frac{1}{2n} log(\frac{2}{\alpha})}\right)^2}$$

Simplificando la expresión

$$P(\mid \hat{p}_n - p \mid > \epsilon_n) \leq 2e^{-2n\left(\frac{1}{2n} log(\frac{2}{\alpha})\right)}$$

$$P(\mid \hat{p}_n - p \mid > \epsilon_n) \leq 2e^{-log(\frac{2}{\alpha})}$$

Recordando la siguiente propiedad logarítmica

$$log_a(\frac{b}{c}) = log_a(b) - log_a(c)$$

Entonces

$$P(\mid \hat{p}_n - p \mid > \epsilon_n) \leq 2e^{log(\alpha) - log(2)}$$

Aplicando la siguiente propiedad de las funciones exponenciales

$$a^{x-y} = \frac{a^x}{a^y}$$

$$P(\mid \hat{p}_n - p \mid > \epsilon_n) \leq 2\left(\frac{e^{log(\alpha)}}{e^{log(2)}}\right)$$

Nuevamente, al considerar la propiedad de las funciones exponenciales

$$a^{log(r)} = r$$

$$P(\mid \hat{p}_n - p \mid > \epsilon_n) \leq 2(\frac{\alpha}{2})$$

Finalmente

$$P(\mid \hat{p}_n - p \mid > \epsilon_n) \leq \alpha$$

Podemos interpretar la probabilidad $P(\mid \hat{p}_n - p \mid > \epsilon_n)$ como la probabilidad de que el valor p sea mayor que ϵ_n y por lo tanto, que $p \notin C_n$. De esta manera

$$P(\mid \hat{p}_n - p \mid > \epsilon_n) = P(p \notin C_n)$$

Por lo tanto

$$P(p \notin C_n) \leq \alpha$$

Por último, si consideramos $P(p \in C_n)$ como el complemento de $P(p \notin C_n)$, podemos expresar la probabilidad de la siguiente manera

$$P(p \in C_n) \geq 1 - \alpha$$

Resultados

El gráfico muestra cómo la cobertura del intervalo de confianza varía conforme aumenta el tamaño de muestra n. Para tamaños de muestra pequeños ($n = 10,\ 50,\ 100$), la cobertura presenta una mayor dispersión, mientras que para tamaños de muestra grandes ($n = 2500, 5000, 10000$), la cobertura empieza a converger a valores entre $0.992 - 0.994$. Por lo que podríamos concluir que con tamaños de muestra pequeños la estimación del intervalo de confianza es menos precisa.

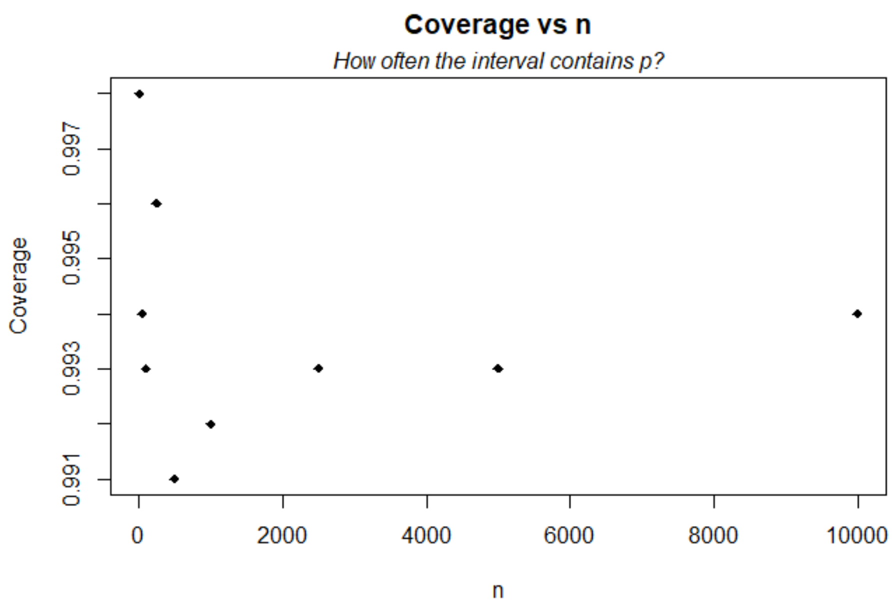


Gráfico de cobertura versus n. Con que frecuencia el intervalo de confianza contiene a p.

