Maximum Likelihood Estimation: Numerical Solution for Gamma Distribution[1]

The Gamma distribution has the probability density function:

$$f\left(y|\alpha,\beta\right) \;=\; \frac{1}{\Gamma\left(\alpha\right)\beta^{\alpha}}y^{\alpha-1}\text{exp}\left\{-\frac{y}{\beta}\right\} \tag{1}$$

where: $y \geq 0$, β is the scale parameter and α is the shape parameter (known parameter).

Maximum Likelihood Estimation

Let $y_1,\ldots,y_n$, denote the data. Assume, $\forall i : y_i$ independente random variables, share the same parameters from a Gamma distribution described in (1).

$$f\left(y_i|\beta\right) \;=\; \frac{1}{\Gamma\left(\alpha\right)\beta^{\alpha}}{y_i}^{\alpha-1}\text{exp}\left\{-\frac{y_i}{\beta}\right\} \tag{2}$$

then, their join probability distribution is:

$$\begin{aligned} f\left(y_1,\ldots,y_n|\beta\right) &= \prod_{i=1}^{n} f\left(y_i|\beta\right) \\ &= \prod_{i=1}^{n} \frac{1}{\Gamma\left(\alpha\right)\beta^{\alpha}}{y_i}^{\alpha-1}\text{exp}\left\{-\frac{y_i}{\beta}\right\} \end{aligned} \tag{3}$$

The likelihood function is:

$$\mathscr{L}\left(\beta|y_1,\ldots,y_n\right) \;=\; f\left(y_1,\ldots,y_n|\beta\right) \tag{4}$$

The log-likelihood function is:

$$\begin{aligned} \ell\left(\beta|y_1,\ldots,y_n\right) &= \log\left(\mathscr{L}\left(\beta|y_1,\ldots,y_n\right)\right) \\ &= \log\prod_{i=1}^{n}\frac{1}{\Gamma\left(\alpha\right)\beta^{\alpha}}{y_i}^{\alpha-1}\text{exp}\left\{-\frac{y_i}{\beta}\right\} \\ &= \sum_{i=1}^{n}\log\left(\frac{1}{\Gamma\left(\alpha\right)\beta^{\alpha}}{y_i}^{\alpha-1}\text{exp}\left\{-\frac{y_i}{\beta}\right\}\right) \\ &= \sum_{i=1}^{n}\left(-\log\,\Gamma\left(\alpha\right)-\alpha\,\log\,\beta+\left(\alpha-1\right)\,\log\,y_i-\frac{y_i}{\beta}\right) \end{aligned} \tag{5}$$

The maximum likelihood estimator (MLE), denoted by $\hat{\beta}$, is such that:

$$\hat{\beta} \;=\; \text{argmax}\left\{\ell\left(\beta|y_1,\ldots,y_n\right)\right\} \tag{6}$$

[1] *To refer this document and the implemented code, please cite as*: Alvarado, M. (2020, September 8). Maximum Likelihood Estimation: Numerical Solution for Gamma Distribution (Version v1.0.0). Zenodo. http://doi.org/10.5281/zenodo.4019900. Also at GitHub: https://github.com/miguel-alvarado-stats/MLE_Gamma.

To maximize the log-likelihood function (5), requires the derivative with respect to β. The resulted function is called the Score function, denoted by $U(\beta|y_1,\ldots,y_n)$.

$$
\begin{aligned}
U(\beta|y_1,\ldots,y_n) &= \frac{\partial \ell(\beta|y_1,\ldots,y_n)}{\partial \beta} \\
&= \sum_{i=1}^{n}\left(-\alpha\frac{1}{\beta}+\frac{y_i}{\beta^2}\right) \\
&= -\frac{n\alpha}{\beta}+\frac{1}{\beta^2}\sum_{i=1}^{n} y_i \qquad (7)
\end{aligned}
$$

Then, the MLE $\hat{\beta}$, is the solution of:

$$
U\left(\beta=\hat{\beta}|y_1,\ldots,y_n\right) = 0 \qquad (8)
$$

Maximum Likelihood Estimation: Newton-Raphson Method

Just for notation, let write equation (8) as:

$$
U(\beta^*) = 0 \qquad (9)
$$

The equation (9), generally, is a nonlinear equation, that can be aproximate by Taylor Series:

$$
U(\beta^*) \approx U\left(\beta^{(t)}\right)+U'\left(\beta^{(t)}\right)\left(\beta^*-\beta^{(t)}\right) \qquad (10)
$$

