MAE $\Rightarrow$ Ejercicios 2

Guillermo Hoyo Bravo

6

a) Sesgo y Varianza

Con $K(x) \Rightarrow \int_x K^2(x)\,dx < \infty \Rightarrow$ si se puede derivar y es continua $\Rightarrow$

$$\Rightarrow E(\hat{f}(x)) = E\left[\frac{1}{nh}\sum_i K\left(\frac{x-x_i}{h}\right)\right] = \frac{1}{h} E\left[K\left(\frac{x-x_i}{h}\right)\right]$$

$$= \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{x-t}{h}\right) f(t)\,dt \underset{\substack{w=\frac{x-t}{h} \\ dw = \frac{-dt}{h}}}{\Longrightarrow} \frac{1}{h}\int_{-\infty}^{\infty} K(w) f(x-wh)\, h(-dw) =$$

$$= -\int_{-\infty}^{\infty} K(w) f(x-wh)\,dw \qquad \text{si } h \to 0 \simeq f(x)\underbrace{\int_{-\infty}^{\infty} K(w)\,dw}_{1} = f(x)$$

- $\underline{\text{Sesgo}}(\hat{f}(x)) = E[\hat{f}(x)] - f(x) = f(x) - f(x) = 0$

Aplicando Taylor $\simeq -\int_{-\infty}^{\infty} K(w)\left[f(x) - wh f'(x) + \frac{w^2h^2}{2} f''(x)\right] dw$

$$= -\underbrace{\int_{-\infty}^{\infty} K(w) f(x)\,dw}_{f(x)=1} + \underbrace{\cancel{h\int_{-\infty}^{\infty} K(w)\,w f'(x)\,dw}}_{\substack{hf'(x)\,E[w]=0 \\ \text{Kernel insesgado}}} - \underbrace{\int f''(x) K(w)\frac{w^2h^2}{2}\,dw}_{\frac{h^2 f''}{2}(x)\ Var(w) = \sigma^2 K}$$

$$= f(x) + \frac{h^2}{2} f''(x)\,\sigma^2 K$$

$$\Rightarrow \text{Sesgo}^2(\hat{f}(x)) = \frac{h^2}{2} f''(x)\,\sigma^2_K$$

- $Var(\hat{f}(x)) = Var\left(\frac{1}{nh}\sum_i K\left(\frac{x-x_i}{h}\right)\right) = \frac{1}{nh^2} Var\left(K\left(\frac{x-x_i}{h}\right)\right)$

$Var\left(K\left(\frac{x-x_i}{h}\right)\right) = E\left[K^2\left(\frac{x-x_i}{h}\right)\right] - \underbrace{E\left[K\left(\frac{x-x_i}{h}\right)\right]}_{\substack{\to f(x) \\ si\, h \to 0}}^2 \approx E\left[K^2\left(\frac{x-x_i}{h}\right)\right] - f(x)$

Si h es grande $\Rightarrow \frac{1}{nh^2}\left(E\left[K^2\left(\frac{x-x_i}{h}\right)\right] - \cancel{f(x)}\right) = \frac{1}{nh^2} E\left[K^2\left(\frac{x-x_i}{h}\right)\right]$

$\xRightarrow[w = \frac{x-t}{h}]{} \frac{1}{h^2 n} E\left[K^2(w)\right] = \frac{1}{nh^2}\int_{-\infty}^{\infty} K^2(w) f(x - wh)\, dw \cdot \cancel{h}$

$\because \int_{-\infty}^{\infty} K^2 < \infty \Rightarrow \frac{1}{nh} \to \infty \Rightarrow Var(\ ) \to 0$

Si h es $\frac{1}{\sqrt{n}}$

$h \longrightarrow 0 \quad \frac{1}{hn} \to \infty$

$\Downarrow \qquad\qquad \Downarrow$

$Sesgo \to 0 \quad Var(\ ) \to 0$

Aplicando el teorema de Taylor:

$\approx \frac{1}{nh}\int K^2(w)\left[f(x) - hwf'(x) + \frac{h^2w^2}{2} f''(x)\right] dw$

$0 \left\{ \begin{array}{l} \text{al dividir por} \\ wh \to \infty = 0 \end{array} \right.$

$= \frac{1}{nh}\left(\int K^2(w) f(w)\, dw - \underbrace{\cancel{\int hwf'(x) K^2(x)\, dw}}_{0} + \int K^2(w) \frac{h^2w^2}{2} f''(x) dw\right.$

