Inferencia estadística

Tarea 1

José Antonio García Ramirez

September 13, 2017

Ejercicio 2

Ej. 13, cap. 2, Wasserman. Sea X $N(0,1)$, y define $Y = e^X$.

b) Genera $10,000$ v.a. normales estandar y obten las correspondientes $10,000$ y's como se indica en el problema. Dibuja el histograma de Y, y comparalo con la PDF que encontraste en el inciso anterior.

Simulamos la muestra de tamaño $10,000$ de X, fijamos la semilla por reproducibilidad

```
set.seed(0)
X <- rnorm(n = 10000, 0 , 1)        # Se generan los datos
Y <- exp(X)
```

Procedemos a visualizar la distribución.

```
Y.frame <- data.frame(x=Y)
p <- ggplot(Y.frame) +              # se genera el histograma con soporte [0,35]
  geom_histogram(aes(x, ..density.., fill = I('Muestral \n 10,000 obs')), bins = 100) +
  theme_minimal() +
  stat_function(fun = function(x) {(1/x)*( 1/ sqrt(2 * pi))*exp(-( (log(x))^2 ) / 2 )},
              aes(colour = I('Teorico')), size = 1 ) +
  scale_fill_manual(name='',values=I('#FEEBE2')) +
  scale_colour_manual(name='',values=I('#7A0177') ) +
  ggtitle(TeX('$Y \\sim Lognormal( \\mu = 0, \\sigma =1  )$')) +
  theme(plot.title = element_text(hjust = 0.5)) +
  ylab('') + ylim(0,.66)
p1 <- ggplot(Y.frame) +             # se genera el histogrma haciendo zoom alrededor de la moda
  geom_histogram(aes(x, ..density.., fill = I('Muestral \n 10,000 obs')), bins = 100) +
  theme_minimal() +
  stat_function(fun = function(x) {(1/x)*( 1/ sqrt(2 * pi))*exp(-( (log(x))^2 ) / 2 )},
              aes(colour = I('Teorico')), size = 1 ) +
  scale_fill_manual(name='',values=I('#FEEBE2')) +
  scale_colour_manual(name='',values=I('#7A0177') ) +
  ggtitle(TeX('$Y \\sim Lognormal( \\mu = 0, \\sigma =1  )$')) +
  theme(plot.title = element_text(hjust = 0.5)) +
  ylab('') + xlim(0,10)
p <- arrangeGrob(p, p1, ncol = 2)
as_ggplot(p) #muestro dos imagenes de la distribución muestral solo cambia el soporte
```

