Statistics 2 Unit 2 Team 8

Nikolaos Kornilakis   Rodrigo Viale   Jakub Trnan   Aleksandra Daneva
Luis Diego Pena Monge

2024-03-20

Exercise 18

The Box-Cox transformation is involved, which is a family of power transformations aimed at stabilizing variance and making the data more closely conform to a normal distribution.

Let $X_1, ..., X_n$ be i.i.d positive random variables s.t the Box-Cox transform $X_i^{(\lambda)} \sim N(\mu, \sigma^2)$. Define

$$g(x, \lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ ln(x) & \lambda = 0 \end{cases}$$

Notice that $X_i^{(\lambda)} = g(X_i, \lambda)$, and that

$$\frac{d}{dx} g(x, \lambda) = \begin{cases} x^{\lambda - 1} & \lambda \neq 0 \\ x^{-1} & \lambda = 0 \end{cases} = x^{\lambda - 1}$$

Then by transformation formula we have that

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2}\left(\frac{g(x,\lambda) - \mu}{\sigma}\right)^2\right) \left|\frac{d}{dx} g(x, \lambda)\right|$$

Then, the likelihood is given by:

$$\begin{aligned} L(\mu, \sigma, \lambda | X_1, ..., X_n) &= \prod_{i=1}^n f_X(X_i) \\ &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2}\left(\frac{g(X_i,\lambda) - \mu}{\sigma}\right)^2\right) \left|\frac{d}{dx} g(X_i, \lambda)\right| \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \prod_{i=1}^n \exp\left(\frac{-1}{2}\left(\frac{g(X_i,\lambda) - \mu}{\sigma}\right)^2\right) \prod_{i=1}^n \left|\frac{d}{dx} g(X_i, \lambda)\right| \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \prod_{i=1}^n \exp\left(\frac{-1}{2}\left(\frac{g(X_i,\lambda) - \mu}{\sigma}\right)^2\right) \prod_{i=1}^n X_i^{\lambda - 1} \end{aligned}$$

Then the log-likelihood function is:

$$\begin{aligned} LL(\mu, \sigma, \lambda | X_1, ..., X_n) &= \ln(L(\mu, \sigma, \lambda | X_1, ..., X_n)) \\ &= -n\ln(\sigma\sqrt{2\pi}) + \sum_{i=1}^n \frac{1}{2\sigma^2}(g(X_i, \lambda) - \mu)^2 + (\lambda - 1)\sum_{i=1}^n \ln(X_i) \\ &= -n\ln(\lambda) - n\ln(\sqrt{2\pi}) + \frac{1}{2\sigma^2}\left[\sum_{i=1}^n g(X_i, \lambda)^2 - 2\mu\sum_{i=1}^n g(X_i, \lambda) + \sum_{i=1}^n \mu^2\right] + (\lambda - 1)\sum_{i=1}^n \ln(X_i) \end{aligned}$$

Exercise 19

First of all, for the exercise to make sense and the thing we want to prove to be provable, assume $p_0 \neq 0$ and $p_1 \neq 0$. Note that F is not differentiable at $x = 0$ or $x = 1$ due to the fact that, since neither are null, there is a jump. Assume $h > 0$.

First of all, notice that for $X_i = 1$, we have that $F(X_i + h|\theta) = p_0 + (1 - p_0 - p_1) + p_1$ and $F(X_i - h|\theta) = p_0 + (1 - p_0 - p_1)F_0(X_i - h|\gamma)$. Since F is not differentiable here, we do not divide by $2h$ (by the proper normalization mentioned in the exercise). Then:

$$\begin{aligned}\lim_{h\to 0+} F(X_i + h|\theta) - F(X_i - h|\theta) &= \lim_{h\to 0+} p_0 + (1 - p_0 - p_1) + p_1 - (p_0 + (1 - p_0 - p_1)F_0(X_i - h|\gamma)) \\ &= p_0 + (1 - p_0 - p_1) + p_1 - (p_0 + (1 - p_0 - p_1)) \\ &= p_1\end{aligned}$$

Similarly, for $X_i = 0$ we have that $F(X_i + h|\theta) = p_0 + (1 - p_0 - p_1)F_0(x_i + h|\gamma)$ and $F(X_i - h|\theta) = (1 - p_0 - p_1)F_0(X_i - h|\gamma)$. Again, here F is not differentiable, then:

$$\begin{aligned}\lim_{h\to 0+} F(X_i + h|\theta) - F(X_i - h|\theta) &= \lim_{h\to 0+} p_0 + (1 - p_0 - p_1)F_0(x_i + h|\gamma) - ((1 - p_0 - p_1)F_0(X_i - h|\gamma)) \\ &= p_0\end{aligned}$$

Finally, if $X_i \in (0, 1)$, F is differentiable at X_i. Therefore for the limit we do normalize by dividing by $2h$. We can assume h to be small enough that $X_i + h < 1$ and $X_i - h \geq 0$ (since we are taking limits). Then:

$$\begin{aligned}\lim_{h\to 0+} \frac{F(X_i + h|\theta) - F(X_i - h|\theta)}{2h} &= \lim_{h\to 0+} \frac{(p_0 + (1 - p_0 - p_1)F_0(X_i + h|\gamma)) - (p_0 + (1 - p_0 - p_1)F_0(X_i - h|\gamma))}{2h} \\ &= (1 - p_0 - p_1)f_0(X_i|\gamma)\end{aligned}$$

Then the likelihood function is given by:

$$\begin{aligned}L(\theta|X_1, ..., X_n) &= \lim_{h\to 0+} \prod_{\{i:X_i=1\}} [F(X_i + h|\theta) - F(X_i - h|\theta)] \prod_{\{i:X_i=0\}} [F(X_i + h|\theta) - F(X_i - h|\theta)] \cdots \\ &\cdots \prod_{\{i:X_i\in(0,1)\}} \left[\frac{F(X_i + h|\theta) - F(X_i - h|\theta)}{2h}\right] \\ &= \prod_{\{i:X_i=1\}} p_1 \prod_{\{i:X_i=0\}} p_0 \prod_{\{i:X_i\in(0,1)\}} (1 - p_0 - p_1)f_0(X_i|\gamma) \\ &= p_1^{n_1} p_0^{n_0} (1 - p_0 - p_1)^{n-n_1-n_0} \prod_{\{i:X_i\in(0,1)\}} f_0(X_i|\gamma)\end{aligned}$$

Exercise 20

Let X be the randomly selected serial number. Then X is distributed as a discrete uniform random variable with support $1, ..., N$, where N is the parameter we wish to estimate. Notice $P(X = k) = \frac{1}{N}$, then

$$E[X] = \sum_{i=1}^{N} \frac{i}{N} = \frac{N(N+1)}{2N} = \frac{N+1}{2}$$

With a sample size of 1 and $X = 888$, $\bar{X} = 888$. Using the method of moments, we solve $888 = \frac{N+1}{2}$ to get our MoM estimate as $\hat{N} = 2 \cdot 888 - 1 = 1775$.