$= \frac{1}{nh} f(x) \int_{-\infty}^{\infty} K^2(w)\, dw = \frac{1}{nh} dK$

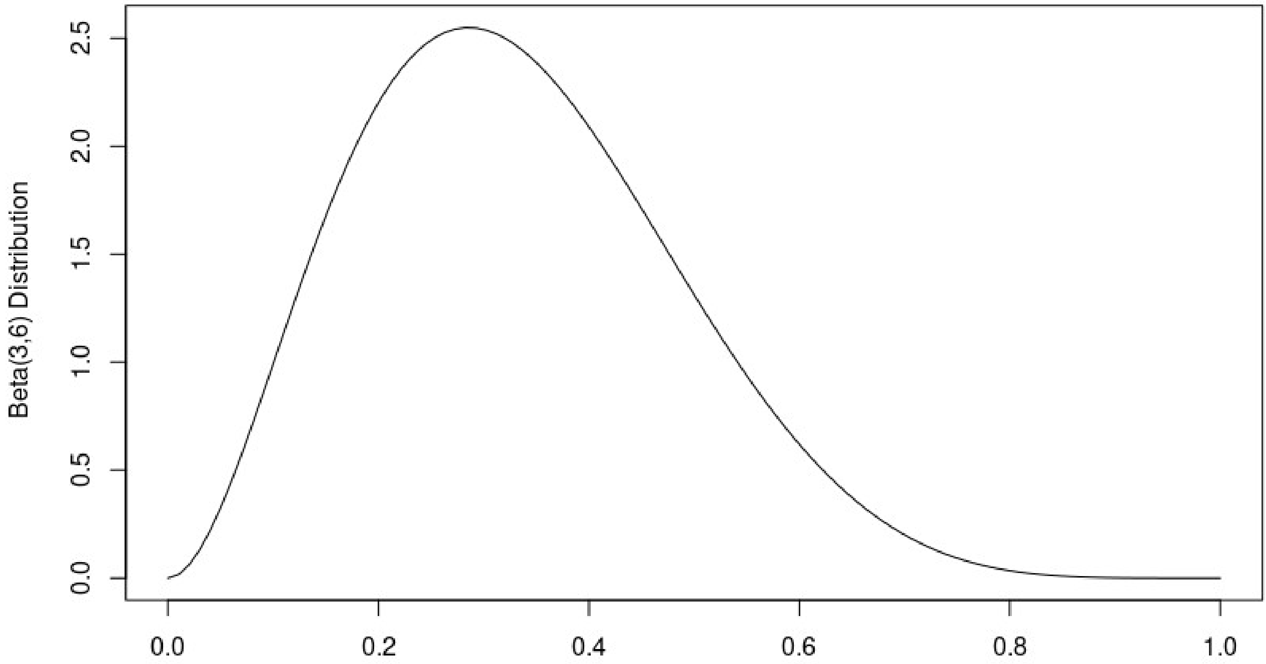
$ECMI(\hat{f}(x)) \to \frac{d\, ECMI(\hat{f}(x))}{d} = \frac{4h^3}{4}\sigma_K^4 \|f''\|_2^2 + \frac{\|K\|_2^2}{nh}$

MAE – Ejercicios – 2º tanda.
Ejercicios: estimación no paramétrica de funciones de densidad