Then, using (10) into (9), and solving for β^*:

$$
\begin{aligned}
U\left(\beta^{(t)}\right)+U'\left(\beta^{(t)}\right)\left(\beta^*-\beta^{(t)}\right) &= 0 \\
\beta^* &= \beta^{(t)}-\frac{U\left(\beta^{(t)}\right)}{U'\left(\beta^{(t)}\right)} \qquad (11)
\end{aligned}
$$

where U' is the derivative of the Score function (7) respect of β.

$$
\begin{aligned}
U'(\beta|y_1,\ldots,y_n) &= \frac{\partial U(\beta|y_1,\ldots,y_n)}{\partial \beta} \\
&= \frac{n\alpha}{\beta^2}-\frac{2}{\beta^3}\sum_{i=1}^{n} y_i \qquad (12)
\end{aligned}
$$

Then, with the Newton-Raphson method: starting with an initial guess $\beta^{(1)}$ successive approximations are obtained using (13), until the iterative process converges.

$$
\beta^{(t+1)} = \beta^{(t)}-\frac{U\left(\beta^{(t)}\right)}{U'\left(\beta^{(t)}\right)} \qquad (13)
$$

In order to example the use of Newton-Rapshon method, we simulate a set of data ($N=15$) that comes from a Gamma distribution whose shape parameter α (which is assumed to be known) is equal to 2 and some positive value for the scale parameter β which, for this example, we will take equal to 2[2].

[2]The package extraDistr provides a tool kit for the Gamma distribution. In particular, the scale parameter β for the Gamma distribution is supplied in the form of scale = 1/rate.

```r
# load packages required
library(extraDistr)

shape <- 2
rate <- 0.5
N <- 15

set.seed(8257)
Y <- rgamma(N, shape, scale = 1/rate)
Y

##  [1] 7.9405506 6.8116147 4.6820307 6.3833320 4.4570320 1.2385520 0.5243205
##  [8] 4.2149631 7.7965736 1.0916002 1.7835166 3.1615198 3.2806975 1.2061107
## [15] 4.5700020
```

We load the code developed into our R function MLE_NR_Gamma stored in the R object with the same name.

```r
# load the function to solve by Newton-Raphson
load("MLE_NR_Gamma.RData")
```

The function MLE_NR_Gamma takes the minimum value of the sample $\beta = \min\{y\}$ as a first guess for the iterative process, besides some other default parameters that can be modified such as the shape parameter α (i.e. shape $= 2$), only needs the data vector Y.

```r
# MLE by Newton-Raphson (NR) for Gamma distribution
MLE_NR_Gamma(Y)

##       ML Estimator Likelihood             Log-Likelihood
##  [1,] "0.5243205"  "4.54775195701934e-34" "-76.7732601263105"
##  [2,] "0.7462714"  "4.2497602616121e-24"  "-53.8151796595443"
##  [3,] "1.0322944"  "8.66065680277306e-19" "-41.5903262039121"
##  [4,] "1.3653770"  "2.31178080438159e-16" "-36.0033433493826"
##  [5,] "1.6864185"  "1.56223856568787e-15" "-34.0926566242412"
##  [6,] "1.8994213"  "2.24929259750235e-15" "-33.7281606292388"
##  [7,] "1.9663410"  "2.29684276062607e-15" "-33.7072409277801"
##  [8,] "1.9713878"  "2.29707167336356e-15" "-33.7071412686589"
##  [9,] "1.9714139"  "2.29707167937519e-15" "-33.7071412660419"
```

Then, the MLE by Newton-Raphson method: $\hat{\beta} = 1.9714139$.

Maximum Likelihood Estimation: Fisher-Scoring Method

A distribution belongs to the exponential family if it can be written in the form:

$$f(y|\beta) = \exp\left\{\frac{a(y)\,b(\beta) - c(\beta)}{\phi} + d(y,\phi)\right\} \tag{14}$$

Since (1) can be written as a member of exponential family as in (14):

$$\begin{aligned} f(y|\beta,\lambda) &= \exp\left\{\log\left(\frac{1}{\Gamma(\alpha)\beta^{\alpha}}y^{\alpha-1}\exp\left\{-\frac{y}{\beta}\right\}\right)\right\} \\ &= \exp\left\{y\left(-\frac{1}{\beta}\right)-\alpha\,\log\beta+(\alpha-1)\;\log y-\log\Gamma(\alpha)\right\} \end{aligned} \tag{15}$$

where, $a(y)=y$, $b(\beta)=-1/\beta$, $c(\beta)=\alpha\,\log\beta$, $d(y,\phi)=(\alpha-1)\;\log y-\log\Gamma(\alpha)$, and $\phi=1$.

