Guia 18 clase

Agustin Muñoz González

22/6/2020

Preparamos el entorno

```
rm(list=ls())
library(ggplot2)
library(tidyr)
library(gganimate)
```

1. Estimar P $(X < 40)$ calculando la empírica en el valor $t = 40$.

Resolución:

```
datos=read.table("GRB_afterglow.dat.csv",header=T, skip=1)
empirica=function(t){
  mean(datos$f<t)
}
##############
empirica(40)
```

```
## [1] 0.6666667
```

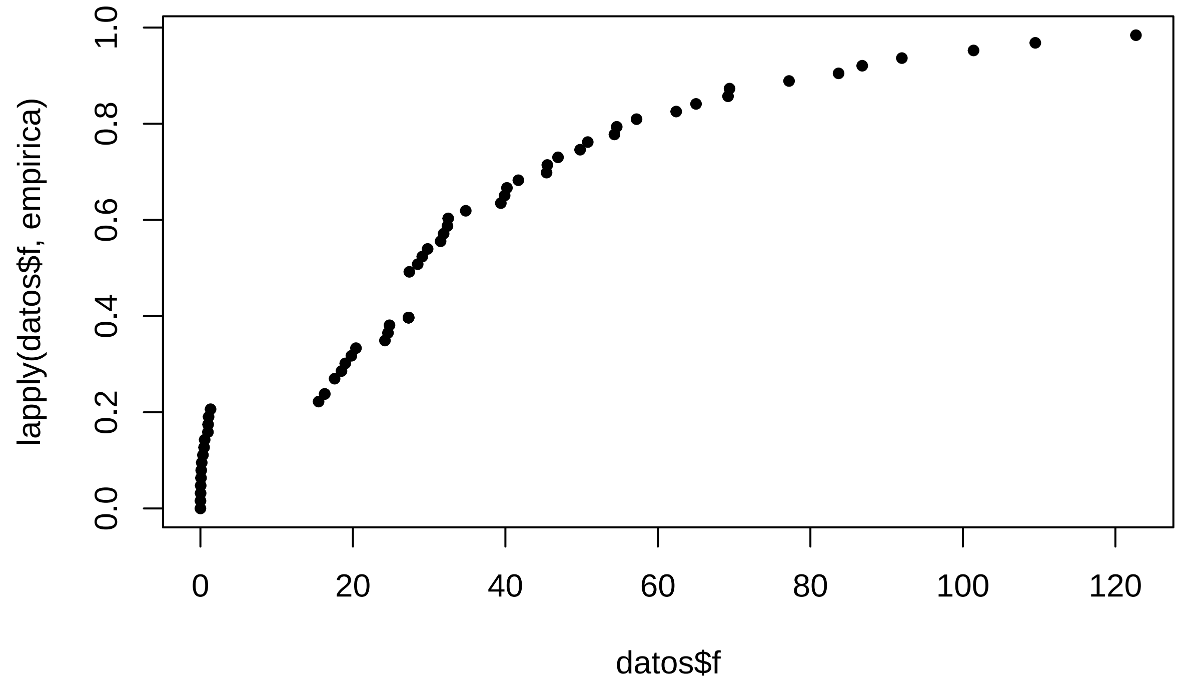
```
ecdf(datos$f)(40)
```

```
## [1] 0.6666667
```