De manera general observamos que la longitud del intervalo disminuye conforme el tamaño de muestra n aumenta. A partir del gráfico generado podemos observar en que valor de n la longitud del intervalo es menor a 0.05. En este caso, el valor de n para el cual la longitud del intervalo es menor a 0.05 se encuentra alrededor de $n \approx 3000$.

En conclusión podemos decir que a medida que aumenta el tamaño de muestra, la longitud del intervalo de confianza se vuelve más estrecho.

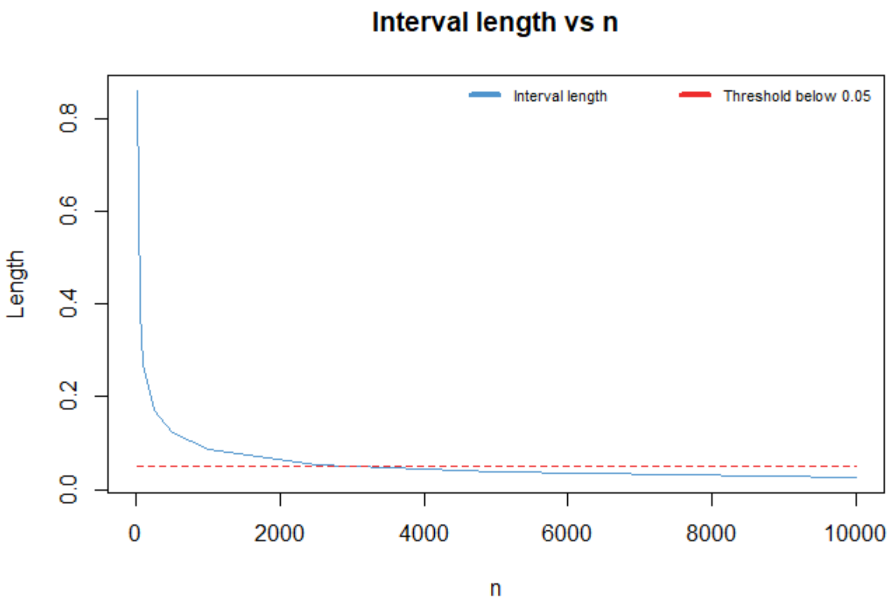
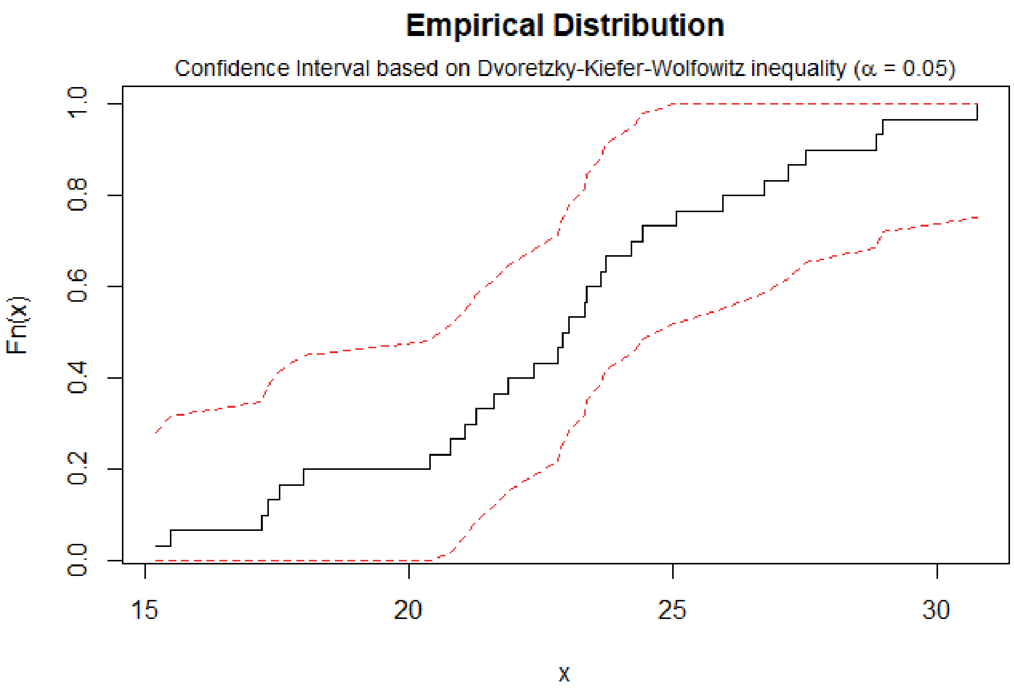