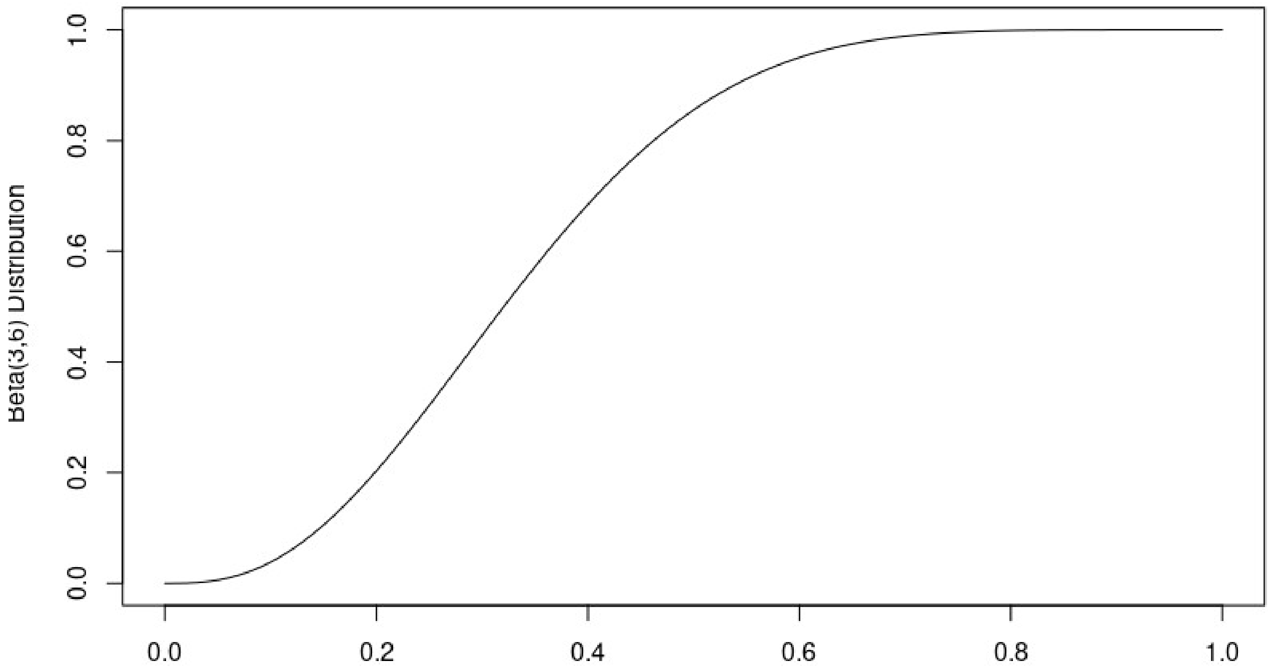
7. Considera una variable aleatoria con distribución beta de parámetros α = 3, β = 6.

a Representa gráficamente la función de densidad y la función de distribución.

Función de Densidad:



Función de Distribución:

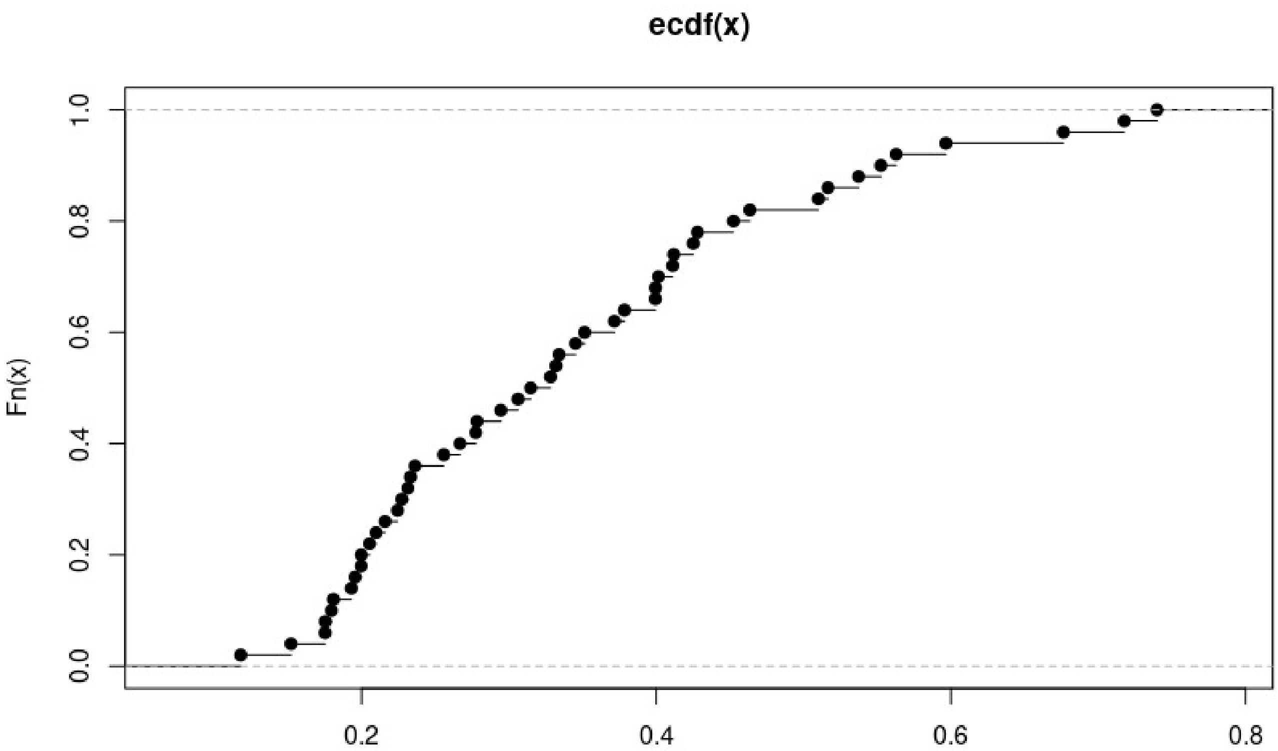


b Simula una muestra de tamaño 20 de esta distribución. A continuación, representa en los mismos gráficos del apartado (a) las estimaciones de F y f obtenidas respectivamente mediante la función de distribución empírica Fn y un estimador del núcleo ˆf obtenidos a partir de la muestra simulada.