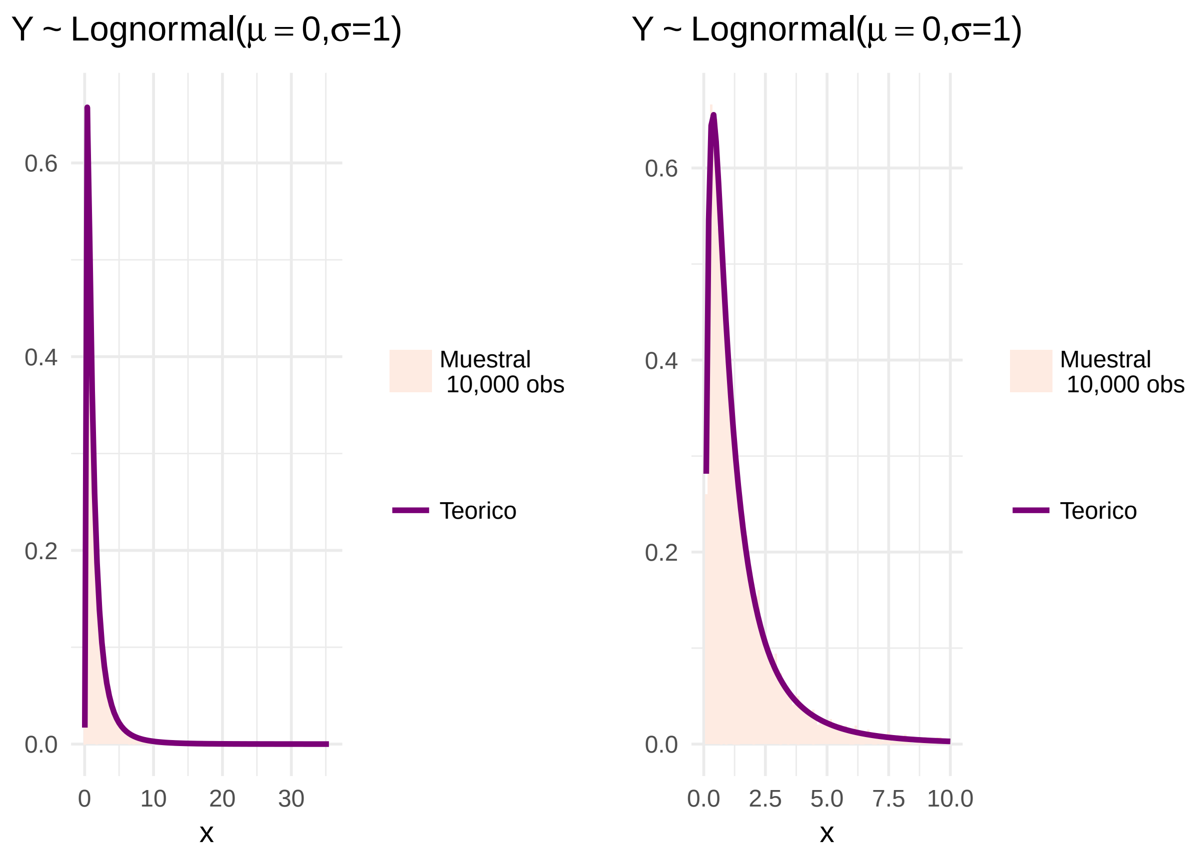


Figura 1

El histograma de los datos simulados $(1,000$ observaciones de Y) se ajusta a la distribución teórica de $LogNormal(0,1)$. La gráfica a la derecha de la figura 1 simplemente es un zoom en los valores menores en los que está definida Y (con el fin de observar mejor el ajuste).

Ejercicio 3

Simulamos $1,000$ observaciones de una v.a. $x\ Uniform(0,1)$ con la cual empleando la transformación $\frac{In(1-x)}{-\beta}$ obtenemos la muestra de observaciones con distribución $Exp(\beta)$.

```
set.seed(0)
n <- 1000
u <- runif(n, 0 ,1)             # se genera la muestra exponencial
alpha <- 1.5
beta <- 2
exponenciales <- -(log(1-u)/ beta) #se grafica la muestra
```