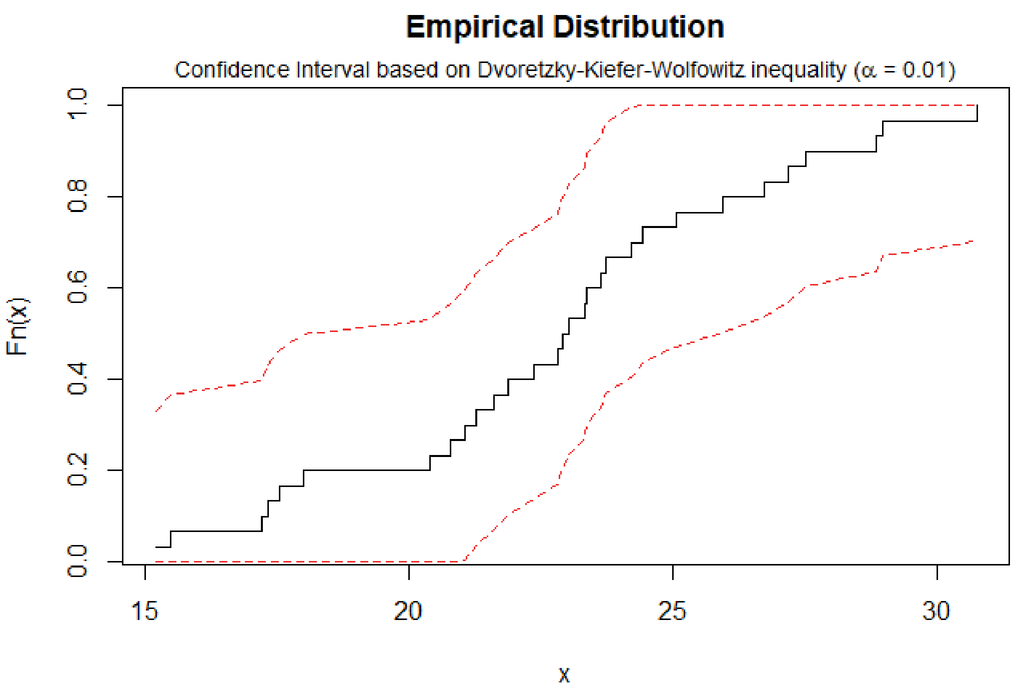
Gráfico de longitud del intervalo versus n indicando región donde la longitud es menor que 0.05.

3. Considera el problema 5 de la Tarea 3. 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 construir la función de distribución empírica.

Resultados



a) Intervalo de confianza Dvoretzky-Kiefer-Wolfowitz $\alpha = 0.05$.



b) Intervalo de confianza Dvoretzky-Kiefer-Wolfowitz $\alpha = 0.01$.

4. En este ejercicio repasará la estimación de densidades.

a) Escriba una función en R que estime una densidad por el método de kerneles. 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.