For maximum likelihood, notice that the likelihood function with one observation is given by $L(N|X) = \frac{1}{N}$. This is not maximizable as it is not bounded as a function of N. However, notice that since it's a strictly decreasing function of N, we maximize the likelihood by minimizing N. However, we know N cannot be less than 888 as that was our selected serial number. Then, $\max_{N \leq 888} \frac{1}{N}$ occurs exactly when $N = 888$. Therefore $N_{MLE} = 888$.

Exercise 21

Let θ be the probability of heads. Then $1 - \theta$ is the probability of tails. Notice that our sample is given by $\{T, T, T, T, T, T, H\}$. Then the likelihood function is given by:

$$L(\theta|X) = (1-\theta)^6 \theta \Rightarrow LL(\theta|X) = 6 \ln(1-\theta) + \ln \theta$$

To maximize the log-likelihood:

$$\begin{aligned} \frac{d}{d\theta} LL(\theta|X) = \frac{-6}{1-\theta} + \frac{1}{\theta} = 0 &\iff \frac{1}{\theta} = \frac{6}{1-\theta} \\ &\iff \frac{1}{\theta} - 1 = 6 \\ &\iff \frac{1}{\theta} = 7 \\ &\iff \theta = \frac{1}{7} \end{aligned}$$

Exercise 23

Let's start by calculating the first moments in general, which will be useful regardless of if we know one parameter or not. Recall $f(x|\mu, \sigma) = \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right)$ for $x \geq \mu$. Then:

$$\begin{aligned} E[X] &= \int_\mu^\infty x \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) dx \\ &\text{Making the substitution } y = \frac{x-\mu}{\sigma} \\ &= \int_0^\infty (\sigma y + \mu) e^{-y} dy \\ &= \sigma \int_0^\infty y e^{-y} dy + \mu \int_0^\infty e^{-y} dy \\ &= \sigma + \mu \end{aligned}$$

$$\begin{aligned} E[X^2] &= \int_\mu^\infty x^2 \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) dx \\ &\text{Making the substitution } y = \frac{x-\mu}{\sigma} \\ &= \int_0^\infty (\sigma y + \mu)^2 e^{-y} dy \\ &= \sigma^2 \int_0^\infty y^2 e^{-y} dy + 2\sigma \int_0^\infty y e^{-y} dy + \mu^2 \int_0^\infty e^{-y} dy \\ &= 2\sigma^2 + 2\sigma\mu + \mu^2 \end{aligned}$$

Then:

$$\begin{cases} M_1 = \sigma + \mu \\ M_2 = 2\sigma^2 + 2\sigma\mu + \mu^2 \end{cases}$$

Solving the system yields:

$$\begin{cases} \mu = M_1 - \sqrt{M_2 - M_1^2} \\ \sigma = \sqrt{M_2 - M_1^2} \end{cases}$$

a)

Assume μ is known. For the method of moments, as we saw above $M_1 = \mu + \sigma$, then $\sigma = M_1 - \mu$. We can estimate σ by taking the empirical first moment from a sample X, as follows:

$$\sigma_{MoM} = \frac{1}{n}\sum_{i=1}^{n} x_i - \mu$$

Now for maximum likelihood firs notice that the likelihood function is given by:

$$L(\sigma|X,\mu) = \prod_{i=1}^{n} \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma^n}\prod_{i=1}^{n}\exp\left(-\frac{x-\mu}{\sigma}\right)$$

Then the log-likelihood is given by:

$$LL(\sigma|X,\mu) = -n\ln(\sigma) - \frac{1}{\sigma}\sum_{i=1}^{n}(X_i - \mu)$$

Which we maximize:

$$\begin{aligned} \frac{d}{d\sigma}LL(\sigma|X,\mu) = \frac{-n}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0 &\iff \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = \frac{n}{\sigma} \\ &\iff \sigma = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu) \\ &\iff \sigma = \frac{1}{n}\sum_{i=1}^{n} x_i - \mu \end{aligned}$$

We see then that $\sigma_{MoM} = \sigma_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i - \mu$

b)

Assume now that both σ and μ are unknown. For the method of moments, recall as we calculated above that:

$$\begin{cases} \mu = M_1 - \sqrt{M_2 - M_1^2} \\ \sigma = \sqrt{M_2 - M_1^2} \end{cases}$$

Plug in the empirical moments $\hat{M}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\hat{M}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ to obtain the MoM estimates:

$$\begin{cases} \mu_{MoM} = \hat{M}_1 - \sqrt{\hat{M}_2 - \hat{M}_1}^2 \\ \sigma_{MoM} = \sqrt{\hat{M}_2 - \hat{M}_1}^2 \end{cases}$$

Now, for the MLE method, notice the same likelihood and log-likelihood as in a) apply, with the difference that now both μ and σ are unknown:

$$LL(\mu, \sigma|X) = -n\ln(\sigma) - \frac{1}{\sigma}\sum_{i=1}^{n}(X_i - \mu)$$

In this case, we want to maximize with respect to both variables, which means we need the gradient to be null.

$$\frac{d}{d\mu}LL = \frac{n}{\sigma}$$
$$\frac{d}{d\sigma}LL = \frac{-n}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)$$

Notice that the partial derivative with respect to μ is always positive, will never be zero. This means that the likelihood increases as a function of μ. However, notice that from the definition of the distribution, we have bounded μ as $X \geq \mu$. This means that the largest possible value that is reasonable for μ given observations $X_1, ..., X_n$ is given by $\mu_{MLE} = min\{X_1, ..., X_n\}$. We can now use this estimation of μ to solve the second equation as in part a):

$$\sigma_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i - \mu_{MLE}$$

Exercise 24

Given the Lomax distribution, note the likelihood function is given by:

$$L(\alpha|X, \lambda) = \prod_{i=1}^{n}\frac{\alpha\lambda^{\alpha}}{(\lambda + X_i)^{\alpha+1}}$$

Then the log-likelihood is given by:

$$LL(\alpha|X, \lambda) = n\ln(\alpha) + n\alpha\ln(\lambda) - (\alpha + 1)\sum_{i=1}^{n}\ln(\lambda + X_i)$$

For the maximization, note:

$$\begin{aligned} \frac{d}{d\alpha}LL(\alpha|X, \lambda) = \frac{n}{\alpha} + n\ln\lambda - \sum_{i=1}^{n}\ln(\lambda + X_i) = 0 &\iff \frac{n}{\alpha} = \sum_{i=1}^{n}\ln(\lambda + X_i) - n\ln\lambda \\ &\iff \alpha = \frac{n}{\sum_{i=1}^{n}\ln(\lambda + X_i) - n\ln\lambda} \\ &\iff \alpha = \frac{n}{\sum_{i=1}^{n}\ln\left(1 + \frac{X_i}{\lambda}\right)} \end{aligned}$$

Then $\hat{\alpha} = \frac{n}{\sum_{i=1}^{n}\ln\left(1+\frac{X_i}{\lambda}\right)}$