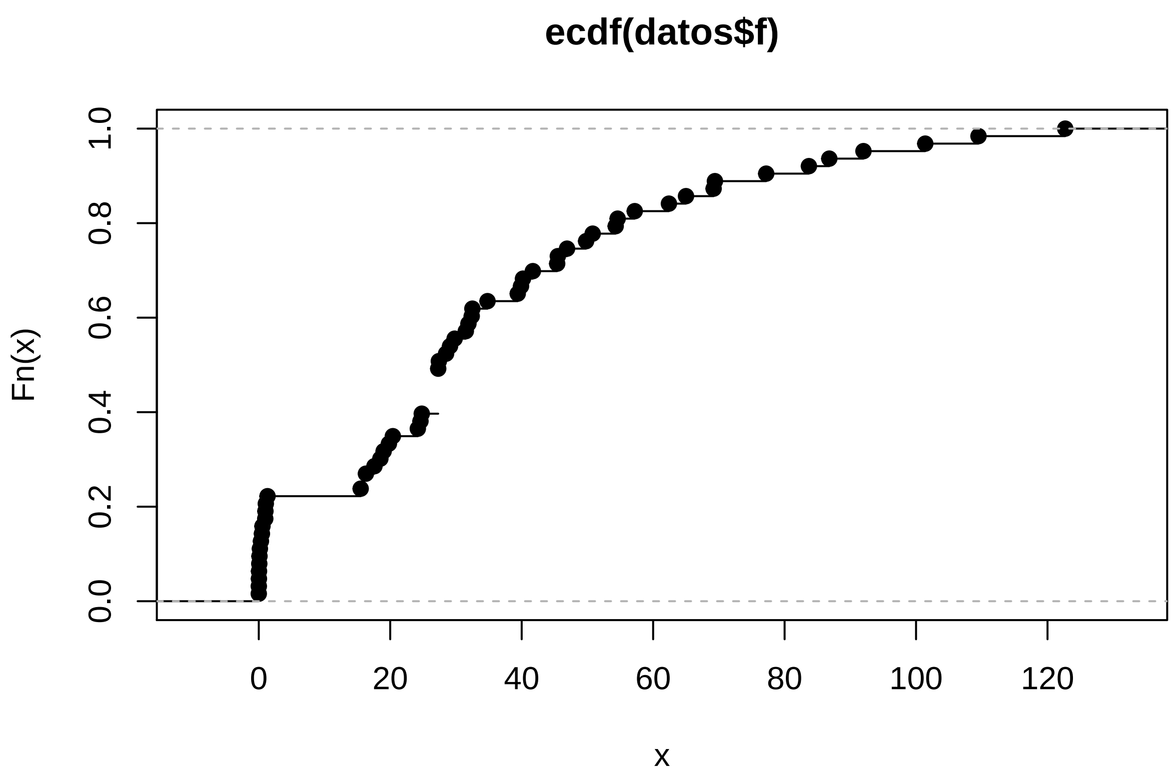
2. Graficar la empírica asociada a los datos flux. Para ello explorar el comando ecdf: empirical cummulative distribution f unction. Puede ejecutar ecdf(datos)(t) para calcuar la empírica de datos en el punto t, y también graficar utilizando plot(ecdf(datos)).

Resolución:

```
plot(datos$f,lapply(datos$f,empirica),pch=20)
```




```r
plot(ecdf(datos$f))
```



3. Estimar a partir de los datos el flux medio.

Resolución:

```r
flux_medio=mean(datos$f)
flux_medio
```

```
## [1] 34.44295
```

4. Estimar a partir de los datos la mediana de flux. ¿Usaría para esto la estimación obtenida en el ítem anterior?

Resolución:

```r
mediana=median(datos$f)
mediana
```

```
## [1] 27.4
```

```r
##############
```

5. Estimar a partir de los datos la varianza de flux.

Resolución:

```r
varianza=sd(datos$f)^2
var(datos$f)
```

```
## [1] 868.8771
```

