Gamma Distribution

The probability density function of Gamma distribution is:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta \cdot x}$$

Where, $a > 0$ is the shape parameter and $\beta > 0$ is the rate parameter. The likelihood function will be:

$$L(a, \beta | x) = \prod_{i=1}^{n} \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x_i^{\alpha-1} \cdot e^{-\beta \cdot x_i}$$

And therefore, the log-likelihood function, is:

$$l(a, b | x) = ln[L(a, \beta | x)] = \sum_{i=1}^{n} ln\left(\frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x_i^{\alpha-1} \cdot e^{-\beta \cdot x_i}\right) =$$

$$= \sum_{i=1}^{n} ln\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right) + \sum_{i=1}^{n} ln(x_i^{\alpha-1}) + \sum_{i=1}^{n} ln(e^{-\beta \cdot x_i}) =$$

$$= \sum_{i=1}^{n} a \cdot ln(\beta) - \sum_{i=1}^{n} ln(\Gamma(\alpha)) + \sum_{i=1}^{n} (\alpha - 1) \cdot \ln(x_i) + \sum_{i=1}^{n} (-\beta \cdot x_i) \cdot ln(e) =$$

$$= n \cdot a \cdot ln(\beta) - n \cdot ln(\Gamma(\alpha)) + (\alpha - 1) \cdot \sum_{i=1}^{n} \ln(x_i) - \beta \cdot \sum_{i=1}^{n} x_i$$

Now this is if we want to maximize our function, if instead we want to minimize it (like for example, if we use the $optim()$ function in R), then we will use the following log-likelihood function:

$$-l(a, b | x) = -ln[L(a, \beta | x)] =$$

$$= -\left[n \cdot a \cdot ln(\beta) - n \cdot ln(\Gamma(\alpha)) + (\alpha - 1) \cdot \sum_{i=1}^{n} \ln(x_i) - \beta \cdot \sum_{i=1}^{n} x_i\right] =$$

$$= -n \cdot a \cdot ln(\beta) + n \cdot ln(\Gamma(\alpha)) - (\alpha - 1) \cdot \sum_{i=1}^{n} \ln(x_i) + \beta \cdot \sum_{i=1}^{n} x_i$$

If we want to estimate the parameters a and β in R, using the log-likelihood function above, then we shall do it like this:

```r
#////////////////////////////////////////////////////////
# Gamma Distribution MLE for shape and scale paramters
#\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\

# Insert the data as a vector
dat <- scan(file = "C:\\Users\\30697\\Desktop\\whales.txt")


# ================================
# Gamma, negative log-likelihood
# ================================
logl <- function(thet, dat){
 #gamma minus log-likelihood
 n <- length(dat)
 a <- thet[1]; b <- thet[2]
 -(n*a*log(b) - n*lgamma(a) + (a-1)*sum(log(dat)) - b*sum(dat))
}


# ================================================
# Maximum Likelihood using Newton-Rhapson
# ================================================
# Load the library (numDeriv)
library(numDeriv)

# Initial Values
thet0 <- c(1,1)
iter <- 0
eps <- 1
```

```r
# Maximum Likelihood Estimation
while (eps > 1e-6){
  iter <- iter + 1
  gradient <- grad(logl, x = thet0, dat = dat)
  hess <- hessian(logl, x = thet0, dat = dat)
  thet <- thet0 - solve(hess)%*%gradient
  thet <- as.vector(thet)
  eps <- sum(abs(thet - thet0))
  thet0 <- thet
}

thet
```
[1] 1.595409 2.632692
```r
iter
```
[1] 7

```r
# ========================================
# Maximum Likelihood using optim()
# ========================================
fit <- optim(par = c(1,1), logl, method = 'BFGS', dat = dat)
fit
```
$par
[1] 1.595409 2.632693

$value
[1] 92.72379

$counts

```
function gradient
   31    10


$convergence
[1] 0


$message
NULL


$hessian
          [,1]           [,2]
[1,]   180.95152  − 79.76624
[2,] − 79.76624     48.33827
```

Our algorithms converged to shape value of $a = 1.595409$ and rate value of $\beta = 2.632693$.

Also, what we know for the Gamma distribution is that:

$$Mean = E(x) = \frac{shape}{rate} = \frac{a}{\beta} = 0.605999$$

$$Variance = Var(x) = \frac{shape}{(rate)^2} = \frac{a}{\beta^2} = 0.2301822$$

Is the Gamma distribution is preferable to the Exponential distribution for these data. Since the Exponential distribution is a special case of the Gamma distribution, where the shape parameter $a = 1$, then this question is the same as saying whether:

$$H_0 : a = 1$$

$$H_1 : a \neq 1$$

We will test this hypothesis, using Wald Test. The Wald statistic is:

$$W = \frac{(\hat{a} - a_0)^2}{Var(\hat{a})}$$

Which under the null hypothesis follows an asymptotic X^2 distribution with one degree of freedom.

What we need is the variance of $\hat{a}$, which we will get by inverting the hessian matrix and taking the first element of the first column. Thus, we calculate it like this in R:

```
# Calculate the variance of a (the shape parameter)
variance <- solve(fit$hessian)[1,1]


# Calculate the Wald statistic
wald <- ((fit$par[1] - 1)^2)/variance
wald
[1] 17.48592


# Perform a hypothesis test based on p-value
pchisq(wald, df = 1, lower.tail = FALSE)
[1] 2.894436e-05
```

As we can see, the $p-value$ is lower than our significance level of 5% ($p-value < 0.05 \Longrightarrow$ $\Longrightarrow 2.894436e - 05 < 0.05$), therefore we reject H_0 in favour of H_1, thus implying that the Gamma distribution seems to be a better choice for this particular set of data.

An alternative way to do Wald Test:

We could also do the test, by using the square root of the Wald Test statistic, which under the null hypothesis follows an asymptotic standardized normal distribution with mean zero and variance one. The test statistic for that is:

$$\sqrt{W} = \frac{\hat{a} - a_0}{Standard\ Error(\hat{a})}$$

What we will do now is test whether this value is in the 95% Confidence Interval. All that in R is done like this:

```
# Calculate the standard error of a (the shape parameter)
standard.error <- sqrt(variance)


# Wald Test
wald2 <- (fit$par[1] - 1)/standard.error
wald2
[1] 4.181617


# 95% Confidence Interval
confidence.interval <- fit$par[1] + c(-1,1) * 1.96 * standard.error


# Hypothesis Test
confidence.interval[1] < wald2 && wald2 < confidence.interval[2]
[1] FALSE
```

Therefore, this statistic is not in our 95% confidence interval. And so, we can say that once again, we reject H_0 in favour of H_1, thus implying that the Gamma distribution seems to be a better choice for this set of data.