Now, let's find the sampling distribution of $\hat{\alpha}/\alpha$. Notice that:

$$
\begin{aligned}
\frac{\hat{\alpha}}{\alpha} &= \frac{n}{\alpha \sum_{i=1}^{n} \ln\left(1+\frac{X_i}{\lambda}\right)} \\
&= \frac{-n}{\sum_{i=1}^{n} \ln\left(\left(1+\frac{X_i}{\lambda}\right)^{-\alpha}\right)} \\
&= \frac{-n}{\sum_{i=1}^{n} \ln\left(1-1+\left(1+\frac{X_i}{\lambda}\right)^{-\alpha}\right)} \\
&= \frac{-n}{\sum_{i=1}^{n} \ln\left(1-\left(1-\left(1+\frac{X_i}{\lambda}\right)^{-\alpha}\right)\right)} \\
&= \frac{-n}{\sum_{i=1}^{n} \ln\left(1-F(X_i)\right)} \\
&\text{Due to PIT }, U_i \sim U(0,1) \\
&= \frac{-n}{\sum_{i=1}^{n} \ln(U_i)} \\
&= \frac{1}{\sum_{i=1}^{n} -\ln\left(U_i^{1/n}\right)}
\end{aligned}
$$

Now, notice that $-\ln\left(U_i^{1/n}\right) \sim Exp(n)$, since:

$$
\begin{aligned}
P\left(-\ln\left(U_i^{1/n}\right) \leq x\right) &= P(U_i \geq e^{-nx}) \\
&= 1-e^{-nx}
\end{aligned}
$$

This means that the characteristic function of $-\ln\left(U_i^{1/n}\right)$ is given by:

$$
\begin{aligned}
E\left[\exp(it(-\ln\left(U_i^{1/n}\right)))\right] &= E\left[U_i^{-it/n}\right] \\
&= \int_0^1 x^{-it/n} dx \\
&= \frac{1}{1-it/n} \\
&= \frac{n}{n-it}
\end{aligned}
$$

Then the characteristic function of the sum of the logarithms is given by:

$$
\begin{aligned}
E\left[\exp\left(it\sum_{i=1}^{n} -ln(U_i^{1/n})\right)\right] &= \prod_{i=1}^{n} E\left[\exp(it(-\ln\left(U_i^{1/n}\right)))\right] \\
&= \left(\frac{n}{n-it}\right)^n \\
&= \left(1-\frac{it}{n}\right)^{-n}
\end{aligned}
$$

This is the characteristic function for a gamma distribution with shape parameter n and scale parameter $1/n$. Recall if $X \sim Gamma(\alpha, \beta)$ where β is the scale, then $1/X \sim Inv.Gamma(\alpha, 1/\beta)$. Then we obtain that:

$$\frac{\hat{\alpha}}{\alpha} = \frac{1}{\sum_{i=1}^{n} - \ln\left(U_i^{1/n}\right)} \sim Inv.Gamma(n, n)$$

To find an expression for the standard error, consider rather the parametrization of an inverse gamma using shape and scale rather than shape and rate. Then we have that $\hat{\alpha}/\alpha \sim Inv.Gamma(n, 1/n)$. That means $\hat{\alpha} \sim \alpha Inv.Gamma(n, 1/n) \sim Inv.Gamma(n, \alpha/n)$. Then:

$$\begin{aligned}
SE(\hat{\alpha}) &= \sqrt{Var[\hat{\alpha}]} \\
&= \sqrt{\frac{(\alpha/n)^2}{(n-1)^2(n-2)}} \\
&= \sqrt{\frac{\alpha^2}{n^2(n-1)^2(n-2)}} \\
&= \frac{\alpha}{n(n-1)\sqrt{n-2}} \\
&\approx \frac{\hat{\alpha}}{n(n-1)\sqrt{n-2}}
\end{aligned}$$

Finally, we can also calculate the asymptotic standard error, for which we will use the fact that the Fisher information matrix's inverse provides us with the covariance matrix of the estimates. This means that each estimate's variance will be given by the diagonal elements of this matrix. Since here we only estimate one parameter, the value of this information will be its reciprocal. Note from here onwards we return to the shape-rate parametrization.

$$\begin{aligned}
I(\theta) &= -E\left[\frac{d^2}{d\alpha^2} LL\right] \\
&= -E\left[\frac{d}{d\alpha}\left(\frac{n}{\alpha} + n\ln\lambda - \sum_{i=1}^{n}\ln(\lambda + X_i)\right)\right] \\
&= -nE\left[\frac{-1}{\alpha^2}\right] \\
&= \frac{n}{\alpha^2}
\end{aligned}$$

Then the asymptotic standard error is given by

$$SE(\hat{\alpha}) = \sqrt{\frac{\alpha^2}{n}} = \frac{\alpha}{\sqrt{n}} \approx \frac{\hat{\alpha}}{\sqrt{n}}$$

Exercise 25

a)

Let $W = max\{X - 40000, 0\}$, where $f(x|\alpha) = \frac{\alpha \cdot 20000^{\alpha}}{(20000+x)^{\alpha+1}}$. I.e. X is Lomax with $\lambda = 20000$. Note that $P(W < 0) = 0$, $P(W = 0) = P(X \leq 40000) = 1 - \left(\frac{20000}{60000}\right)^{\alpha}$, and for $y > 0$ we have:

$$\begin{aligned}
F_w(y) &= P(W \leq y) \\
&= P(max\{X - 40000, 0\} \leq y) \\
&= P(X - 40000 \leq y) \\
&= P(X \leq 40000 + y) \\
&= F_X(40000 + y)
\end{aligned}$$

Then notice that we have a discontinuous "density'' function, as there is clearly a mass at 0. Then the density is fiven by:

$$f_W(y) = \begin{cases} 0 & y < 0 \\ 1 - \left(\frac{20000}{60000}\right)^\alpha & y = 0 \\ f_x(40000 + y) & y > 0 \end{cases}$$

To calculate the mean and variance, note we don't pay mind to the discontinuity since the realization of W there is 0. It does not affect any moment's values.

$$\begin{aligned} E[W] &= \int_0^\infty x \frac{\alpha \cdot 20000^\alpha}{(60000 + x)^{\alpha+1}} dx \\ &= \alpha \cdot 20000^\alpha \int_0^\infty \frac{x}{(60000 + x)^{\alpha+1}} dx \\ &= \alpha \cdot 20000^\alpha \frac{60000^{1-\alpha}}{\alpha(\alpha - 1)} \\ &= 20000^\alpha \frac{60000^{1-\alpha}}{\alpha - 1} \end{aligned}$$

Similarly:

$$\begin{aligned} E[W^2] &= \int_0^\infty x^2 \frac{\alpha \cdot 20000^\alpha}{(60000 + x)^{\alpha+1}} dx \\ &= \alpha \cdot 20000^\alpha \int_0^\infty \frac{x^2}{(60000 + x)^{\alpha+1}} dx \\ &= 20000^\alpha \frac{2^{11-5\alpha} \cdot 1875^{2-\alpha}}{\alpha^2 - 3\alpha + 2} \end{aligned}$$