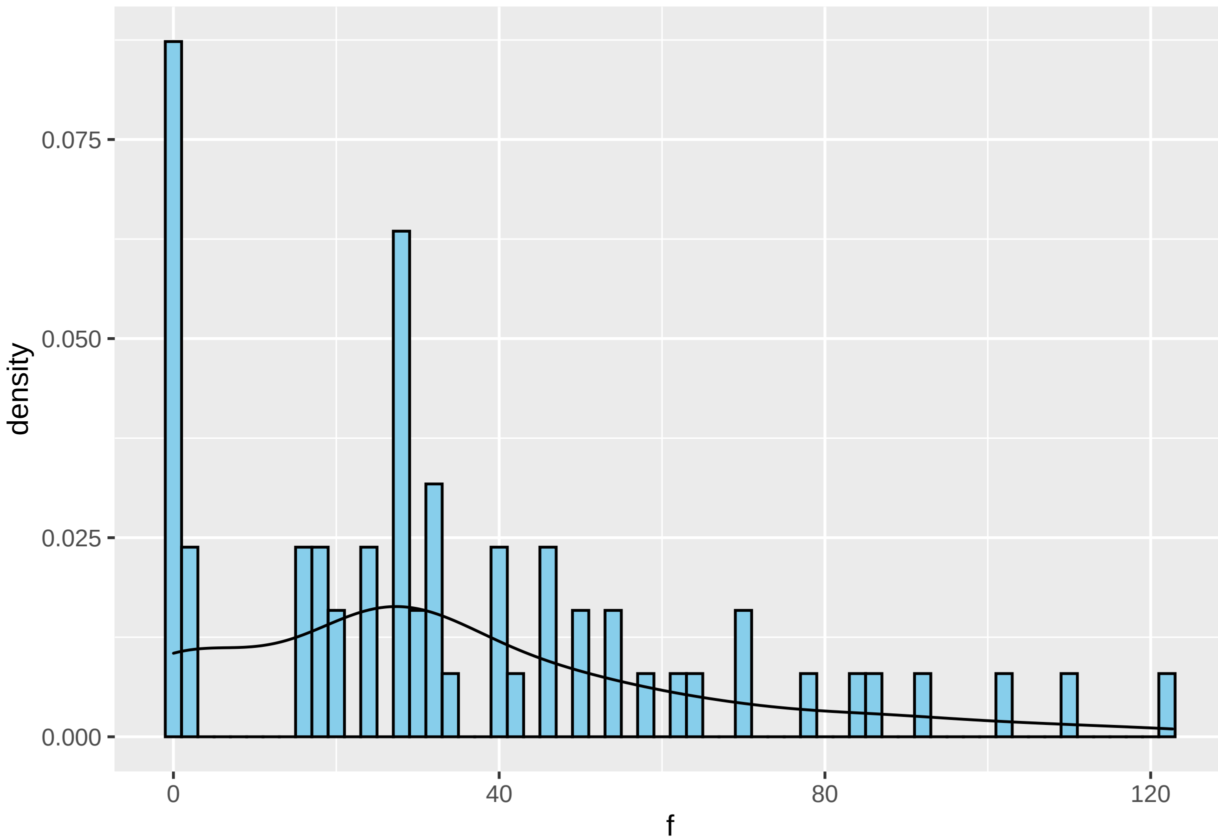
```r
varianza
```

```
## [1] 868.8771
```

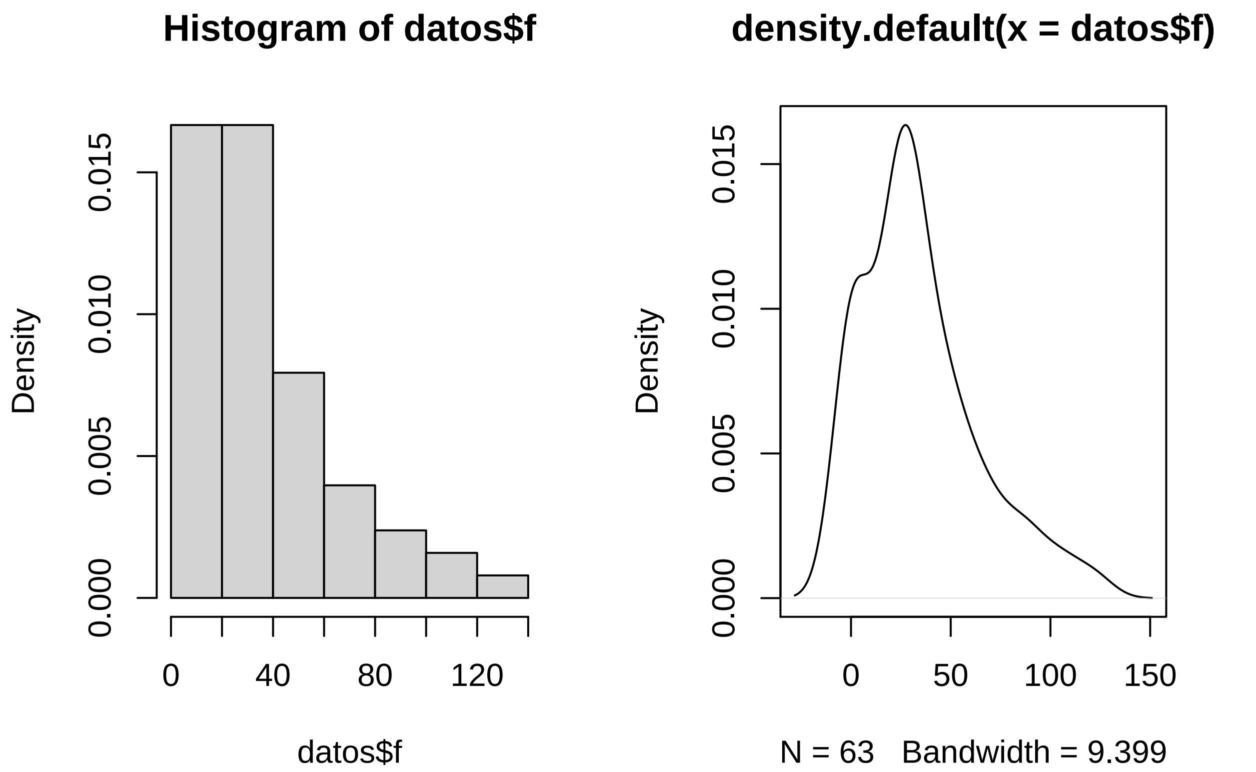
6. Realizar un histograma para los datos flux. ¿Los datos parecen tener alguna distribución conocida? Explorar el comando density.