Then, since the Gamma distribution belongs to the exponential family, it can be show that the variance of U, denoted by $\mathcal{J}$, is:

$$\mathcal{J} = \mathsf{Var}\{U\} = -\mathsf{E}\{U'\} \tag{16}$$

where:

$$\mathsf{E}\{U'\} = -\frac{1}{\phi}\left(b''(\beta)\frac{c'(\beta)}{b'(\beta)}-c''(\beta)\right) \tag{17}$$

For MLE, it is common to approximate U' by its expected value $\mathsf{E}\{U'\}$. In this case:

$$\begin{aligned} \mathcal{J} &= -\mathsf{E}\{U'\} \\ &= \mathsf{E}\{-U'\} \\ &= \mathsf{E}\left\{-\sum_{i=1}^{n}U_i'\right\} \\ &= \sum_{i=1}^{n}-\mathsf{E}\{U_i'\} \\ &= \sum_{i=1}^{n}-\frac{1}{\phi}\left(b''(\beta)\frac{c'(\beta)}{b'(\beta)}-c''(\beta)\right) \end{aligned} \tag{18}$$

where, using (1), the previous derivaties:

$$\begin{aligned} b'(\beta) &= \frac{1}{\beta^2} \\ b''(\beta) &= -\frac{2}{\beta^3} \\ c'(\beta) &= \frac{\alpha}{\beta} \\ c''(\beta) &= -\frac{\alpha}{\beta^2} \\ \frac{1}{\phi} &= 1 \end{aligned}$$

Then, replacing them into (18):

$$\begin{aligned} \mathcal{J} &= \sum_{i=1}^{n}-\left(\left(-\frac{2}{\beta^3}\right)\left(\frac{\frac{\alpha}{\beta}}{\frac{1}{\beta^2}}\right)-\left(-\frac{\alpha}{\beta^2}\right)\right) \\ &= \sum_{i=1}^{n}\frac{\alpha}{\beta^2} \\ &= \frac{n\alpha}{\beta^2} \end{aligned} \tag{19}$$

Finally:

$$
\begin{aligned}
\mathcal{J} &= -\mathsf{E}\left\{U'\right\} = \frac{n\alpha}{\beta^2} \\
-\mathcal{J} &= \mathsf{E}\left\{U'\right\} = \frac{n\alpha}{\beta^2}
\end{aligned}
\tag{20}
$$

Then, approximating U' by its expected value $\mathsf{E}\left\{U'\right\}$, the equation (13) results into:

$$
\beta^{(t+1)} = \beta^{(t)} + \frac{U\left(\beta^{(t)}\right)}{\mathcal{J}\left(\beta^{(t)}\right)} \tag{21}
$$

In order to example the use of Fisher-Scoring method, we use the same data used in the Newton-Rapshon method. We load the code developed into our R function MLE_FS_Gamma, stored in the R object with the same name.

```
# load the function to solve by Fisher-Scoring
load("MLE_FS_Gamma.RData")
```

The function MLE_FS_Gamma takes the minimum value of the sample $\beta = \min\left\{y\right\}$ as a first guess for the iterative process, besides some other default parameters that can be modified such as the shape parameter α (i.e. shape $= 2$), only needs the data vector Y.

```
# MLE by Fisher-Scoring (FS) for Gamma distribution
MLE_FS_Gamma(Y)

##      ML Estimator Likelihood            Log-Likelihood
## [1,] "0.5243205"  "4.54775195701934e-34" "-76.7732601263105"
## [2,] "1.9714139"  "2.29707167937518e-15" "-33.7071412660419"
## [3,] "1.9714139"  "2.29707167937519e-15" "-33.7071412660419"
```

Then, the MLE by Fisher-Scoring method: $\hat{\beta} = 1.9714139$.

In summary, the same estimate for the MLE is achieved by both approaches: the Newton-Raphson and the Fisher-Scoring method, however the latter does it more efficiently.

Naive Approach

The main idea behind the Maximum Likelihood (ML) method is to choose those estimates for the unknown parameters that maximize the join probability of our observed data (our sample). Keeping in mind this idea, if we want to get the MLE and avoiding to implement a numerical solution, a naive approach is to set a large range of possible values for unknown parameters, evaluate the log-likelihood function (also, the likelihood function) at each point and the point for which the log-likelihood function (also, the likelihood function) reaches its maximum value, will be our MLE looking for.

We set a set of values for the scale parameter β, by setting rate $= 0.2$ to rate $= 1$, spaced by 0.001; since scale $= 1/\text{rate}$ and evaluating (4) and (5) at each point from the set of values.