Which yields that the variance is given by:

$$\begin{aligned} Var[W] &= E[W^2] - E[W]^2 \\ &= 20000^\alpha \frac{2^{11-5\alpha} \cdot 1875^{2-\alpha}}{\alpha^2 - 3\alpha + 2} - \left(20000^\alpha \frac{60000^{1-\alpha}}{\alpha - 1}\right)^2 \end{aligned}$$

b)

To calculate the MLE estimator of α, note that e can use a trick here. Note all payments are possitive, so we can transform our variable from W_i to X_i and use the fact that X is distributed as a Lomax with known $\lambda = 20000$. So, we compute $X_i = W_i + 40000$, and now as proven in previous execises:

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^n ln\left(1 + \frac{X_i}{20000}\right)}$$

```
W <- c(14000, 21000, 6000, 32000, 2000)
X <- W + 40000
n <- length(X)

alpha_MLE <- n/sum(log( 1 + X/20000 ))
alpha_MLE
```

```
## [1] 0.7623775
```

To estimate the standard error, we will use both formulas found in the previous exercise. First, the distribution "direct'' estimation of the standard error:

$$SE(\hat{\alpha}) \approx \frac{\hat{\alpha}}{n(n-1)\sqrt{n-2}}$$

```
se_alpha <- alpha_MLE/(n*(n-1)*sqrt(n-2))
se_alpha
```

```
## [1] 0.02200794
```

And finally an asymptotic estimator using Fisher's information. Note, however, that with a sample of 5 observations it's a little optimistic to assume the normality convergence argument used for this result.

$$SE(\hat{\alpha}) \approx \frac{\hat{\alpha}}{\sqrt{n}}$$

```
a_se_alpha <- alpha_MLE/sqrt(n)
a_se_alpha
```

```
## [1] 0.3409456
```

Exercise 26

Note that the actual payment is given by $W = min\{X, M\}$. Now, we have two cases, if $y < M$, then $P(W \leq y) = P(X \leq y) = 1 - \left(\frac{\lambda}{\lambda+y}\right)^{\alpha}$. If $y \geq M$ then $P(W \leq y) = 1$

Now, notice that $P(X \geq M) = \left(\frac{\lambda}{\lambda+M}\right)^{\alpha}$. then the likelihood function is given by:

$$L = \prod_{i=1}^{n} \frac{\alpha\lambda^{\alpha}}{(\lambda+x)^{\alpha+1}} \prod_{i=1}^{m} \left(\frac{\lambda}{\lambda+M}\right)^{\alpha}$$

Then the log-likelihood is:

$$LL = \sum_{i=1}^{n} [\ln(\alpha) + \alpha\ln(\lambda) - (\alpha+1)\ln(\lambda+X_i)] + m\ln(\lambda) - m\alpha\ln(\lambda+M)$$

For the maximization, note:

$$\begin{aligned}
\frac{d}{d\alpha}LL &= \frac{n}{\alpha} + \sum_{i=1}^{n} \ln\left(\frac{\lambda}{\lambda+X_i}\right) + m\ln\left(\frac{\lambda}{\lambda+M}\right) \\
&= \frac{n}{\alpha} - \sum_{i=1}^{n} \ln\left(1+\frac{X_i}{\lambda}\right) - m\ln\left(1+\frac{M}{\lambda}\right) \\
&= 0 \qquad\qquad \iff \alpha = \frac{n}{\sum_{i=1}^{n}\ln\left(1+\frac{X_i}{\lambda}\right) + m\ln\left(1+\frac{M}{\lambda}\right)}
\end{aligned}$$

Let's calculate the MLE with the given data:

```
X <- c(14.9, 775.7, 805.2, 993.9, 1127.5, 1602.5, 1998.3)
M <- rep(2000, 3)

n <- length(X)
```

```r
m <- length(M)
lambda <- 8400

alpha_MLE = n/ ( sum(log(1 + X/lambda)) + m*log(1+2000/lambda) )
alpha_MLE
```

```
## [1] 4.833747
```

Exercise 27

For the method of moments, as shown in exercise 17:

$$\begin{cases} \mu_1 = \frac{\lambda}{\alpha-1} \\ \mu_2 - \mu_1^2 = \frac{\lambda^2\alpha}{(\alpha-1)^2(\alpha-2)} \end{cases}$$

Solving for α, λ yields that:

$$\begin{aligned} \lambda = \mu_1(\alpha - 1) \Rightarrow \mu_2 - \mu_1^2 &= \frac{\mu_1^2\alpha}{\alpha - 2} \\ \Rightarrow \frac{\mu_2 - \mu_1^2}{\mu_1^2} &= \frac{\alpha}{\alpha - 2} \\ \Rightarrow \frac{\mu_1^2}{\mu_2 - \mu_1^2} &= 1 - \frac{2}{\alpha} \\ \Rightarrow \alpha = 2 - 2\frac{\mu_1^2}{\mu_2} \\ \Rightarrow \lambda = \mu_1 \left(1 - 2\frac{\mu_1^2}{\mu 2}\right) \end{aligned}$$

Next, as used in multiple exercises so far, recall the log-likelihood is given by:

$$LL = n\ln(\alpha) + \alpha n \ln(\lambda) - (\alpha + 1)\sum_{i=1}^{n} \ln(\lambda + X_i)$$

We want to maximize this, so we want $\nabla LL = 0$, which means the equations that must be satisfied by the MLE estimators are:

$$\begin{cases} \frac{d}{d\alpha}LL = \frac{n}{\alpha} + n\ln(\lambda) - \sum_{i=1}^{n} \ln(\lambda + X_i) = 0 \\ \frac{d}{d\lambda}LL = \frac{\alpha n}{\lambda} - \sum_{i=1}^{n} \frac{\alpha+1}{\lambda+X_i} = 0 \end{cases}$$

Exercise 28

We will prove this through characteristic functions:

$$
\begin{aligned}
CF(X^\gamma) &= E[\exp(itX^\gamma)] \\
&= \int_0^\infty \exp(itx^\gamma) c\gamma x^{\gamma-1} \exp(-cx^\gamma) dx \\
&= c \int_0^\infty \gamma x^{\gamma-1} \exp(x^\gamma(it-c)) dx \\
&\text{Substituting } u = x^\gamma \\
&= c \int_0^\infty \exp(u(it-c)) du \\
&= c \int_0^\infty \exp(-u(c-it)) du \\
&\text{Substituting } y = u(c-it) \\
&= c \int_0^\infty e^{-y} \frac{dy}{c-it} \\
&= \frac{c}{c-it}
\end{aligned}
$$