Resolución:

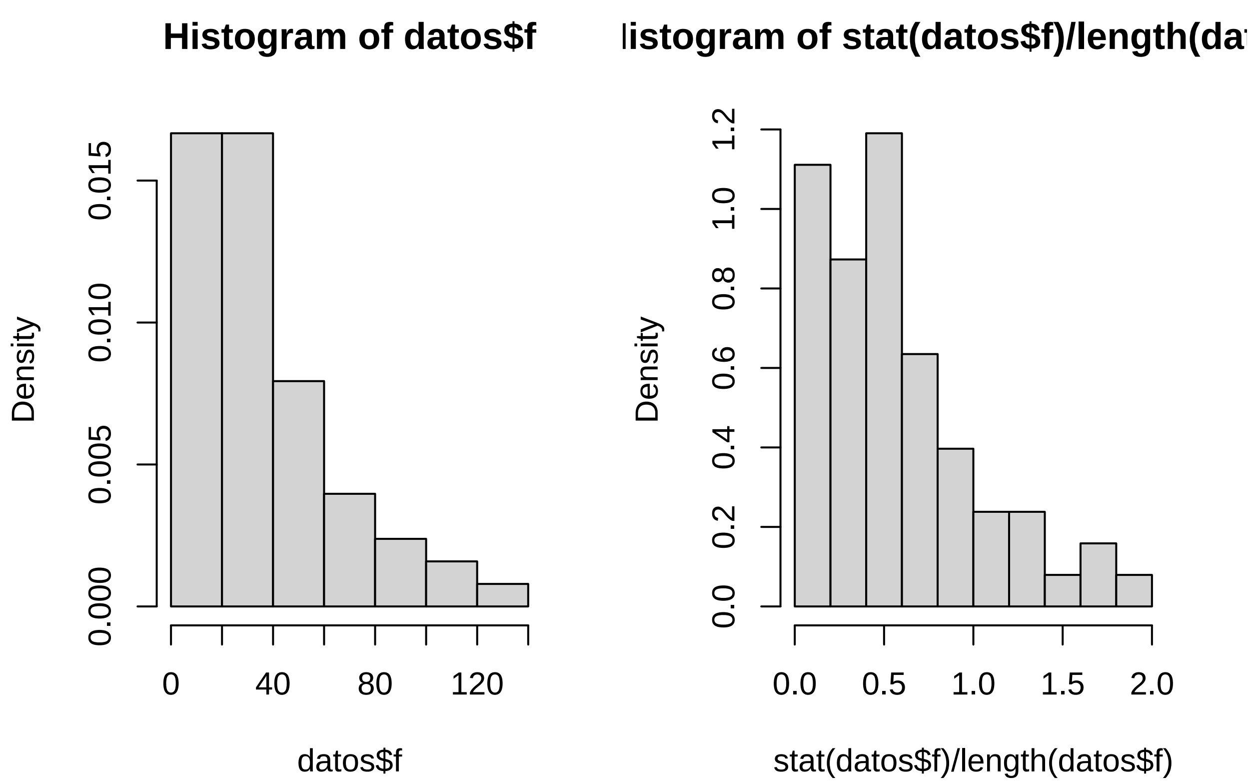
```r
library(ggplot2)
ggplot(datos)+
  geom_histogram(aes(x=f,y=..density..), binwidth=2,alpha=1,
                     fill="skyblue",color="black")+
  geom_density(aes(x=f))
```




```
par(mfrow=c(1,2))
hist(datos$f,freq=F)
plot(density(datos$f))
```




```r
par(mfrow=c(1,1))
par(mfrow=c(1,2))
hist(datos$f,freq=F)
hist(stat(datos$f)/length(datos$f),freq=F)
```



```r
par(mfrow=c(1,1))
```