Y en la figura 2, contrastamos la distribución de la muestra contra la distribución teórica

```
ggplot(data.frame(y=exponenciales)) +
  geom_histogram(aes(y, ..density.., fill = I('Muestral \n 1000 obs'))) +
  theme_minimal() +
  stat_function(fun = function(x) {dexp(x, beta) },
              aes(colour = I('Teorico')), size = 1 ) +
  scale_fill_manual(name='',values=I('#CEE4EB')) +
  scale_colour_manual(name='',values=I('#7A0177') ) +
```

```
  ggtitle(TeX('$Y \\sim Exp( \\mu = 1.5 )$')) +
  theme(plot.title = element_text(hjust = 0.5)) +
  ylab('')+xlim(c(0,2))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

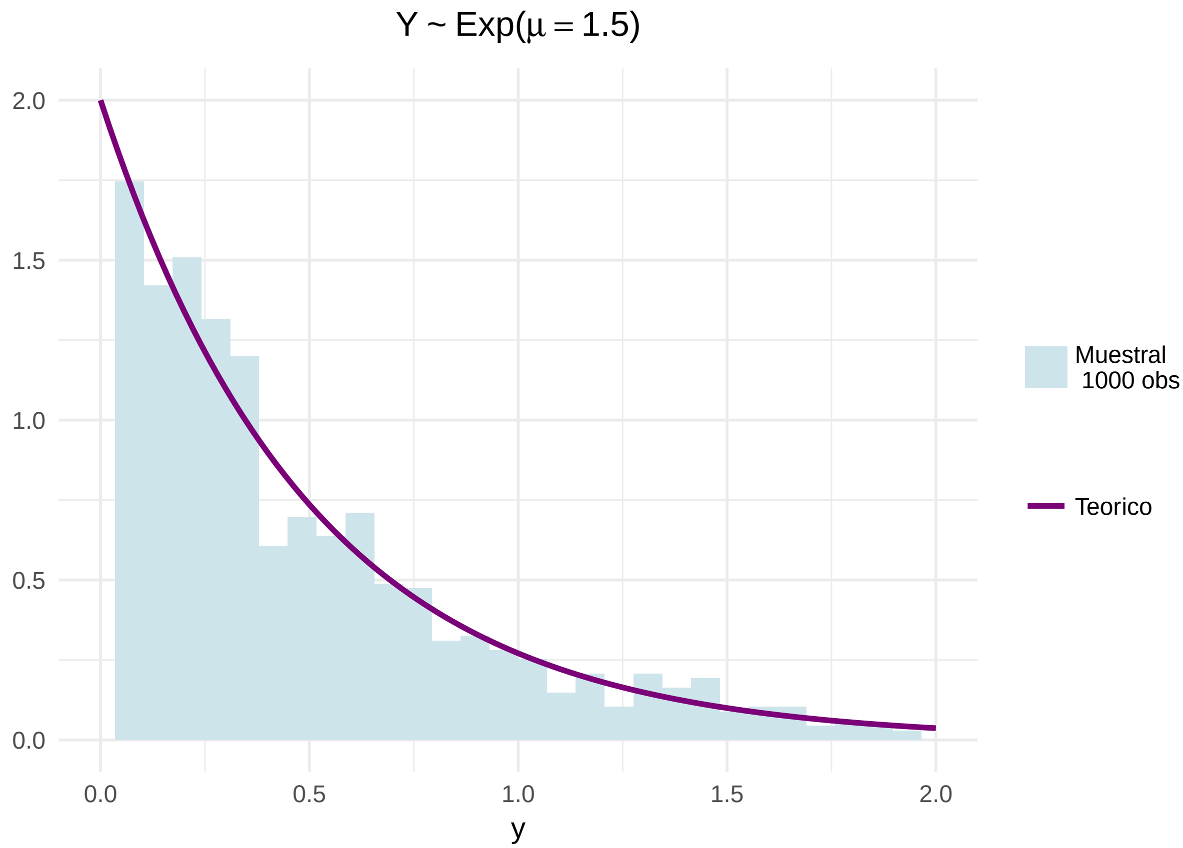


Figura 2

De manera análoga con la muestra uniforme generada y la transformación $\theta(-In(1-U))^{\frac{1}{\beta}}$ se genera una muestra de la distribución $Weibull(\theta, \beta)$, la distribución teórica se contrasta contra la muestral en la figura 3