This is exactly the characteristic function of an exponential with rate parameter c. This could be useful in terms of method of moments estimation since, if γ is known then c can be estimated using the (potentially abusing notation) γ-th moment of X. This is due to the fact that if $Y \sim Exp(\lambda)$ then $E[Y] = 1/\lambda$. Therefore, knowing γ we can estimate:

$$E[X^\gamma] = \frac{1}{c} \approx \frac{1}{n}\sum_{i=1}^{n} X_i^\gamma$$

If γ is unknown, then it would be possible to use numeric methods to approximate the values through the facct that if $Y \sim Exp(\lambda)$, then $E[Y^k] = \frac{k!}{\lambda^k}$. This gives us a link between the moments of the Weibull and the exponential distribution. One potential way to do this would be to numerically solve the system of equations:

$$
\begin{cases}
E[X^\gamma] \approx \hat{\mu}_\gamma = \frac{1}{n}\sum_{i=1}^{n} X_i^\gamma = \frac{1}{c} \\
E[X] \approx \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \left(\frac{1}{c}\right)^{1/\gamma}\Gamma(1+1/\gamma)
\end{cases}
$$

However this is a bit complicated. Let's look now at the quantiles instead. First of all, note that the quantile function is given by:

$$q = 1 - e^{-cx^\gamma} \iff x = \left(\frac{-\ln(1-q)}{c}\right)^{1/\gamma}$$

In particular, note that:

$$q = 1 - e^{-cx^\gamma} \iff -cx^\gamma = ln(1-q)$$

Let $X_{(1)}$ be the sample 1st quartile, $X_{(3)}$ the third. Then:

$$
\begin{aligned}
cX_{(1)}^{\gamma} = -ln(0.75) \Rightarrow c &= \frac{-ln(0.75)}{X_{(1)}^{\gamma}} \\
&\Rightarrow \frac{-ln(0.75)}{X_{(1)}^{\gamma}} X_{(3)}^{\gamma} = -ln(0.25) \\
&\Rightarrow \left(\frac{X_{(3)}}{X_{(1)}}\right)^{\gamma} = \frac{ln(0.25)}{ln(0.75)} \\
&\Rightarrow \gamma ln\left(\frac{X_{(3)}}{X_{(1)}}\right) = ln\left(\frac{ln(0.25)}{ln(0.75)}\right) \\
&\Rightarrow \gamma = ln\left(\frac{ln(0.25)}{ln(0.75)}\right)/ln\left(\frac{X_{(3)}}{X_{(1)}}\right) \\
&\Rightarrow c = \frac{-ln(0.75)}{X_{(1)}^{ln\left(\frac{ln(0.25)}{ln(0.75)}\right)/ln\left(\frac{X_{(3)}}{X_{(1)}}\right)}}
\end{aligned}
$$

Exercise 29

Notice that the likelihood is given by:

$$L(\mu, \sigma|X) = \prod_{i=1}^{n} \frac{1}{\sigma} e^{-\left(\frac{X_i-\mu}{\sigma}\right)} e^{-e^{-\left(\frac{X_i-\mu}{\sigma}\right)}}$$

Then the log-likelihood is:

$$
\begin{aligned}
L(\mu, \sigma|X) &= \sum_{i=1}^{n} \left(-ln(\sigma) - \frac{X_i-\mu}{\sigma} - e^{-\frac{X_i-\mu}{\sigma}}\right) \\
&= -nln(\sigma) - \sum_{i=1}^{n} \frac{X_i-\mu}{\sigma} - \sum_{i=1}^{n} e^{-\frac{X_i-\mu}{\sigma}} \\
&= -nln(\sigma) - \frac{1}{\sigma}\sum_{i=1}^{n} X_i + \frac{n\mu}{\sigma} - \sum_{i=1}^{n} e^{-\frac{X_i-\mu}{\sigma}}
\end{aligned}
$$

```r
Rivers <- c(5500, 4380, 2370, 3220, 8050, 4560, 2100, 6840, 5640, 3500, 1940,
            7060, 7500, 5370, 13100, 4920, 6500, 4790, 6050, 4560, 3210, 6450,
            5870, 2900, 5490, 3490, 9030, 3100, 4600, 3410, 3690, 6420, 10300,
            7240, 9130)

n <- length(Rivers)

LL <- function(mu,sigma) {
  return((-n)*log(sigma) - 1/sigma*(sum(Rivers) - n*mu) - sum(exp(-(Rivers-mu)/sigma)))
}

# initial values for parameters
mean <- mean(Rivers)
sd <- sd(Rivers)

mu <- optim(c(mean,sd), function(c) -LL(c[1],c[2]), method = "BFGS")$par[1]
sigma <- optim(c(mean,sd), function(c) -LL(c[1],c[2]), method = "BFGS")$par[2]
mu
```

```
## [1] 4501.432
```

```r
sigma
```

```
## [1] 1933.551
```

When we need to compare the goodness of fit using QQ plot, we need to calculate a quantile function. For the quantile function we first need the CDF of our distribution which can be calculated by integrating the density function which is given.

$$
\begin{aligned}
F(y) &= \int_{-\infty}^{y} \frac{1}{\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)} e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}} dx \\
&\text{Substitute } u = exp\left(-\frac{x-\mu}{\sigma}\right) \\
&= \int_{exp\left(-\frac{y-\mu}{\sigma}\right)}^{\infty} e^{-u} du \\
&= exp\left(-exp\left(-\frac{y-\mu}{\sigma}\right)\right)
\end{aligned}
$$

Then the quantile function is given by:

$$F^{-1}(q) = \mu - \sigma ln(-ln(q))$$

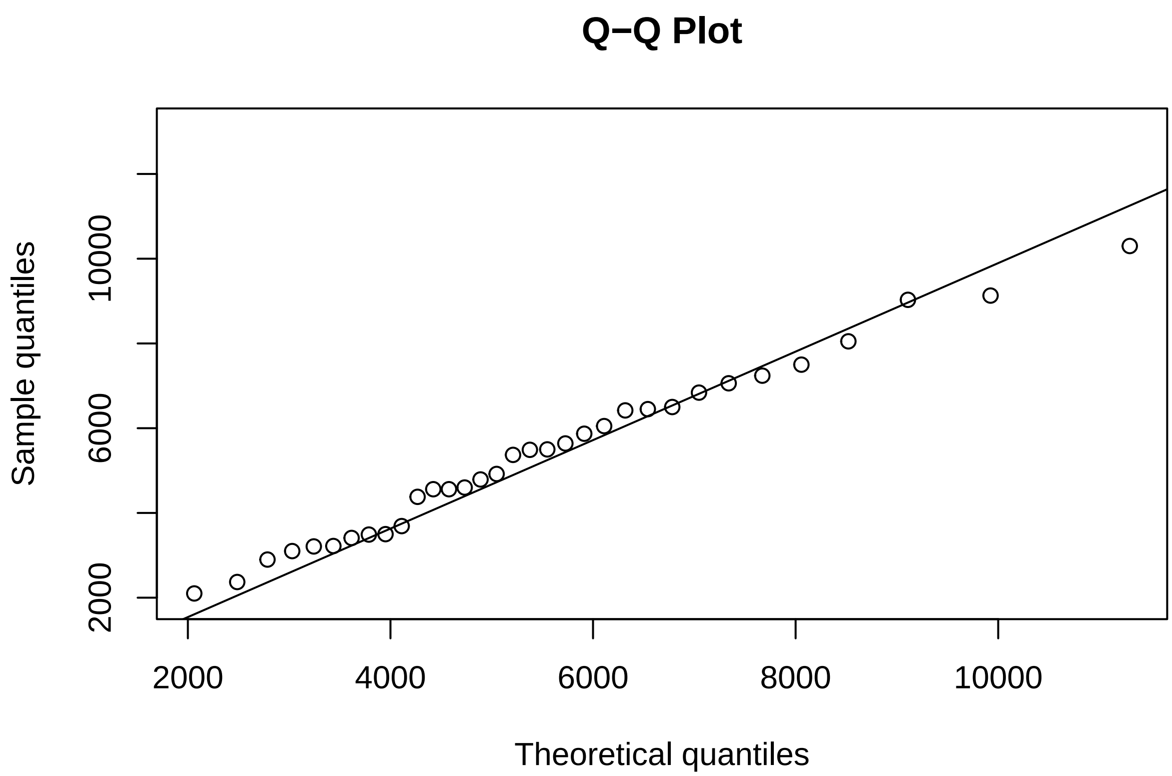
We can now calculate theoretical quantiles and compare them:

```r
CDFinverse <- function(a) mu-sigma*log(-log(a))

quantiles <- seq(0,1,0.01)

theo_q <- CDFinverse(quantiles)

qqplot(theo_q, Rivers,
       xlab = "Theoretical quantiles",
       ylab = "Sample quantiles",
       main = "Q-Q Plot")
qqline(y = quantile(Rivers, probs = quantiles),
       distribution = CDFinverse)
```