2. Parte 2

2.1. Estimación bajo modelo exponencial: $E(\lambda)$

Distribución exponencial de parámetro $\lambda > 0$, $X_i \sim \mathcal{E}(\lambda)$.

$$f(x) = \lambda e^{-\lambda x}\mathcal{I}_{x\geq 0},\ F(t) = 1 - e^{-\lambda t},\quad \text{para } t \geq 0\cdot$$

En tal caso, $E(X) = \frac{1}{\lambda}$ y $V(X) = \frac{1}{\lambda^2}$. Notar que la verosimilitud, cuando $x_i \geq 0, \forall i$, resulta

$$L(\lambda, x) = \prod_i f(x_i, \lambda) = \prod_i \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_i x_i}.$$

Por lo tanto, la log–verosimilitud es:

$$l(\lambda, x) = n\log(\lambda) - \lambda \sum_i x_i.$$

de donde se deduce que el estimador de máxima verosimilitud está dado por $\widehat{\lambda} = \frac{1}{\overline{X}_n}$.

Utilizando el modelo propuesto, el estimador de máxima verosimilitud obtenido y el método plug-in, es decir reemplazando el parámetro por su valor estimado, resolver los siguientes items.

7. Estimar por el método de Máxima Verosimilitud $P(X \leq 40)$. Comparar con la estimación obtenida en la Sección 1.

Resolución:

Defino el estimador de maxima verosimilitud EMV del parámetro λ.

```
EMV=function(datos){
  1/mean(datos)
}
```

Quiero estimar $P(X \leq 40) = F_\lambda(40)$. Por el método plug-in basta hacer F_{EMV} con F la función de distribución de la variable exponencial.

Defino primero la función de distribución.

```
F_X=function(t,lambda){
  1-exp(-lambda*t)
}
```

Estimo $P(X \leq 40)$ usando F_X(40, EMV(datos)). Recuerdo además los valores que obtuvimos en la sección 1 con la empírica y con ecdf.

```
F_X(40,EMV(datos$f))
```

```
## [1] 0.6869339
```

```
pexp(40,EMV(datos$f))
```

```
## [1] 0.6869339
```

```
##############
empirica(40)
```

```
## [1] 0.6666667
```