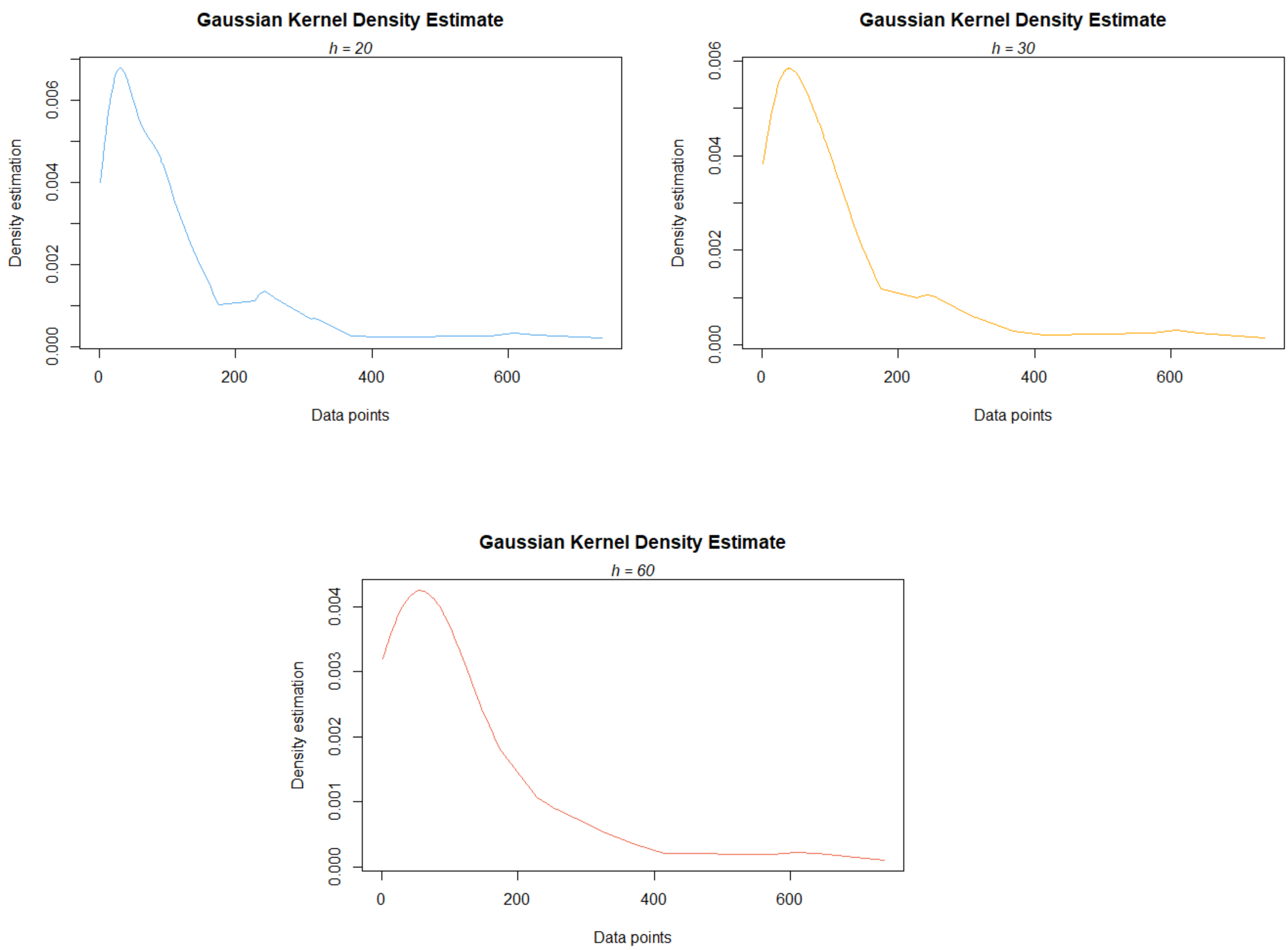
b) Cargue en R al archivo *Tratamiento.csv*, el cual contiene la duración de los períodos de tratamiento (en días) de los pacientes de control en un estudio de suicidio. Utilice la función del inciso anterior para estimar la densidad del conjunto de datos para h = 20, 30, 60. Grafique las densidades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

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*.

