

L1A2

Emil K Svensson

8 November 2016

Assignment 2

2.2

What is the distribution type of x ?

The distribution given is an exponential distribution.

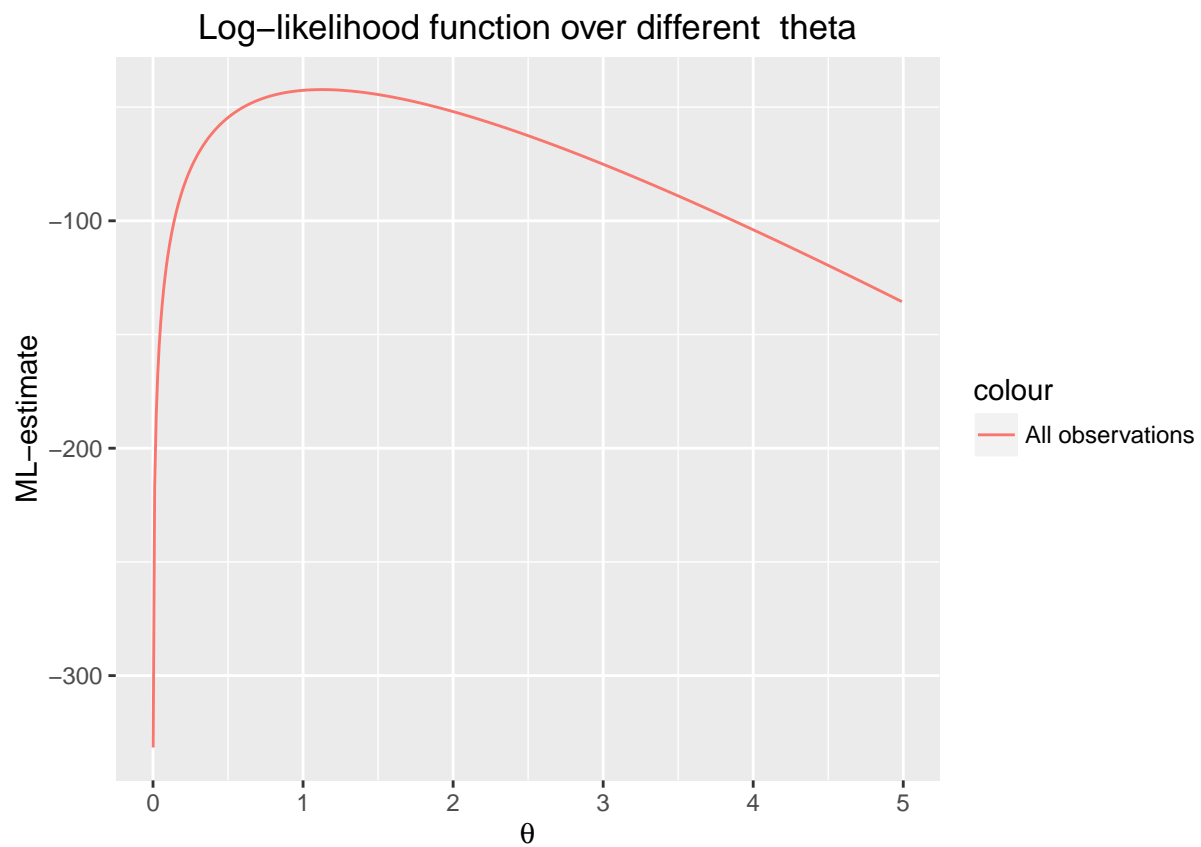
Write a function that computes the log-likelihood for a given θ and data vector x

The log-likelihood was computed to $\log p(x|\theta) = n \log(\theta) - \theta \sum_{i=1}^N x_i$

```
explog <-function(theta,data){  
  
  return(      length(data)*log(theta,base = exp(1)) - (theta*sum(data))      )  
  
}
```

Plot the curve showing the dependence of log-likelihood on θ where the entire data is used for fitting.