We observe a good fit with between the observed data and the theoretical distribution with the estimated parameters.

Exercise 31

The claim amounts are exponentially distributed with rate parameter λ. However, since any excess claim above 10000 units is handled by reinsurance, the claim amount is constant at 10000 for claims that are $\geq$ 10000.

Let x_i represent the amount of the i-th claim from the 68 claims beneath 10000 and let y_i represent the i-th claim from the 12 claims above 10000.

Note that density function of an exponential distribution is $f(x) = \lambda e^{-\lambda x}$.

To find the MLE $\hat{\lambda}$, we first need the joint density,

$$f(x_1, \cdots, x_{68}, y_1, \cdots, y_{12}|\lambda) = \prod_{i=1}^{68} \lambda e^{-\lambda x_i} * e^{-12\lambda y_1}$$

Now, we apply the logarithm to the joint density to get the log-likelihood, and substitution $y = 10000$:

$$\ell(\lambda) = log(\lambda^{68} \prod_{i=1}^{68} e^{-\lambda x_i} * e^{-12 y_i \lambda})$$

$$= 68log(\lambda) - \lambda \sum_{i=1}^{68} x_i - 12 y_i \lambda$$

$$= 68log(\lambda) - \lambda \sum_{i=1}^{68} x_i - 120,000\lambda$$

Computing the derivative with respect to parameter λ of the log-likelihood function and setting it to zero will maximize the likelihood function,

$$\ell^{'}(\lambda) = \frac{68}{\lambda} - \sum_{i=1}^{68} x_i - 120000 = 0$$

Which yields the MLE

$$\hat{\lambda} = \frac{68}{\sum_{i=1}^{68} x_i + 120000}$$

and we know that the sum of the 68 claims below 10000 is 220000, so we can substitute and get the

$$= \frac{68}{220000 + 120000} = \frac{68}{340000}$$

As per slide 55, using the asymptotic normality of the Maximum Likelihood Estimator, $\sqrt{(nI(\hat{\lambda})}(\hat{\lambda} - \lambda_0)$ is approximately standard normally distributed ($\mathbb{N}(1, 0)$). We can use this fact and calculate a 95 percent confidence interval ($\alpha = 0.05$) with n=80. The confidence is defined as following:

$$1 - \alpha = \mathbb{P}\Big(z_{\frac{\alpha}{2}} \leq \sqrt{nI(\hat{\lambda})}(\hat{\lambda} - \lambda) \leq z_{1-\frac{\alpha}{2}}\Big)$$

Simplifying and isolating the λ variable we get the result that can be calculated in R.

$$\mathbb{P}\Big(\hat{\lambda} + \frac{z_{\frac{\alpha}{2}}}{\sqrt{nI(\hat{\lambda})}} \leq \lambda \leq \hat{\lambda} - \frac{z_{\frac{\alpha}{2}}}{\sqrt{nI(\hat{\lambda})}}\Big)$$

Now, note that:

$$\begin{aligned} I(\lambda) &= -E\left[\frac{d^2}{d\lambda^2}\ell(\lambda)\right] \\ &= -E\left[\frac{-68}{\lambda^2}\right] \\ &= \frac{68}{\lambda^2} \end{aligned}$$

Then the approximate 95% confidence interval is given by:

$$\hat{\lambda} \pm z_{0.05/2} / \sqrt{n\frac{68}{\hat{\lambda}^2}}$$

```
MLE <- 68/340000

CI <- c(MLE + qnorm(0.025)/sqrt(80*68/MLE^2), MLE - qnorm(0.025)/sqrt(80*68/MLE^2))
CI
```

```
## [1] 0.0001946853 0.0002053147
```

Exercise 32

a)

Let $X_1, ..., X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$. Recall that the MLE estimator for μ is given by $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$. We therefore have a sum of independent normal distributions multiplied by a constant, all of them with the same mean and variance. This is again a normal random variable, but rescaled due to the $1/n$ multiplicative factor.

Recall the sum of normally distributed independent random variables simply adds up the means and the variances. Since in our case all of them are the same (due to i.i.d), we have that:

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{n\mu}{n} = \mu$$
$$Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Thus, $\hat{\mu} \sim N(\mu, \sigma^2/n)$. When applying bootstrap, due to the fact that μ and σ are still unknown parameters, we use the MLE estimates as plug-ins, and sample from this to implement the bootstrapping. Therefore, we estimate the distribution by:

$$N(\hat{\mu}, \hat{\sigma}^2/n)$$

b)

Recall from a) that the bootstrap estimate of the distribution of $\hat{\mu}$ is normal with mean $\hat{\mu}$. When subtracting a scalar from a normal distribution, we essentially shift the distribution (which is still normal), but we do not affect its variance whatsoever. Therefore the variance of the estimate of the distribution remains $\hat{\sigma}^2/n$. However, recall that the MLE for the mean of a normal r.v. is an unbiased estimator, i.e. $E[\hat{\mu}] = \mu$. Therefore, $E[\hat{\mu} - \mu] = E[\hat{\mu}] - \mu = 0$. This means we shift the normal distribution horizontally exactly so its mean is 0. Therfore the bootstrap estimate of the sampling of $\hat{\mu} - \mu$ follows a $N(0, \hat{\sigma}^2/n)$ distribution.

c)

Using the bootstrap estimated distribution, we have shown in the previous point that $\hat{\mu} - \mu \sim N(0, \hat{\sigma}^2/n)$, and as seen in the class notes the confidence interval will be given by:

$$\left(\bar{X} - Q_{N(0,\hat{\sigma}^2/n)}(1-\alpha/2), \bar{X} + Q_{N(0,\hat{\sigma}^2/n)}(1-\alpha/2)\right)$$