```
Weibulls <- beta * ( (-log(1-u) )^(1/alpha)  ) # se genera la muestra weibull
ggplot(data.frame(y=Weibulls)) +           # se dibuja el histograma
  geom_histogram(aes(y, ..density.., fill = I('Muestral \n 1000 obs')), bins = 20) +
  theme_minimal() +
  stat_function(fun = function(x) {dweibull(x, alpha, beta) },
              aes(colour = I('Teorico')), size = 1 ) +
  scale_fill_manual(name='',values=I('#E7E1EF')) +
  scale_colour_manual(name='',values=I('#41B6C4') ) +
  ggtitle(TeX('$Y \\sim Weibull( \\alpha = 1.5, \\beta = 2 )$')) +
  theme(plot.title = element_text(hjust = 0.5)) +
  ylab('')
```

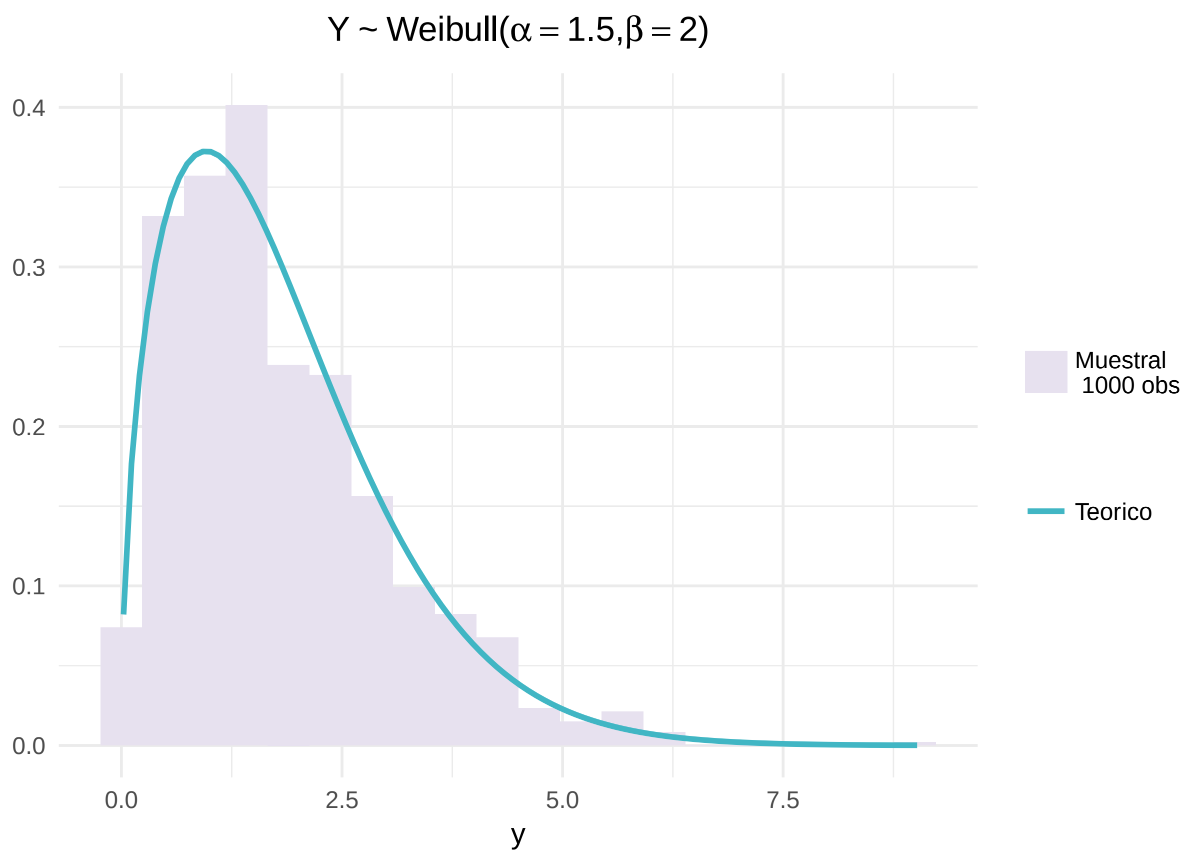


Figura 3

Ejercicio 18

El libro de texto dice que: >Let $X_1, .., X_n$ $N(0,1)$ r.v. and let $\bar{X_n} = \frac{\sum_{i=0}^{n} X_i}{n}$. Plot $\bar{X_n}$ versus $n = 1, ..., 10000$. >Repeat for $X_1, ..., X_n$ $Cauchy(0,1)$. Explain why there is such a differenece.

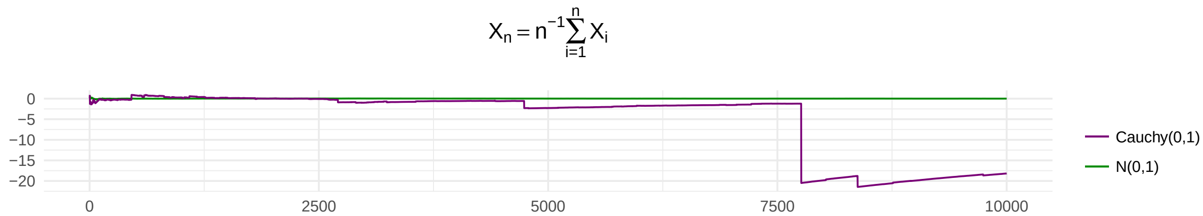
Se simulan los datos con distribución y con distribución $Cauchy(0,1)$

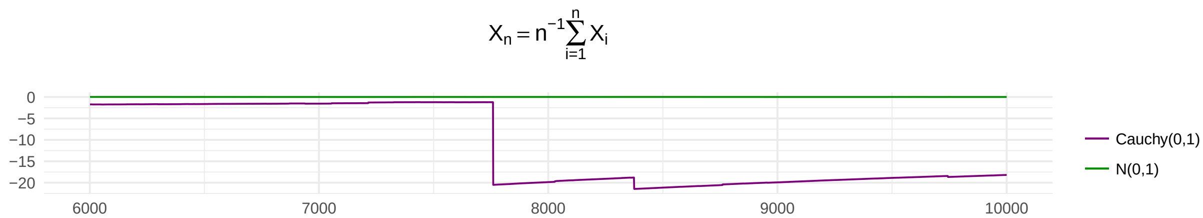
```
set.seed(240000)                      # fijamos la semilla por reproducibilidad
X <- rnorm(10000, 0, 1)
cum.mean <- function(i, x = X)
{     # Funcion para calcular la media acumulada recibe un entero para indicar el indice hasta
      # donde se considera tambien recibe el vector sobre el que se calcula la media
      return(mean(x[1:i]))
}
Cum.mean <- function(j) {unlist(lapply( 1:j, cum.mean)) }
medias.normal <-Cum.mean(length(X))
data <- data.frame(x=medias.normal)
data$origen <- 'N(0,1)'
data$indice <- 1:dim(data)[1]
X <- rcauchy(10000, location = 0, scale = 1)
medias.cauchy <- Cum.mean(length(X))
data2 <- data.frame(x=medias.cauchy)
data2$origen <- 'Cauchy(0,1)'
```

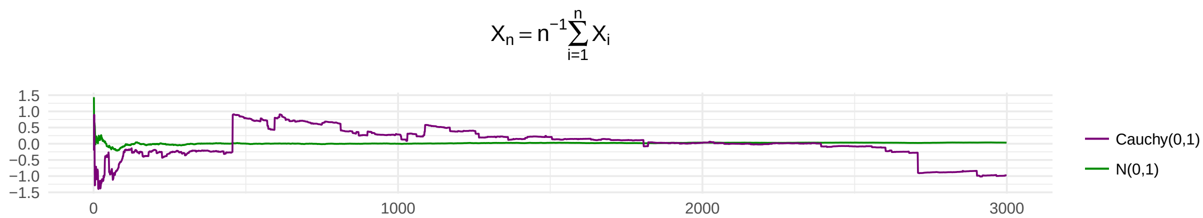
```r
data2$indice <- 1:dim(data2)[1]
data <- rbind(data, data2)
```