Resultados

A simple vista podemos considerar que la curva de densidad cuando $h = 60$ sobreajusta los datos, esto en comparación con las gráficas obtenidas cuando $h = 20$ y $h = 30$. Por otro lado, podríamos considerar que la mejor estimación de la curva de densidad es cuando $h = 30$, esto ya que es el valor que proporcione una estimación más suave de la densidad.

Estimación de Densidad por el método de Kernel (KDE)



5. Considera el siguiente experimento en dos etapas: primero se escoge un punto X con distribución uniforme en (0, 1); despúes se elige un punto Y con distribución uniforme en (-X, X). El vector aleatorio (X, Y) representa el resultado del experimento. ¿Cúal es su densidad conjunta de (X, Y)? ¿Cúal es la densidad marginal de Y? ¿Cúal es la densidad condicional de X dada Y?

Solución

Dado que X tiene una distribución uniforme $X \sim Uniforme(0,1)$, su función de densidad es constante, y puede ser expresada de la siguiente manera

$$f_X(x) = \begin{cases} 1 & \text{si } 0 \leq x \leq 1 \\ 0 & \text{otro caso} \end{cases}$$

Por otro lado, la función de densidad $f_{Y|X}(y|x)$ se puede expresar como

$$f_{Y \mid X}(y \mid x) = \begin{cases} \frac{1}{2x} & \text{si } -X \leq y \leq X \text{ y } 0 \leq x \leq 1 \\ 0 & \text{otro caso} \end{cases}$$

Para variables aleatorias continuas la función de densidad de probabilidad conjunta esta definida de la siguiente manera

$$f_{X,Y}(x,y) = f_{Y \mid X}(y \mid x)\, f_X(x) = f_{X \mid Y}(x \mid y)\, f_Y(y)$$

donde $f_{Y \mid X}(y \mid x)$ es la probabilidad condicional de Y dado X, $f_{X \mid Y}(x \mid y)$ es la probabilidad condicional de X dado Y, y $f_X(x)$, $f_Y(y)$ son las densidades marginales de X y Y, respectivamente.

Por lo tanto la densidad condicional de Y dado X esta definida como

$$f_{Y \mid X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Sabemos que la densidad marginal de X $f_X(x)$ es 1 con $0 \leq x \leq 1$, por lo tanto

$$f_{Y \mid X}(y \mid x) = \frac{f_{X,Y}(x,y)}{1} = f_{X,Y}(x,y)$$

Entonces

$$f_{Y \mid X}(y \mid x) = f_{X,Y}(x,y)$$

De esta manera tenemos que la **densidad conjunta** $f_{X,Y}(x,y)$ esta definida como

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2x} & \text{si } 0 \leq x \leq 1 \text{ y } -X \leq y \leq X \\ 0 & \text{otro caso} \end{cases}$$

Comprobamos que la función de densidad conjunta $f_{X,Y}(x,y)$ satisface

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx\, dy = 1$$

Tenemos que

$$\int_0^1 \int_{-X}^{X} \frac{1}{2x} dx\, dy$$

$$\int_0^1 \frac{x-(-x)}{2x} dx$$

$$\int_0^1 \frac{2x}{2x} dx$$

$$\int_0^1 dx$$

Finalmente

$$x\Big|_0^1 = 1 - 0 = 1$$

Por lo tanto se cumple que

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2x} dx\, dy = 1$$

Las distribuciones marginales de las variables aleatorias continuas X y Y estan definidas de la siguiente manera

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Por lo que podemos calcular la distribución marginal $f_Y(y)$ de la siguiente manera

$$f_Y(y) = \int_{-\infty}^{\infty} \frac{1}{2x} dx$$

Dado que el intervalo de la densidad $f_{X,Y}(x,y)$ se encuentra entre $-x \leq y \leq x$ con $x \in (0,1)$ entonces podemos considerar que $-1 \leq y \leq 1$, lo cual podemos expresarlo como $\mid y \mid \leq 1$, de manera similar podemos decir que $\mid y \mid \leq x$, por lo que al hacer el cambio de coordenadas para calcular la densidad marginal de Y podemos considerar el intervalo de x como $\mid y \mid \leq x \leq 1$.