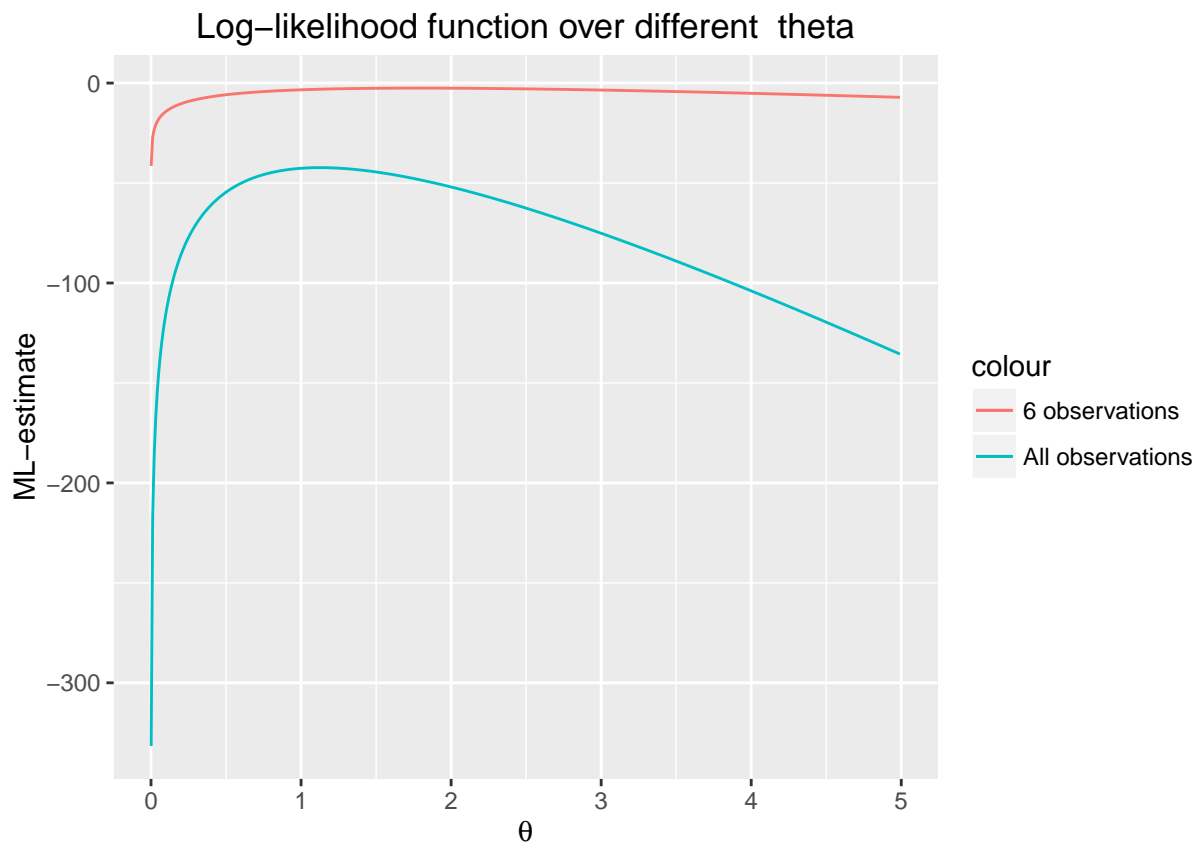
```
theta <- seq(0.001,5,by=0.01) #Generates a sequence of different  
thetaML <- sapply( X = theta, FUN = explog, data = machines)  
#apply the different thetas to the log-likelihood with the machine data-set.  
plotData <- data.frame(x=theta,y=thetaML) #Put the data in a frame.  
  
library(ggplot2)  
#Plotting  
  
p <- ggplot() + geom_line(data=plotData,aes(x=x,y=y,col="All observations")) +  
  labs(title=paste("Log-likelihood function over different theta"),  
        x=expression(theta),y="ML-estimate")  
plot(p)
```



The maximum log-likelihood is estimated to about 1.1 according to the plot.

2.3

```
thetaML3 <- sapply( X = theta, FUN = explog, data = machines[1:6])  
#Using only the 6 first observations of machines  
plotData$y2 <- thetaML3  
  
p + geom_line(data = plotData, aes(x=x,y=y2,col="6 observations"))
```