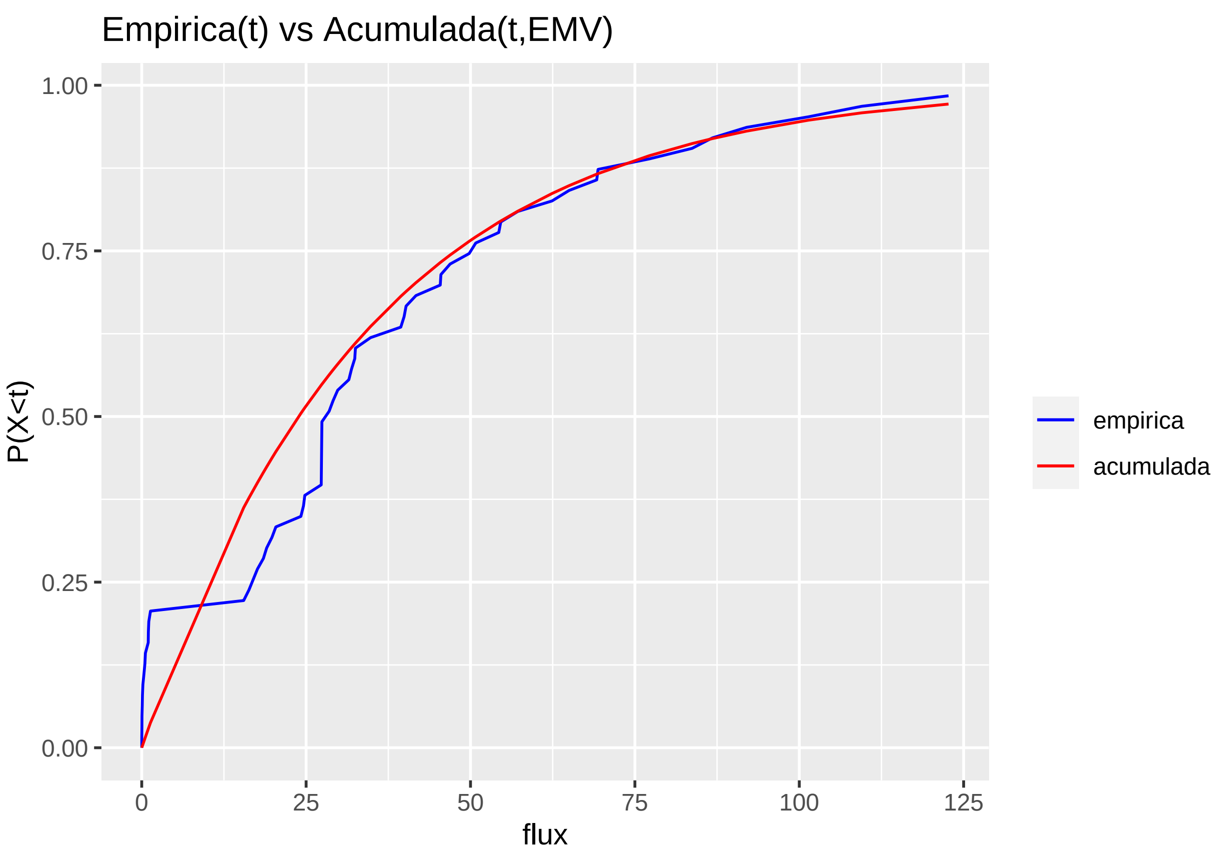
```
ecdf(datos$f)(40)
```

```
## [1] 0.6666667
```

8. Graficar la empírica asociada a los datos flux y superponer la función de distribución acumulada exponencial con el parámetro que considere pertinente.

Resolución:

```r
datos_plot=data.frame(cbind('flux'=datos$f,
'empirica'=sapply(datos$f,empirica),
'acumulada'=sapply(datos$f,F_X,lambda=EMV(datos$f))))
ggplot(datos_plot)+
  geom_line(aes(x=flux,y=empirica,color='empirica'))+
  geom_line(aes(x=flux,y=acumulada,color='acumulada'))+
  scale_colour_manual("",
                      breaks = c("empirica", "acumulada"),
                      values = c("blue", "red")) +
  xlab("flux") +
  scale_y_continuous("P(X<t)") +
  labs(title="Empirica(t) vs Acumulada(t,EMV)")
```