De esta manera tenemos que

$$f_Y(y) = \int_{|y|}^{1} \frac{1}{2x} dx$$

Resolviendo la integral tenemos que

$$f_Y(y) = \frac{1}{2}\int_{|y|}^{1} \frac{1}{x} dx$$

$$f_Y(y) = \frac{1}{2} ln(x)\Big|_{|y|}^{1}$$

Ya que consideramos el valor absoluto de y, multiplicamos por 2 el resultado final, entonces

$$f_Y(y) = ln(x)\Big|_{y}^{1}$$

Evaluando la integral

$$f_Y(y) = ln(1) - ln(y)$$

$$f_Y(y) = -ln(y)$$

De esta manera expresamos la **densidad marginal** $f_Y(y)$ como

$$f_Y(y) = \begin{cases} -ln(y) & \text{si } 0 \leq y \leq 1 \\ 0 & \text{otro caso} \end{cases}$$

Conociendo la densidad marginal $f_Y(y)$ y la densidad conjunta $f_{X,Y}(x,y)$, podemos calcular la densidad condicional de X dada Y de la siguiente manera

$$f_{X \mid Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Reemplazando

$$f_{X \mid Y}(x \mid y) = \frac{\frac{1}{2x}}{-ln(y)}$$

Finalmente tenemos que la **densidad condicional** $f_{X \mid Y}(x \mid y)$ esta definida como

$$f_{X \mid Y}(x \mid y) = -\frac{1}{2x\, ln(y)}$$

6. Cargue en R al conjunto de datos *Maíz.csv*, el cual contiene el precio mensual de la tonelada de maíz y el precio de la tonelada de tortillas en USD. En este ejercicio tendrá que estimar los coeficientes de una regresión lineal simple.

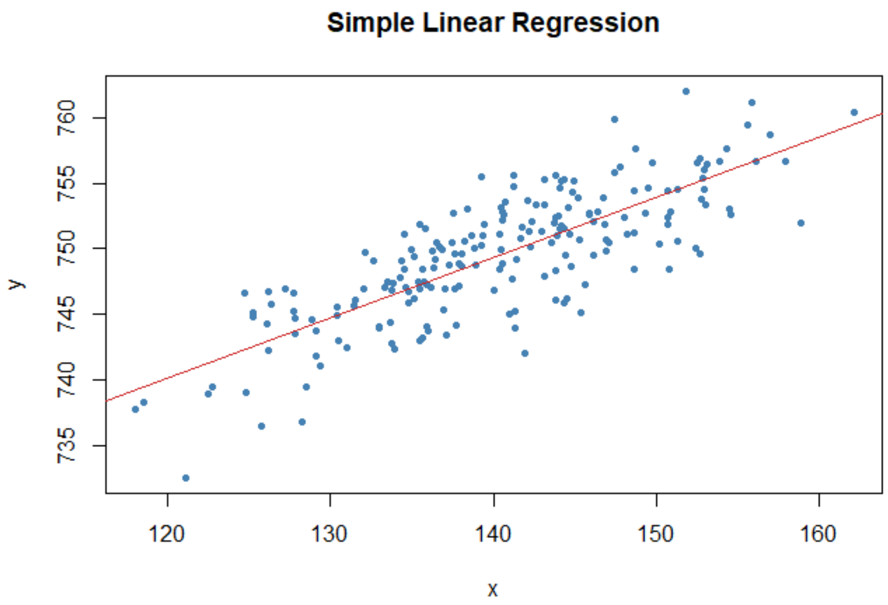
a) Calcule de forma explícita la estimación de los coeficientes via mínimos cuadrados y ajuste la regresión correspondiente. Concluya.

b) Calcule de forma explícita la estimación de los coeficientes via regresión no-paramétrica tipo kernel (ver Nadaraya, E. A. (1964). "On Estimating Regression". Theory of Probability and its Applications. 9 (1): 141-2. doi:10.1137/1109020) y ajuste la regresión correspondiente. Concluya.

c) Compare ambos resultados. ¿Qué diferencias observa?