```r
p <- ggplot(data, aes(x=indice, y=x)) +geom_line(data =subset(data, indice > 1 & origen =='N(0,1)'),
                                                   aes(colour=origen))  +
            geom_line(data =subset(data, indice > 1 & origen !='N(0,1)'), aes(colour=origen))  +
            theme_minimal() +
            ggtitle(TeX('$X_n = n^{-1}\\sum_{i=1}^nX_i}$' )) +
            theme(plot.title = element_text(hjust = 0.5)) +
            scale_colour_manual(name='', values=c(I('#7A0177'), I('green4') ) )  +
            ylab('') + xlab('')
      # se repite la grafica anterior pero haciendo zoom en las ultimas 4000 ejecuciones
p1 <- ggplot(data, aes(x=indice, y=x)) +geom_line(data =subset(data, indice > 6000 & origen =='N(0,1)')
                                                    aes(colour=origen))  +
  geom_line(data =subset(data, indice > 6000 & origen !='N(0,1)'), aes(colour=origen))  +
  theme_minimal() +
  ggtitle(TeX('$X_n = n^{-1}\\sum_{i=1}^nX_i}$' )) +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_colour_manual(name='', values=c(I('#7A0177'), I('green4') ) )  +
  ylab('') + xlab('') + xlim( c(6000, 10000))
      # se repite la primer grafica haciendo zoom en las primeras 3000 ejecuciones
p2 <- ggplot(data, aes(x=indice, y=x)) +geom_line(data =subset(data, indice < 3000 & origen =='N(0,1)')
                                                    aes(colour=origen))  +
  geom_line(data =subset(data, indice < 3000 & origen !='N(0,1)'), aes(colour=origen))  +
  theme_minimal() +
  ggtitle(TeX('$X_n = n^{-1}\\sum_{i=1}^nX_i}$' )) +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_colour_manual(name='', values=c(I('#7A0177'), I('green4') ) )  +
  ylab('') + xlab('') + xlim( c(1, 3000))

p <- arrangeGrob(p, p1, p2, ncol = 1)
as_ggplot(p) #muestro tres imagenes de la distribución muestral solo cambia el soporte
```


$X_n = n^{-1}\sum_{i=1}^{n} X_i$
0
−5
−10
−15
−20
0
2500
5000
7500
10000
Cauchy(0,1)
N(0,1)


$X_n = n^{-1}\sum_{i=1}^{n} X_i$
0
−5
−10
−15
−20
6000
7000
8000
9000
10000
Cauchy(0,1)
N(0,1)


$X_n = n^{-1}\sum_{i=1}^{n} X_i$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
1000
2000
3000
Cauchy(0,1)
N(0,1)