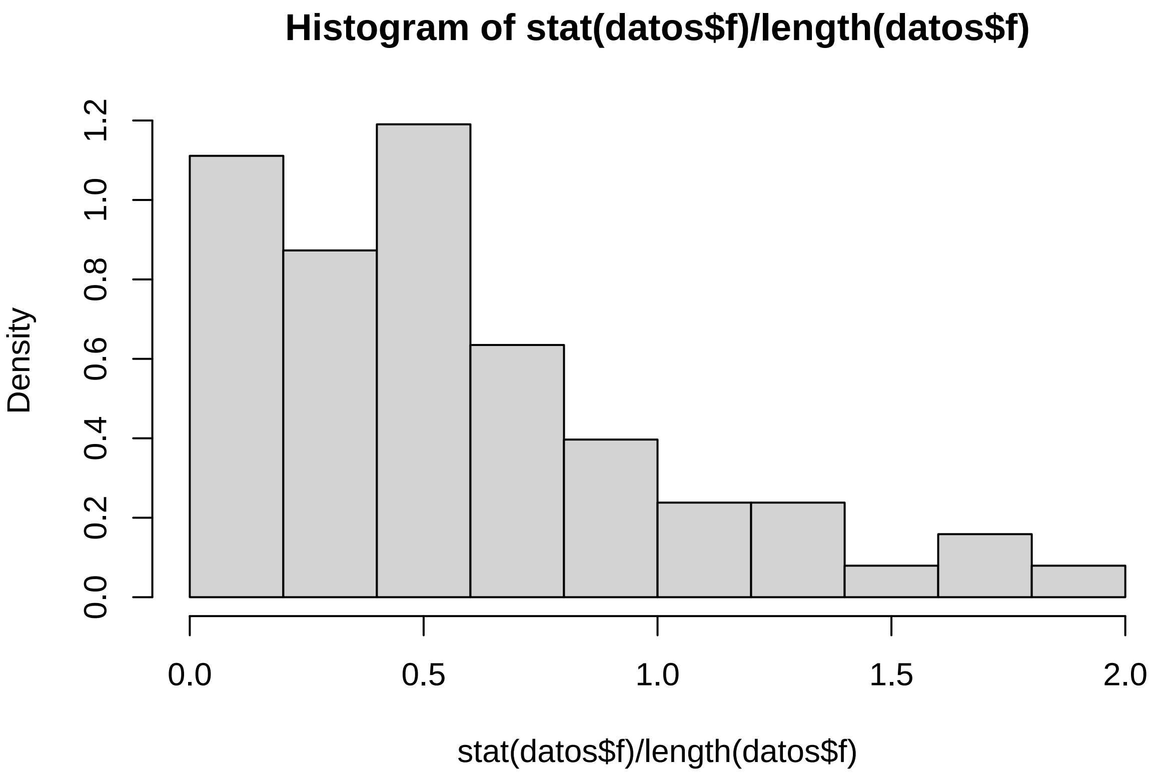
9. Realizar un histograma para los datos de flux y superponer la función de densidad exponencial con el parámetro que considere pertinente.

Resolución:

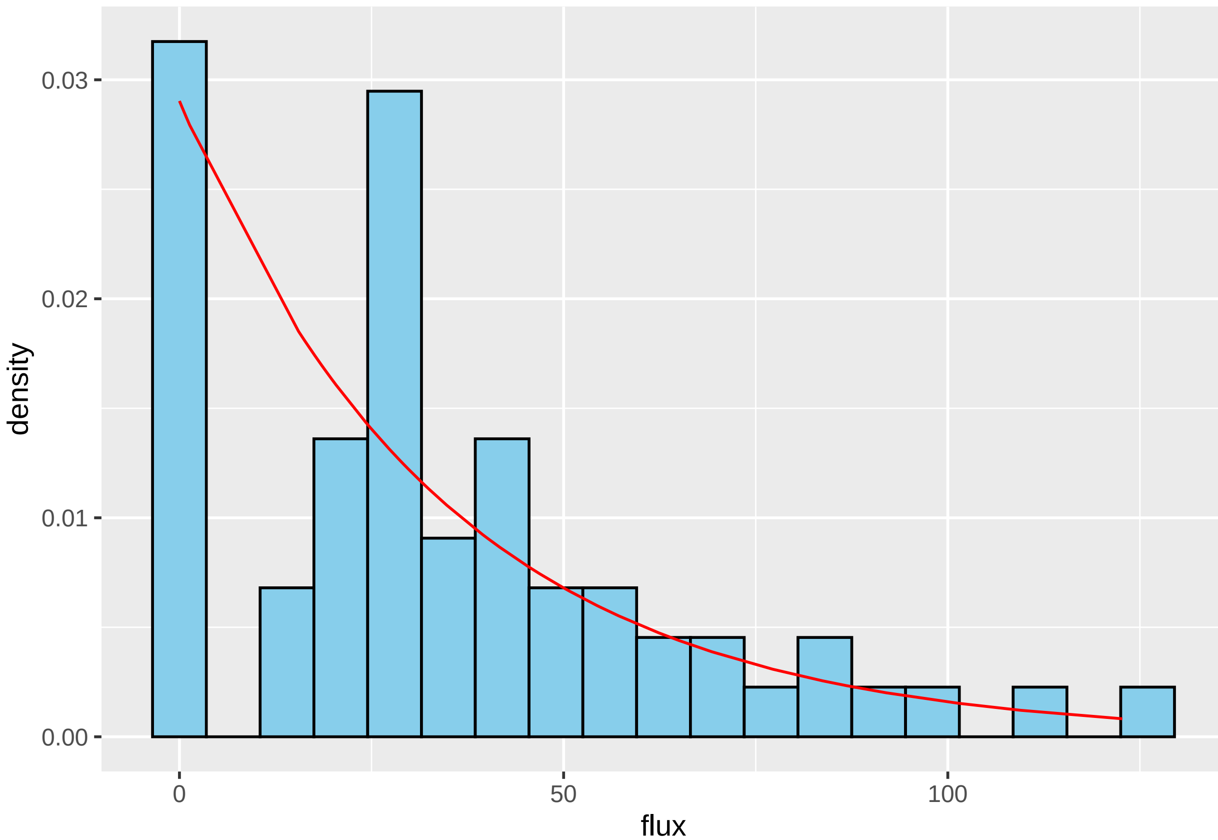
En primer lugar definimos la función de densidad de la exponencial.

```r
f_X=function(x,lambda){
  lambda*exp(-lambda*x)*ifelse(x>=0,1,0)
}
```

```
datos_plot=data.frame(cbind(datos_plot,'densidad'=sapply(datos$f,f_X,lambda=EMV(datos$f))))
# le paso los breaks del otro hist
a=hist(stat(datos$f)/length(datos$f),freq=F)$breaks
```




```
ggplot(datos_plot)+
  geom_histogram(aes(x=flux,y=..density..), binwidth=7,
                 fill="skyblue",color="black")+
  geom_line(aes(x=flux,y=densidad),color='red')
```



