Assignment for Statistical Computing

Rene Franck ESSOMBA (ESSREN001)

April, 14 2012

1 Question 1

The data that we are using in our analysis is **aircondit** provided in the **boot** package. The observations are the times in hours between failures of air-conditioning equipment, hence they are assumed to be exponentially distributed. Our aim is to estimate the MLE using a boostrap method and estimate the bias and standard error of it. Also, we will find a basic confidence interval and a percentile confidence interval and compare the two.

1.1 Statistical background

The pdf of an exponential distribution is $f(x) = \lambda e^{-\lambda x}, x > 0$, and the maximum likelihood estimate is $\hat{\lambda} = \frac{n}{\sum x_i}$

1.2 Problem Solving

```
> data <- c(3,5,7,18,43,85,91,98,100,130,230,487)
> m.data <- mean(data)
> data.mle <- 1/m.data        #analytical method
> data.mle

[1] 0.00925212
```

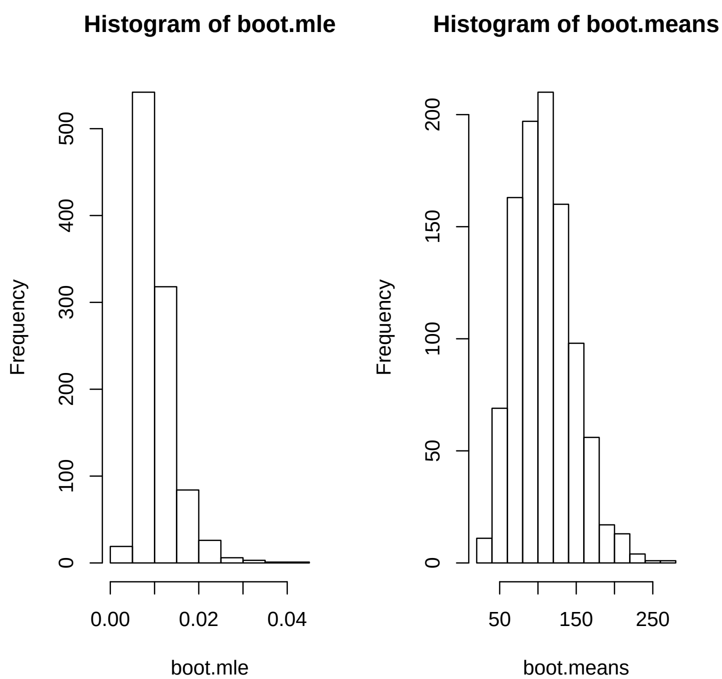
Using an optimization approach, we get the following:

```
> llik <- function(lambda)
+ { L <- lambda * exp(-lambda * data)
+   ll <- sum(log(L))
+   return(-ll)                       # max(llik) = min(-llik)
+ }
> optim(1,llik)$par

[1] 0.009228516

> nlm(llik,1)$estimate
```

```
[1] 0.00925162
```

Now we are going to generate B = 1000 bootsrap samples and for each of them calculate the mean and the MLE (analytically) and store them in **boot.means** & **boot.mle** vectors. Here is the code and the outputs:

```
> B <- 1000
> data.vect <- c(3,5,7,18,43,85,91,98,100,130,230,487)
> boot.means <- numeric(B)
> boot.mle <- numeric(B)
> for (i in 1:B)
+ { boot.samp <- sample(data.vect,replace=T)
+ boot.means[i] <- mean(boot.samp)
+ boot.mle[i] <- 1/boot.means[i]
+ }
> par(mfrow =c(1,2))
> hist(boot.mle)
> hist(boot.means)
```



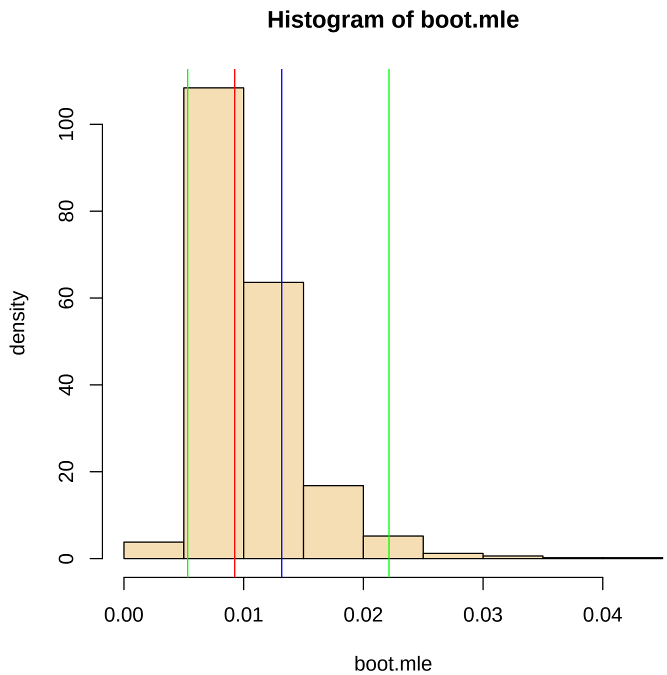
Hence the estimates of the bias and standar error of the estimate are:

```
> bmle <- mean(boot.mle) - data.mle
> bmle                                          # bias of the estimate
```

```
[1] 0.001271437

> data.mle - bmle

[1] 0.007980683

> sd(boot.mle)                        # standard error of the estimate

[1] 0.004328079
```

The 95% bootstrap confidence interval & percentile confidence interval are derived as follows:

```
> pci <- quantile(boot.mle, p = c(0.025,0.975))
> LL <- as.numeric(2*data.mle - pci[2])
> UL <- as.numeric(2*data.mle - pci[1])
> bbci <- c(LL,UL)
> pci # percentile confidence interval

       2.5%       97.5%
0.005323281 0.022141245

> bbci # basic bootstrap confidence interval

[1] -0.003637004  0.013180960

> hist(boot.mle,freq=F,ylab="density",col="wheat")
> abline(v=data.mle,col="red")
> abline(v=pci[1],col="green")
> abline(v=pci[2],col="green")
> abline(v=bbci[1], col="blue")
> abline(v=bbci[2], col="blue")
```



There is a significant difference between the intervals, this is because of the skew distribution, in this case the percentile confidence interval is more reliable in the sense that it will more accurately capture the true MLE 95% of the time (see the graph above).

1.3 Summary

For this little exercise, we were asked to estimate the MLE of the distribution of a data set that was exponentially distributed. For this purpose, we generated 1000 bootstrap samples for the MLE and computed the bias and the standard error of the estimate. We found two confidence intervals for the MLE and concluded that the percentile confidence interval was more accurate.

2 Question 2

The observations are an i.i.d sample from a Cauchy(θ,1) distribution. Our aim is to estimate the MLE for θ using the Newton-Raphson method and the bisection method.

2.1 Relevant background

The pdf of a Cauchy(θ,1) distribution is in the form: $f(x) = \frac{1}{\pi}\left(\frac{1}{(x-\theta)^2+1}\right)$, where $\theta$is the location parameter and the scale parameter is 1, therefore the log likehood function will be: $g(\theta) = \sum(\log(\pi) - \log((x_i - \theta)^2 + 1))$

Its first derivative is: $g'(\theta) = \sum\left(\frac{2(x-\theta)}{1+(x-\theta)^2}\right)$

Its second derivative is: $g''(\theta) = \sum\left(\frac{2[(x-\theta)^2-1]}{[1+(x-\theta)^2]^2}\right)$