For the real confidence interval, recall from class notes that the α confidence interval for μ using the t distribution is given by:

$$\left(\bar{X} - \frac{S}{\sqrt{n}} Q_{t_{n-1}}(1-\alpha/2), \bar{X} + \frac{S}{\sqrt{n}} Q_{t_{n-1}}(1-\alpha/2)\right)$$

Exercise 33

Let θ be the real parameter tht we estimate through $\hat{\theta}$. Recall that if the distribution of $\Delta = \hat{\theta} - \theta$, then:

$$\begin{aligned}
1 - \alpha &= P(Q_\Delta(\alpha/2) \leq \hat{\theta} - \theta \leq Q_\Delta(1-\alpha/2)) \\
&= P(Q_\Delta(\alpha/2) - \hat{\theta} \leq -\theta \leq Q_\Delta(1-\alpha/2) - \hat{\theta}) \\
&= P(\hat{\theta} - Q_\Delta(\alpha/2) \geq \theta \geq \hat{\theta} - Q_\Delta(1-\alpha/2))
\end{aligned}$$

Now, we don't know the distribution of Δ, but we can approximate it through $\theta^* - \hat{\theta}$. Since $\hat{\theta}$ is a scalar, then the quantiles of $\theta^* - \hat{\theta}$ will be equal to the quantiles of θ^* minus the scalar $\hat{\theta}$. We can therefore estimate:

$$\hat{\theta} - Q_\Delta(\alpha/2) \approx \hat{\theta} - (\theta_L - \hat{\theta}) = 2\hat{\theta} - \theta_L$$
$$\hat{\theta} - Q_\Delta(1-\alpha/2) \approx \hat{\theta} - (\theta_U - \hat{\theta}) = 2\hat{\theta} - \theta_U$$

We can plug in this estimator in the previous equation to get that:

$$1 - \alpha = P(2\hat{\theta} - \theta_L \geq \theta \geq 2\hat{\theta} - \theta_U)$$

Which implies the bootstrap confidence interval can be written as $(2\hat{\theta} - \theta_U, 2\hat{\theta} - \theta_L)$.

Now, let's assume symmetry of θ^* around $\hat{\theta}$. Then the $\alpha/2$ and $1 - \alpha/2$ quantiles are equidistant to $\hat{\theta}$. In other words, $\hat{\theta} - \theta_L = \theta_U - \hat{\theta}$, which means $\hat{\theta} = (\theta_U + \theta_L)/2$. Substituting this in the expression for the confidence interval we obtain that:

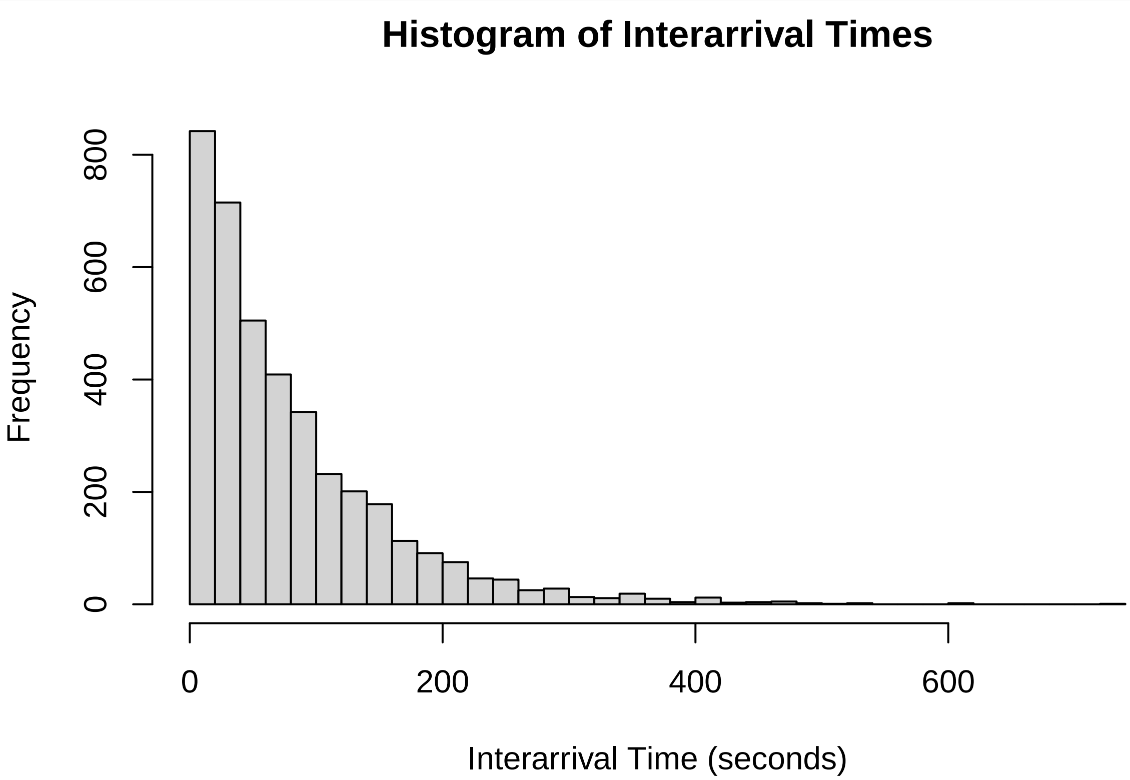
$$\begin{aligned}(2\hat{\theta} - \theta_U, 2\hat{\theta} - \theta_L) &= (2(\theta_U + \theta_L)/2 - \theta_U, 2(\theta_U + \theta_L)/2 - \theta_L) \\ &= (\theta_L, \theta_U)\end{aligned}$$

Exercise 34

a)

```
setwd("~/Documents/QFin/Q3/Statistics 2/Assignments/HW2")
g <- read.table("gamma-arrivals.txt") |>
  unlist() |>
  as.numeric()

hist(g,
     main   = "Histogram of Interarrival Times",
     xlab   = "Interarrival Time (seconds)",
     breaks = 50)
```



From the shape of the distribution, a Gamma could be a potential fit.

b)

We know from the lecture notes that the Method of Moments estimates for the Gamma distribution are

$$\hat{\alpha} = \frac{\bar{X}^2}{\hat{\sigma}^2} \text{ and } \hat{s} = \frac{\hat{\sigma}^2}{\bar{X}}$$

where $\hat{\alpha}$ is the sample shape estimate and $\hat{s}$ is the sample scale estimate. For the Maximum Likelihood estimates, we use package MASS, which allows us to fit a Gamma distribution to our data using MLE. Implementing this in R:

```
mom_shape <- (mean(g))^2 / var(g)
mom_scale <- var(g) / mean(g)

fit_mle <- MASS::fitdistr(g, "gamma")
mle_shape <- fit_mle$estimate["shape"]
mle_scale <- 1 / fit_mle$estimate["rate"]

dif_shape <- mom_shape - mle_shape
dif_scale <- mom_scale - mle_scale

data.frame(Parameter = c("Shape", "Scale"),
           MoM       = c(mom_shape, mom_scale),
           MLE       = c(mle_shape, mle_scale),
           Dif       = c(dif_shape, dif_scale),
           row.names = NULL)
```

```
##   Parameter       MoM       MLE         Dif
## 1     Shape  1.012095  1.026332 -0.01423717
## 2     Scale 78.979962 77.884342  1.09562021
```

As can be seen, there is very little difference between both estimates.