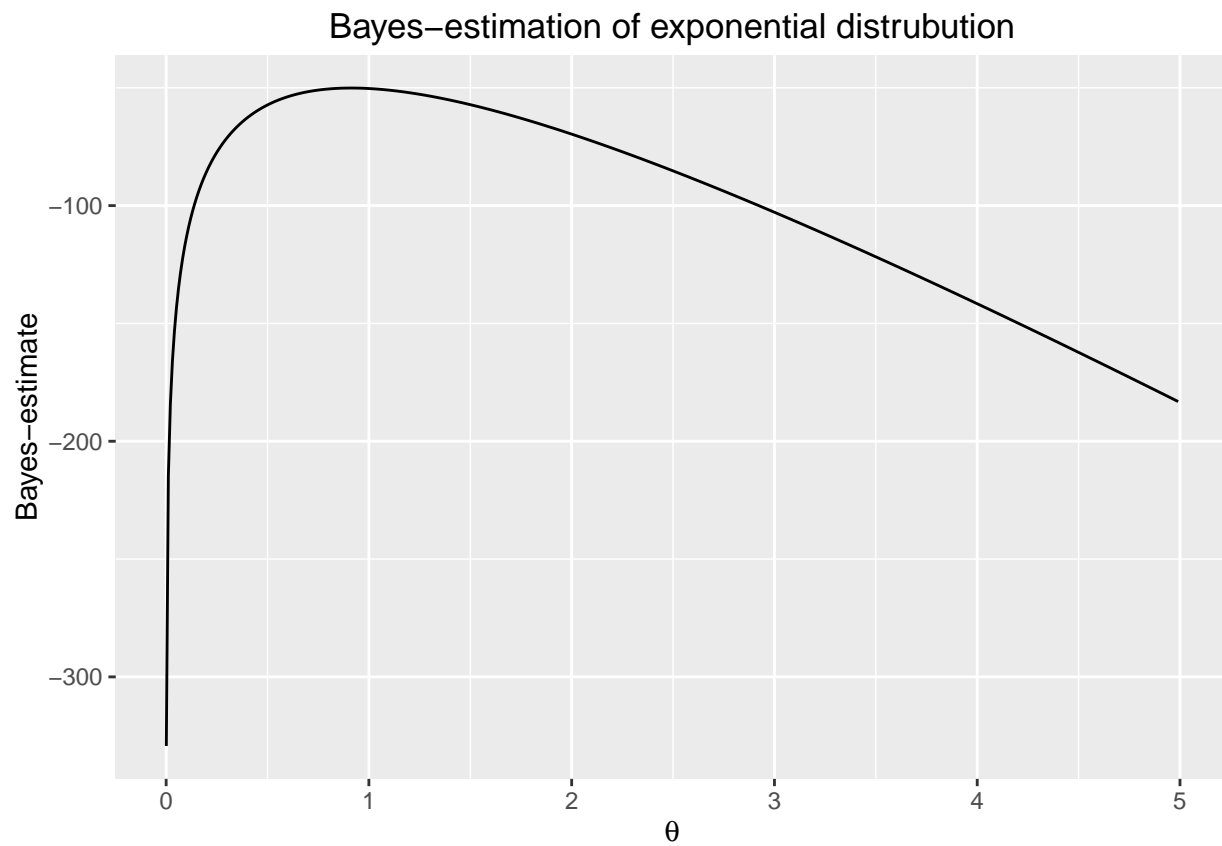


The light red line shows the likelihood function of different theta using six observations and the light blue line shows the estimate using all observations. Both lines increases at the starting values of theta and reaches their peaks at about the same time. The light red line has a higher intercept than the light blue curve, they differ with about 300 units.

The difference is that the estimate using all observations dies down at a much faster rate than the one using six observations and displays a much clearer max-value than the one with six observations. This shows a large dependence of data that the Maximum Likelihood estimators use.

2.4

```
postBayes <- function(data,theta){  
  lambda <- 10  
  return(length(data) * log(theta) - theta * sum(data) +  
         log(lambda) - (lambda*theta))  
}  
  
thetaBayes <- sapply(theta,FUN = postBayes,data = machines)  
  
plotDataBayes <- data.frame(x=theta,y=thetaBayes)  
B <- ggplot() + geom_line(data=plotDataBayes,aes(x=x,y=y)) +  
  labs(title="Bayes-estimation of exponential distrubution",  
       x=expression(theta),y="Bayes-estimate")  
plot(B)
```



2.5

```
library(gridExtra)

#extract the exact theta for the ML-estimate
maxTheta <- theta[thetaML == max(thetaML)]
# Generate 50 random variables from the exponential distrubution with
# the ML-estimate extracted
set.seed(12345)
randomExpData<- data.frame(x=rexp(50,rate = maxTheta))

#Plotting
ab<- ggplot(data = randomExpData,aes(x=x)) + geom_histogram(bins = 10)
abc<- ggplot(data =data.frame(x=machines),aes(x=x)) + geom_histogram(bins = 10)

grid.arrange( ab, abc, ncol=2)
```