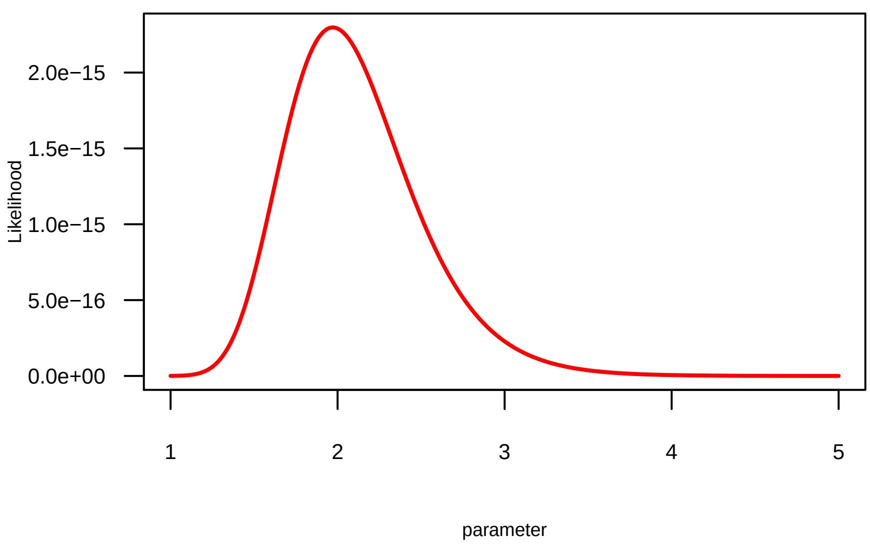
```r
# package dplyr is required to use the pipe operator
library(dplyr)

# set a large range of values for the parameter
rate <- seq(0.2, 1, 0.001)
tabla <- matrix(c(NA,NA,NA,NA), nrow = length(rate), ncol = 4)
tabla[,1] <- rate
tabla[,2] <- 1/tabla[,1]

# evaluate the likelihood and the log-likelihood at each point
for (i in 1:length(rate)){
  tabla[i,3] <- prod(dgamma(Y, shape = 2, scale = 1/rate[i], log = FALSE))
  tabla[i,4] <- sum(dgamma(Y, shape = 2, scale = 1/rate[i], log = TRUE))
}
colnames(tabla) <- c("rate", "beta", "Likelihood", "Log-Likelihood")
df_tabla <- as.data.frame(tabla)
```

Then, the plot of likelihood function evaluated at each point:

Figure 1: Family of Likelihood Function



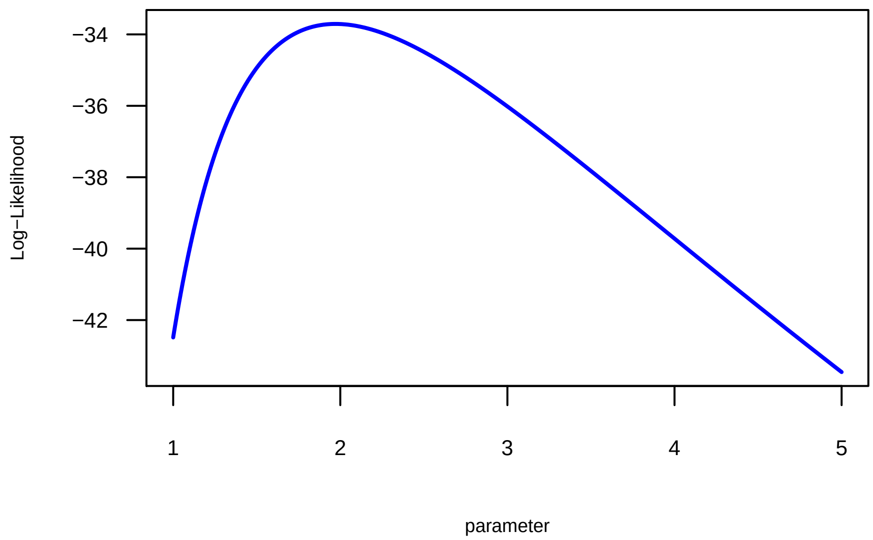
The point for which the likelihood function is maximum, is our MLE from a naive approach:

```r
mxl <- max(tabla[,3])
df_tabla %>% filter(`Likelihood` == mxl)

##    rate     beta   Likelihood Log-Likelihood
## 1 0.507 1.972387 2.297063e-15      -33.70714
```

Also, the plot of log-likelihood function evaluated at each point:

Figure 2: Family of Log−Likelihood Function



The point for which the log-likelihood function is maximum, which is the same point at the likelihood function reaches its maximum value, is our MLE from a naive approach:

```
mxllog <- max(tabla[,4])
df_tabla %>% filter(`Log-Likelihood` == mxllog)

##    rate     beta   Likelihood Log-Likelihood
## 1 0.507 1.972387 2.297063e-15      -33.70714
```

As we can see, using this naive approach, we reach a value that is close enough to that which is reached using the numerical solutions: the Newton-Raphson and the Fisher-Scoring method.