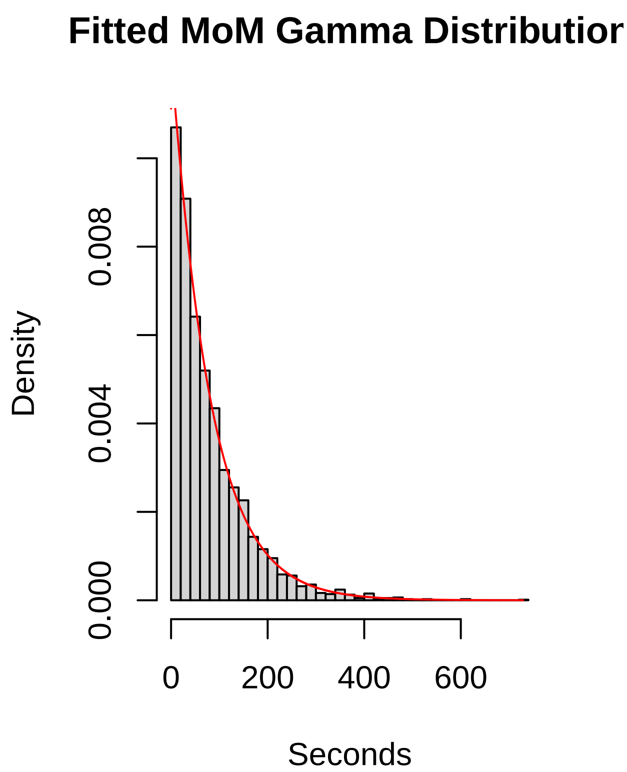
c)

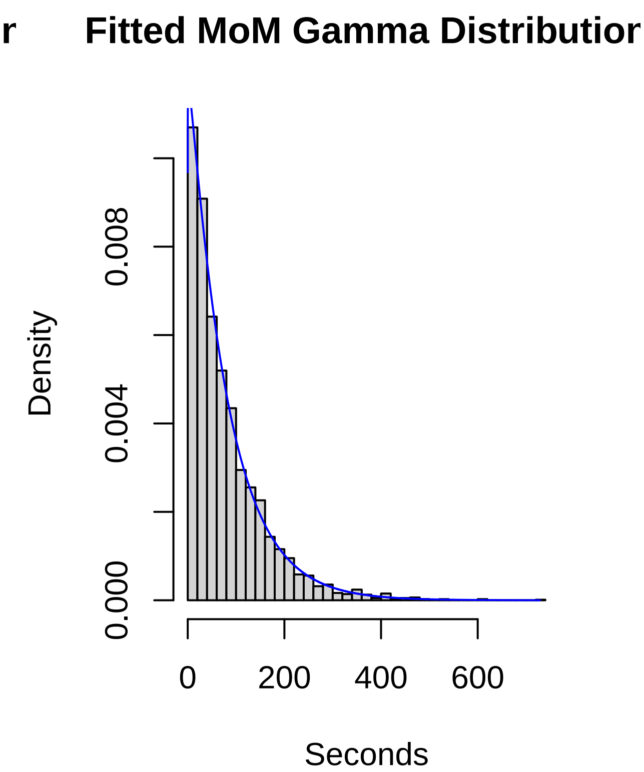
```
par(mfrow = c(1, 2))

x <- seq(0.001, max(g), by = 0.05)

hist(g,
     probability = T,
     main        = "Fitted MoM Gamma Distribution",
     xlab        = "Seconds",
     breaks      = 50)
lines(x, dgamma(x, mom_shape, scale = mom_scale), col = "red")

hist(g,
     probability = T,
     main        = "Fitted MoM Gamma Distribution",
     xlab        = "Seconds",
     breaks      = 50)
lines(x, dgamma(x, mle_shape, scale = mle_scale), col = "blue")
```





Both plots seem to reasonably fit the interarrivals data distribution.

d)

```r
N <- 1e3

mom_est <- matrix(NA, nrow = N, ncol = 2)
mle_est <- matrix(NA, nrow = N, ncol = 2)

set.seed(8)

# Perform bootstrap (takes a long time to run)
for (i in 1:N) {
  bootstrap_data <- sample(g, replace = TRUE)

  sample_mean <- mean(bootstrap_data)
  sample_variance <- var(bootstrap_data)

  mom_shape_boot <- (sample_mean^2) / sample_variance
  mom_scale_boot <- sample_variance / sample_mean

  fit_mle_boot <- suppressWarnings(MASS::fitdistr(bootstrap_data, "gamma"))

  mle_shape_boot <- (fit_mle_boot$estimate["shape"][[1]])
  mle_scale_boot <- (1 / fit_mle_boot$estimate["rate"][[1]])

  mom_est[i, ] <- c(mom_shape_boot, mom_scale_boot)
  mle_est[i, ] <- c(mle_shape_boot, mle_scale_boot)
}
```

```r
# Calculate confidence intervals
alpha <- 0.05
mom_lower_bounds <- apply(mom_est, 2, function(x) quantile(x, probs = alpha/2))
mom_upper_bounds <- apply(mom_est, 2, function(x) quantile(x, probs = 1 - alpha/2))

mle_lower_bounds <- apply(mle_est, 2, function(x) quantile(x, probs = alpha/2))
mle_upper_bounds <- apply(mle_est, 2, function(x) quantile(x, probs = 1 - alpha/2))

data.frame(`Method.Parameter` = c("MoM.Shape", "MLE.Shape", "MoM.Scale", "MLE.Scale"),
           `Lower.Bound`      = c(mom_lower_bounds[1], mle_lower_bounds[1], mom_lower_bounds[2], mle_lo
           `Upper.Bound`      = c(mom_upper_bounds[1], mle_upper_bounds[1], mom_upper_bounds[2], mle_up
           row.names = NULL)
```

```
##   Method.Parameter Lower.Bound Upper.Bound
## 1        MoM.Shape   0.9511951    1.077534
## 2        MLE.Shape   0.9901454    1.066614
## 3        MoM.Scale  73.2963238   84.517904
## 4        MLE.Scale  74.0153344   81.726616
```

It can seen from the table that confidence intervals for the MLE are narrower than the ones for the MoM.

e)

In a Poisson process, the times between events are expected to follow an exponential distribution. When considering the Gamma distribution, it becomes identical to the exponential distribution when the shape parameter α is set to 1. Since the confidence intervals for α, derived from both the method of moments and the maximum likelihood estimation, include 1, the suggestion that the interarrival times align with those of a Poisson process cannot be statistically dismissed.