

# Programming in R

## Task 4

Sofia Gianelli Nan, Gabriela Levenfeld Sabau and Miguel Díaz-Plaza Cabrera

2023-10-14

### Table of Contents

Exercise 1.....	2
Exercise 2.....	3
Exercise 3.....	4

## Exercise 1.

Write a program in R for numerically computing both gamma distribution parameters' maximum likelihood (MLE) estimators.

The code and the output generated by it are shown below.

```
# Simulate some gamma-distribution data
set.seed(1)
N <- 1000;
sampleData <- rgamma(N, shape = 2, rate = 0.2)

# Define the likelihood function for the gamma distribution
gamma_likelihood = function(alpha, beta){
  R <- dgamma(sampleData, shape = alpha, rate = beta, log = TRUE)
  log_LL <- -sum(R)
}

# Perform MLE
library(stats4)
mle(minuslogl=gamma_likelihood, start=list(alpha=0.1, beta=0.1),
method="L-BFGS-B", lower=c(0,0), upper=c(Inf, Inf))

##
## Call:
## mle(minuslogl = gamma_likelihood, start = list(alpha = 0.1, beta =
## 0.1),
##      method = "L-BFGS-B", lower = c(0, 0), upper = c(Inf, Inf))
##
## Coefficients:
##      alpha      beta
## 1.8756167 0.1887568
```

Although the code is commented, we want to emphasize some insight below.

First of all, we simulate 1000 data points from a gamma distribution with known parameters using 'rgamma', as it is specified on the statement. The set up for this parameters are:  $\alpha = 2$  and  $\beta = 0.2$ . Then the next part of the code consists of defining the likelihood function.

And, finally, we use an optimization function to get the Maximum Likelihood Estimator (MLE) in which is a requirement to set up some parameters. The only thing necessary is to take into account that both alpha ( $\alpha$ ) and beta ( $\beta$ ) must be greater than zero.

The output generate consist on the estimated values of  $\alpha$  and  $\beta$  (1.8756167 and 0.1887568, respectively) which represent the best-fitting gamma distribution to the provided data. As a conclusion, we can say that the estimated parameters are reasonably close to the true values, indicating that the MLE process has performed well.

## Exercise 2.

**Make a short summary of the main characteristics of the library MaxLik.**

The **MaxLik** library is a package in R used for computing the maximum likelihood estimation (MLE) where it can be calculate with different methods. It also provides tools and mechanisms for other optimization tasks.

The *maxLik* function can be used as follows:

```
maxLik(logLik, grad = NULL, hess = NULL, start, method, constraints  
= NULL, ...)
```

It requires a user defined log-likelihood function, with options for specifying gradient and Hessian. You can choose from various optimization methods, with support for constraints, and pass additional arguments to the selected optimization routine.

Finally, *MaxLike* has some useful methods such as: *coef*, *logLik* and *summary*, among others. The latter (*summary*) will be used in the following section, since it allows displaying a summary of the MLE on estimates, standard errors, p, and z-values.

### Exercise 3.

Apply this package to calculate the MLE estimates of the previous gamma-distributed data. Compare the results with the user-written function.

```
# Download the necessary package
if (!requireNamespace("maxLik", quietly = TRUE))
install.packages("maxLik", dependencies = TRUE)
library(maxLik)

## Loading required package: miscTools

##
## Please cite the 'maxLik' package as:
## Henningsen, Arne and Toomet, Ott (2011). maxLik: A package for maximum
## likelihood estimation in R. Computational Statistics 26(3), 443-458. DOI
## 10.1007/s00180-010-0217-1.
##
## If you have questions, suggestions, or comments regarding the 'maxLik'
## package, please use a forum or 'tracker' at maxLik's R-Forge site:
## https://r-forge.r-project.org/projects/maxlik/

# Simulate some gamma-distribution data
set.seed(1)
N <- 1000;
sampleData <- rgamma(N, shape = 2, rate = 0.2)

# Define the likelihood function for the gamma distribution
gamma_likelihood = function(parameters){
  alpha <- parameters[1]
  beta <- parameters[2]
  R <- dgamma(sampleData, shape = alpha, rate = beta, log = TRUE)
  log_LL <- sum(R)
}

# Performe MLE using MaxLik Library
ml_NR <- maxLik(gamma_likelihood, start = c(alpha = 1, beta = 0.1),
method = "NR")
print(summary(ml_NR))

## -----
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 5 iterations
## Return code 8: successive function values within relative tolerance
## limit (reltol)
## Log-Likelihood: -3198.393
## 2 free parameters
## Estimates:
##      Estimate Std. error t value Pr(> t)
## alpha 1.875567  0.077446  24.22  <2e-16 ***
## beta  0.188750  0.008929  21.14  <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## -----

ml_NM <- maxLik(gamma_likelihoood, start = c(alpha = 1, beta = 0.1),
method = "NM")
print(summary(ml_NM))

## -----
## Maximum Likelihood estimation
## Nelder-Mead maximization, 59 iterations
## Return code 0: successful convergence
## Log-Likelihood: -3198.393
## 2 free parameters
## Estimates:
##      Estimate Std. error t value Pr(> t)
## alpha 1.875631   0.077677  24.15  <2e-16 ***
## beta  0.188782   0.008952  21.09  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## -----
```

As shown on the code above, the structure is quite similar to the one made before. However, we have had to make some modifications in order to make the code works.

- Download and install the necessary packages.
- The biggest change is found on computing the log-likelihood function where instead of having two parameter, we have created the variable “parameters” in which  $\alpha$  and  $\beta$  are storage.
- And, lastly, the MLE is calculated with the MaxLik library as required by the statement. In addition, we have considered it appropriate to test at least two of the methods offered by this library to find the MLE.

After executing the R code, we can conclude that all three methods generate similar results for the parameters of the gamma distribution. A summary table is presented below to facilitate understanding of the results. As can be seen, the estimates obtained using *stats4* library are very closed to those obtained using *MaxLik* library. No matter which method is used, Newton-Raphson or Nelder-Mead method. Furthermore, the maximum likelihood estimator for  $\alpha$  and  $\beta$  are consistent across the methods which suggest all processes has performed well.

Method	Alpha	Beta
stats4 (L-BFGS-B)	1.875617	0.1887568
MaxLik (Newton-Raphson)	1.875567	0.1887500
MaxLik (Nelder-Mead)	1.875631	0.1887820