10. Estimar por el método de Máxima Verosimilitud el flux medio a partir de los datos.

Resolución:

El flux medio de una exponencial es el punto medio del grafico de densidad, i.e. la esperanza. Y comparo con lo que nos dio antes

```
1/EMV(datos$f)
```

```
## [1] 34.44295
```

```
qexp(0.5, rate = EMV(datos$f), lower.tail = TRUE, log.p = FALSE)
```

```
## [1] 23.87403
```

```
# OJO quantil 1/2 es la MEDIANA no la MEDIA
####################
flux_medio
```

```
## [1] 34.44295
```

11. Si $X \sim \mathcal{E}(\lambda)$, entonces su mediana resuelve la ecuación

$$1 - e^{\lambda t} = 0.5,$$

y por consiguiente vale

$$-\frac{\log(0.5)}{\lambda} = \frac{\log(2)}{\lambda}.$$

Estimar por el método de Máxima Verosimilitud la mediana de flux.

Resolución:

Calculo log(2)/EMV(datos$f) vs mediana

```r
log(2)/EMV(datos$f)
```

```
## [1] 23.87403
```

```r
###################
mediana
```

```
## [1] 27.4
```

12. Estimar por el método de Máxima Verosimilitud la varianza de flux.

Resolución:

La varianza de la exponencial es 1/lambda^2 entonces la aproximamos por 1/EMV(datos$f) y comparamos con lo que nos dio en la sección 1.

```r
1/EMV(datos$f)^2
```

```
## [1] 1186.317
```

```r
##############
varianza
```

```
## [1] 868.8771
```

EN LOS ITEMS 11 Y 12 NOS DAN COSAS DISTINTAS A LAS QUE HICIMOS EN LA SECCION 1 PORQUE ACA ESTAMOS USANDO QUE SABEMOS QUE ES UNA DIST EXPONENCIAL, O SEA ES UN TRAJE A MEDIDA ESTO! ES UNA ESTIMACION MEJOR ESTA ULTIMA!! UNA FORMA DE CHEQUEAR ESTO ES VER EL QUE TIENE MENOR ECM (ERROR CUADRATICO MEDIO)

```r
ECME=function(datos){
  theta=3
  mean((datos-theta)^2)
}
```