Resultados

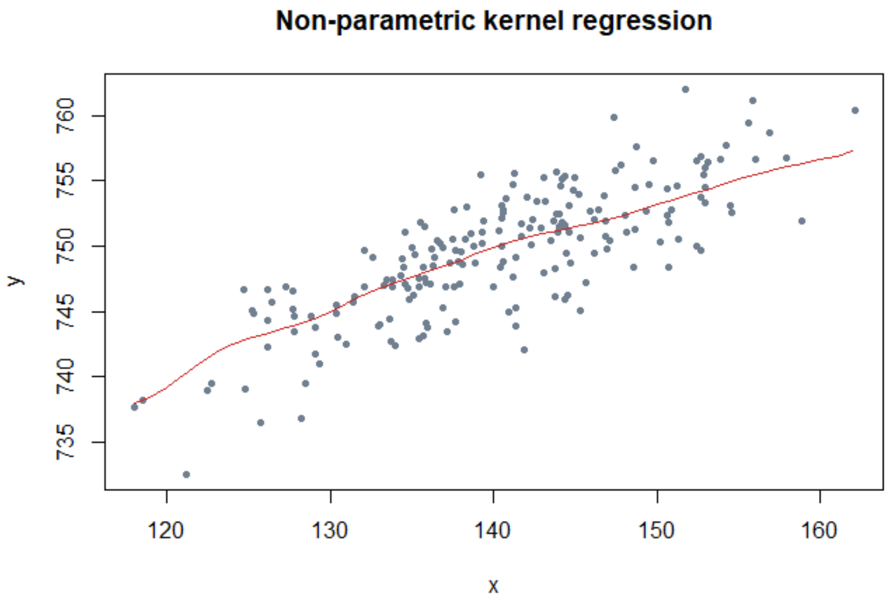
Podemos concluir que existe una relación positiva entre el precio mensual de la tonelada de maíz (x) y el precio mensual de la tonelada de la tortilla (y), esto basándonos en los valores de los coeficientes $\beta_0 = 684.95, \beta_1 = 0.46$; donde β_0 representa la intersección de la recta con el eje y, mientras que β_1 representa el valor de la pendiente de la recta (cómo cambia Y al incrementar X en una unidad).



Regresión lineal simple. Precio mensual de la **tonelada de maíz** (x) en relación con el precio mensual de la **tonelada de tortilla** (y).

A diferencia de la regresión lineal simple, la regresión no paramétrica aproxima un modelo que se ajusta mejor a los datos de acuerdo a como estan distribuidos. Esto coincide con los valores de los coeficientes obtenidos en el modelo de regresión lineal simple, donde el valor del coeficiente $\beta_1 = 0.46$, lo cual si bien indica una relación positiva, no es del todo correcto afirmar que existe una relación 1:1, por lo que el modelo ajustado a los datos no puede ser del todo una linea recta como se estima en la regresión lineal simple.

En general, podemos concluir que la regresión no paramétrica tipo kernel estima con mayor precisión el modelo que se ajusta a los datos.



Regresión no paramétrica tipo kernel. Precio mensual de la **tonelada de maíz** (x) en relación con el precio mensual de la **tonelada de tortilla** (y).

7. Se requiere verificar la programación del tiempo que debe durar la luz verde en un semáforo que se encuentra en una cierta intersección, con vuelta a la izquierda. Como no se tiene información al respecto, se envía a un estudiante graduado a hacer observaciones sobre el número de carros X y el número de camiones Y que llegan en un ciclo (entre verde y verde); y con esto se construye la tabla de distribuciones conjunta. Se plantean una serie de preguntas y con ello confirman qué tan eficiente fue el ciclo planeado.

Los resultados fueron los siguientes

		x		
		0	1	2
y	0	0.025	0.015	0.010
	1	0.050	0.030	0.020
	2	0.125	0.075	0.050
	3	0.150	0.090	0.060
	4	0.100	0.060	0.040
	5	0.050	0.030	0.020

a) Verifica que es una tabla válida de probabilidades conjuntas.