Codigo CDF:

```
n <- 50
x <- rbeta(n,3,6)
f_ecdf <- ecdf(x)
plot(f_ecdf)
```

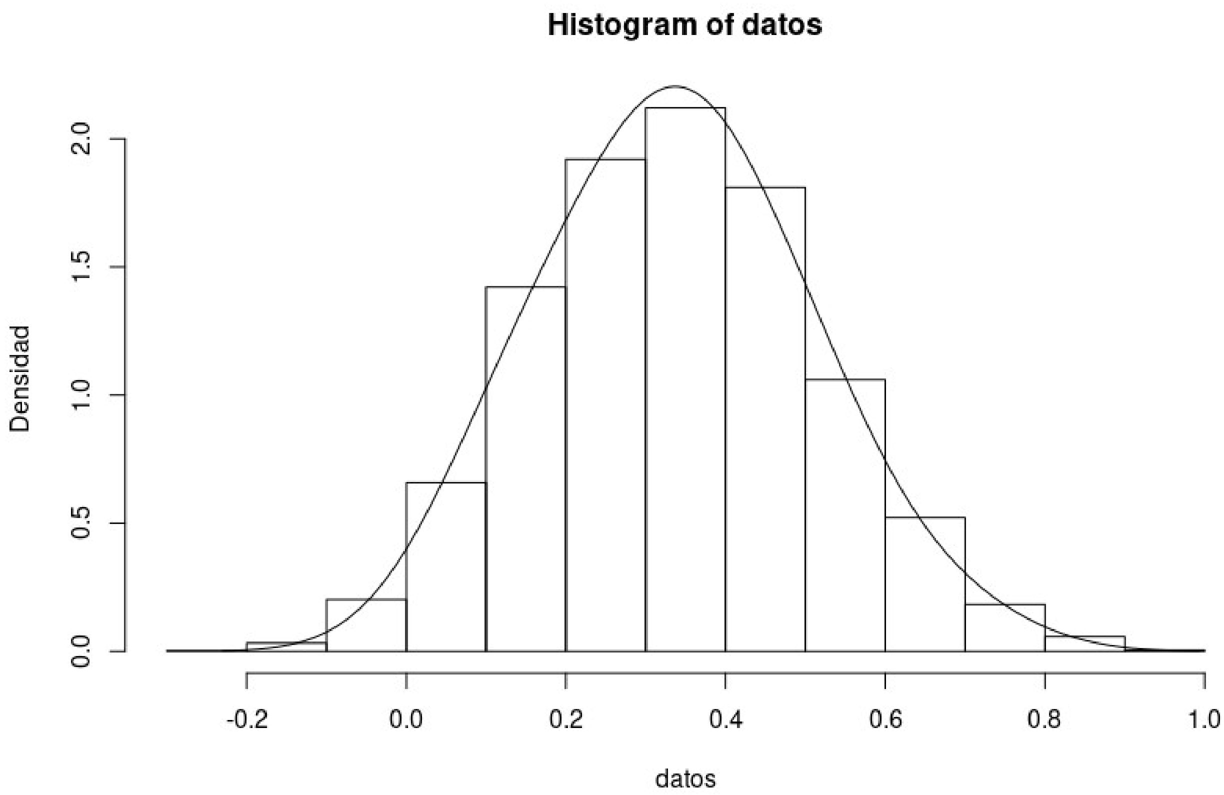
Gráfica CDF:



PDF Estimador del Nucleo:

```r
rcore <- function(n, muestra, h){
y = sample(muestra, n, rep = TRUE) + rnorm(n, sd = h)
return(y)
}

h<- 0.1
muestra <- rbeta(n,3,6)
d <- rcore(10000, muestra, h)
hist(d, freq = FALSE, ylab = 'Frecuencia')
```