Para comprobar que es una tabla legítima de probabilidades conjuntas, debemos verificar que se cumple lo siguiente

$$0 \leq f(x,y) \leq 1 \forall (x,y) \qquad \sum_x \sum_y f(x,y) = 1$$

En este caso podemos comprobar fácilmente que para cada $\forall (x,y) 0 \leq f(x,y) \leq 1$. Por otro lado, si sumamos todas las probabilidades $f(x,y)$ para $x = 0,1,2$ y $y = 0,1,2,3,4,5$; es fácil corroborar que

$$\sum_x \sum_y f(x,y) = 1$$

b) Calcula las distribuciones marginales para carros y camiones.

		x			
		0	1	2	$P\{X = x\}$
y	0	0.025	0.015	0.010	0.050
	1	0.050	0.030	0.020	0.100
	2	0.125	0.075	0.050	0.250
	3	0.150	0.090	0.060	0.300
	4	0.100	0.060	0.040	0.200
	5	0.050	0.030	0.020	0.100
	$P\{Y = y\}$	0.500	0.300	0.200	**1.00**

c) ¿Cuál es la probabilidad de que se tengan exactamente un carro y un camión en un ciclo dado?

La probabilidad de que se tengan exactamente un carro y un camión en un ciclo dado esta dada por

$$f(1,1) = 0.030$$

d) Supongamos que la vuelta a la izquierda tiene una capacidad máxima de 5 carros y un camíon es equivalente a 3 carros. ¿Cúal es la probabilidad de que se sature la línea en un ciclo dado?

Dado que nos interesa la probabilidad

$$P(X > 5) = P(Y > 1)$$

Por lo tanto la probabilidad de que se sature la línea en un ciclo dado es

$$P(X > 5) = 0.250 + 0.300 + 0.200 + 0.100 = 0.850$$

e) ¿Cuál es la probabilidad de que se tenga exactamente un carro, dado que ya se tiene un camión en un ciclo predeterminado y se sabe que no habrá ningún otro camión más? Contesta esta misma pregunta para $Y = 0, 1, 2, 3, 4, 5$.

Sean X y Y dos variables aleatorias con distribución conjunta $f(x,y)$, y marginales definidas por $f_X(x)$ y $f_Y(y)$. Entonces la distribución condicional de X dado $Y = y$ es

$$P\{X = x|Y = y\} = \frac{f(x,y)}{f_Y(y)}$$

Por lo tanto podemos calcular la probabilidad $P\{X = 1|Y = 1\}$ de la siguiente manera

$$P\{X = 1|Y = 1\} = \frac{f(1,1)}{f_Y(1)} = \frac{0.030}{0.100} = 0.300$$

De manera similar, calculamos las probabilidades de que se tenga exactamente un carro dado que se tienen y camiones sustituyendo las distribuciones conjuntas $f(x,y)$ y marginales $f_Y(y)$ correspondientes.

$$P\{X = 1|Y = 0\} = \frac{f(1,0)}{f_Y(0)} = \frac{0.015}{0.050} = 0.300$$

$$P\{X = 1|Y = 2\} = \frac{f(1,2)}{f_Y(2)} = \frac{0.075}{0.250} = 0.300$$

$$P\{X = 1|Y = 3\} = \frac{f(1,3)}{f_Y(3)} = \frac{0.090}{0.300} = 0.300$$

$$P\{X = 1|Y = 4\} = \frac{f(1,4)}{f_Y(4)} = \frac{0.060}{0.200} = 0.300$$

$$P\{X = 1|Y = 5\} = \frac{f(1,5)}{f_Y(5)} = \frac{0.030}{0.100} = 0.300$$

Observamos que

$$P\{X = 1|Y = y\} = P\{X = 1\}$$

Por lo que podemos concluir que la variable aleatoria Y no agrega información adicional sobre el comportamiento de la variable aleatoria X. Por lo tanto podemos afirmar que X y Y son independientes.