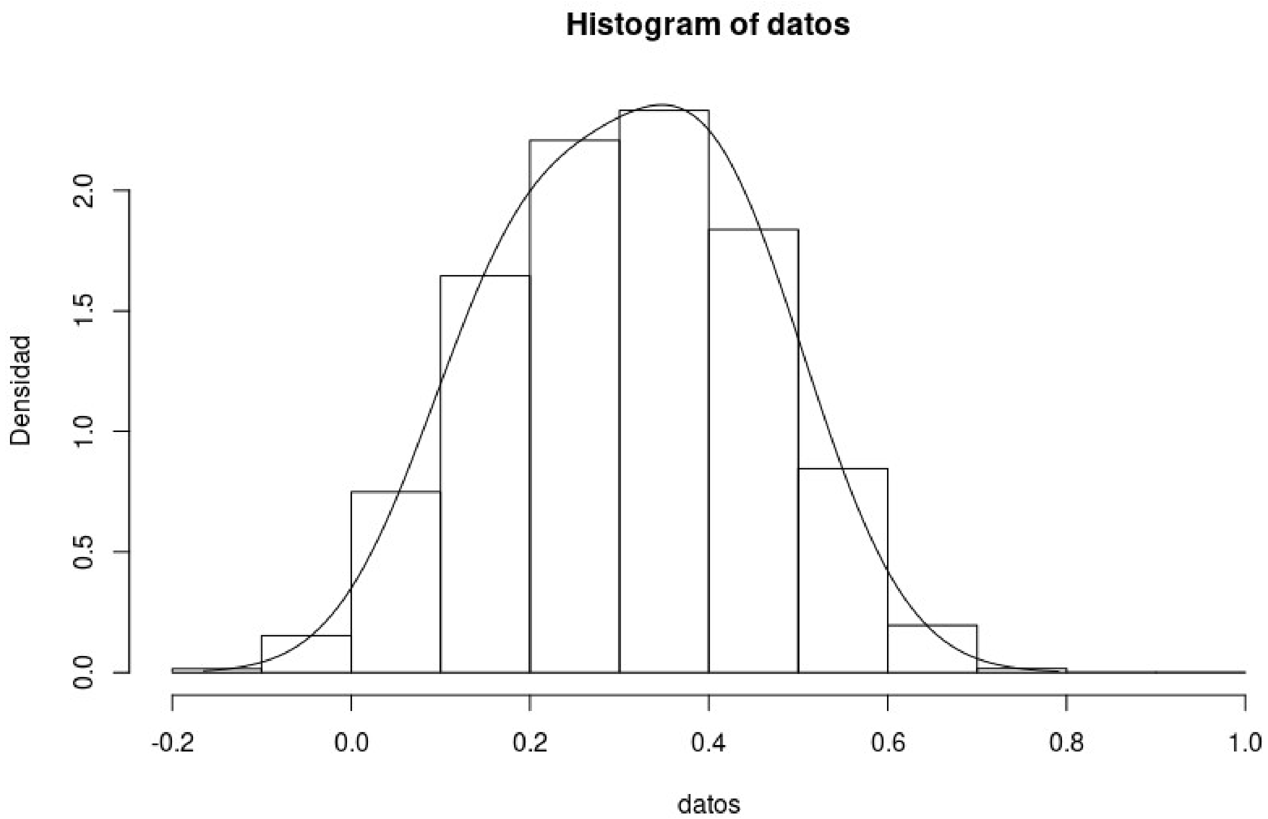


Código Gráfica Estimador del Nucleo:

```
rcore <- function(n, muestra, h){
y = sample(muestra, n, rep = TRUE) + rnorm(n, sd = h)
return(y)
}

h<- 0.1
muestra <- rbeta(n,3,6)
d <- rcore(10000, muestra, h)
hist(d, freq = FALSE, ylab = 'Densidad')

estimador_nucleo <- density(muestra,bw = h)
lines(estimador_nucleo$x, estimador_nucleo$y)
```



c Verifica empíricamente el grado de aproximación alcanzado en las estimaciones de F y f. Para ello, genera 200 muestras de tamaño 20 y para cada una de ellas evaluar el error (medido en la norma del supremo, es decir, el máximo de las diferencias entre las funciones) cometido al aproximar F por Fn y f por ˆf. Por ´ultimo, calcula el promedio de los 200 errores obtenidos.

Ahora se compara el estimador y el resultado teórico con la máxima distancia de cada repretición.

Código:

```
errs <- NULL
repeticiones <- 200
n <-20

for(i in 1:reps){
  muestra<-rbeta(n,3,6)
  dens<-density(muestra,kernel="gaussian")

  real<-dbeta(dens[['x']],3,6)
  x<-rbind(dens[["y"]],real)

  errs<-rbind(errs, suppressMessages(distance(x,method="chebyshev")))
}

mean(errs)
```

Resultado:

Tras la simulación es obtenido el valor promedio de errores: **0.8136935**

Guillermo Hoyo Bravo