MA677 final project

Lihao Liao

Question 4.25

The density of the order statistic $U_{(i)}$is given by

$$
\begin{aligned}
f_{U_{(i)}}(u) &= \frac{n!}{(i-1)!(n-i)!}\{F(x)\}^{i-1}\{1-F(x)\}^{n-i}f(x) \\
&= \frac{n!}{(i-1)!(n-i)!}u^{i-1}(1-u)^{n-i} \quad 0<u<1 \\
&= constant * u^{i-1}(1-u)^{n-i} \quad 0<u<1
\end{aligned}
$$

Thus $U_{(i)}$ follows a beta distribution, denoted by $U_{(i)} \sim Beta(i, n+1-i)$

The median of Beta distribution $Beta(\alpha, \beta)$ is given by

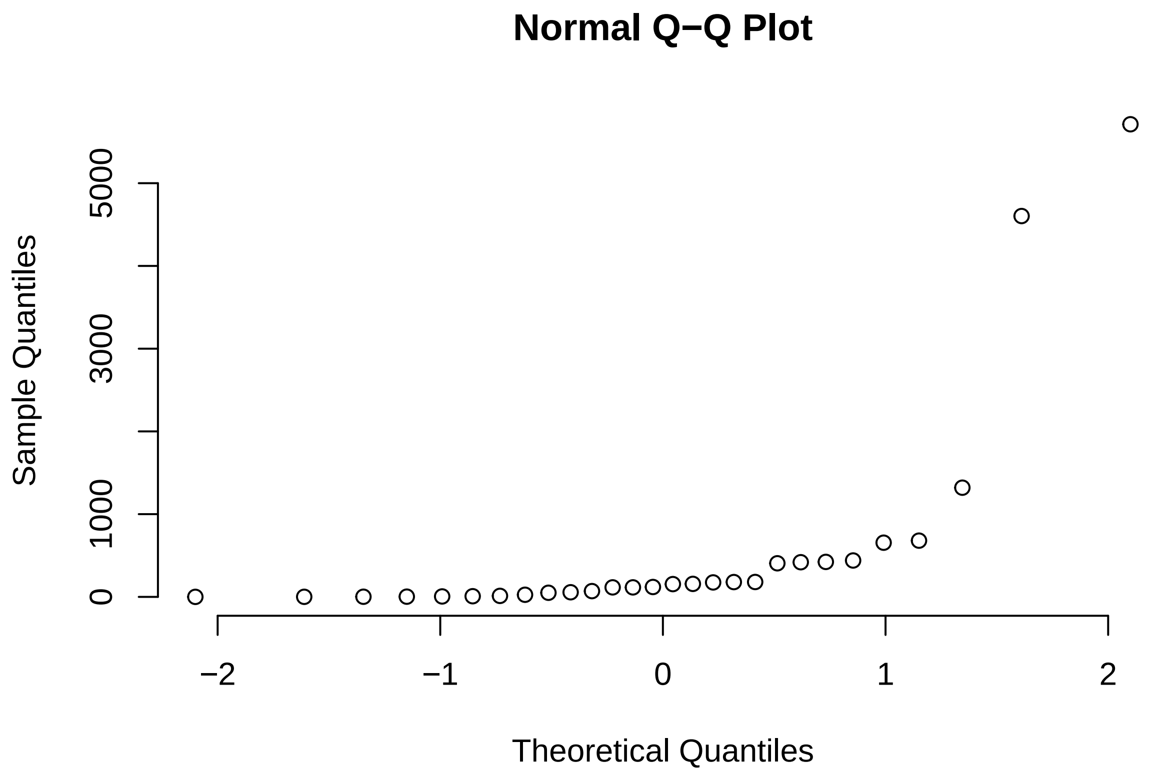
$$
\begin{aligned}
median &\approx \frac{\alpha - 1/3}{\alpha + \beta - 2/3} \\
median(U_{(i)}) &\approx \frac{i - 1/3}{i + n + 1 - i - 2/3} \\
&= \frac{i - 1/3}{n + 1/3}
\end{aligned}
$$

Question 4.39

```
X = c(0.4, 1.0, 1.9, 3.0, 5.5,  8.1, 12.1, 25.6,
115.0  , 119.5,   154.5 , 157.0 ,  175.0,
419.0 ,  423.0,   440.0,  655.0  , 680.0,
  50.0 , 56.0, 70.0 , 115.0,
 179.0,     180.0,   406.0,
1320.0,    4603.0 , 5712.0)
```

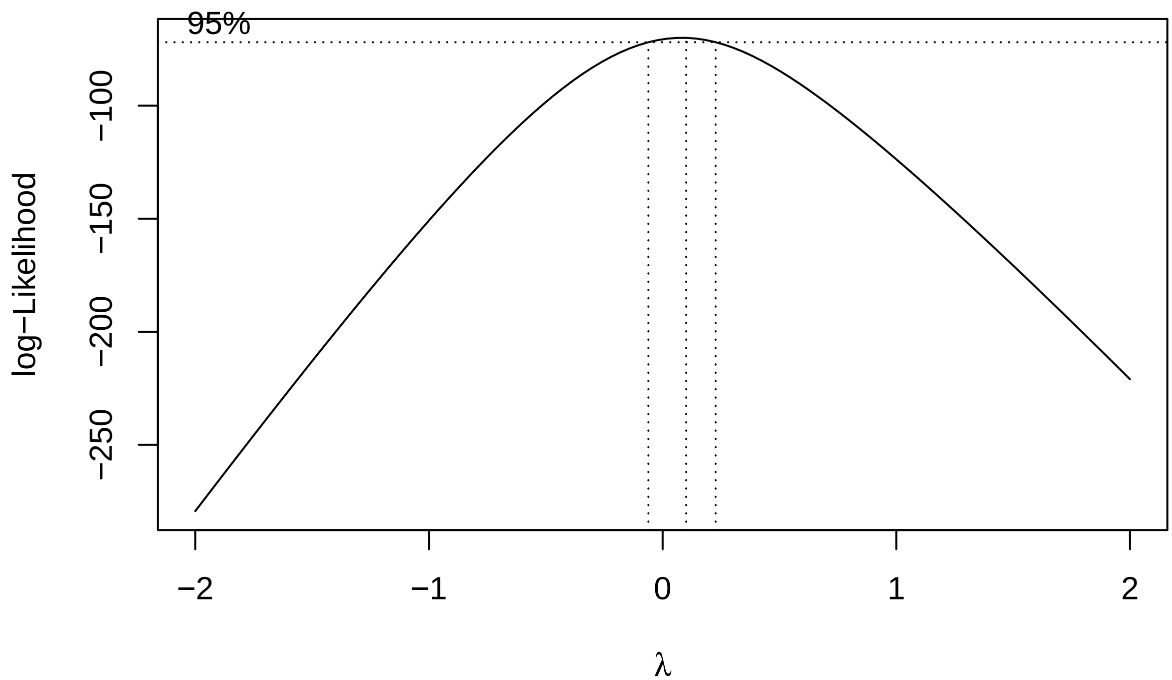
From the following plot we can see that, the sample data are skewed to the right.

```
qqnorm(X, pch = 1, frame = FALSE)
```



The profile log-likelihood of λ is maximized at $\hat{\lambda} = 0.101$, which indicates that the log-transform $\lambda = 0$ is a sensible transformation.

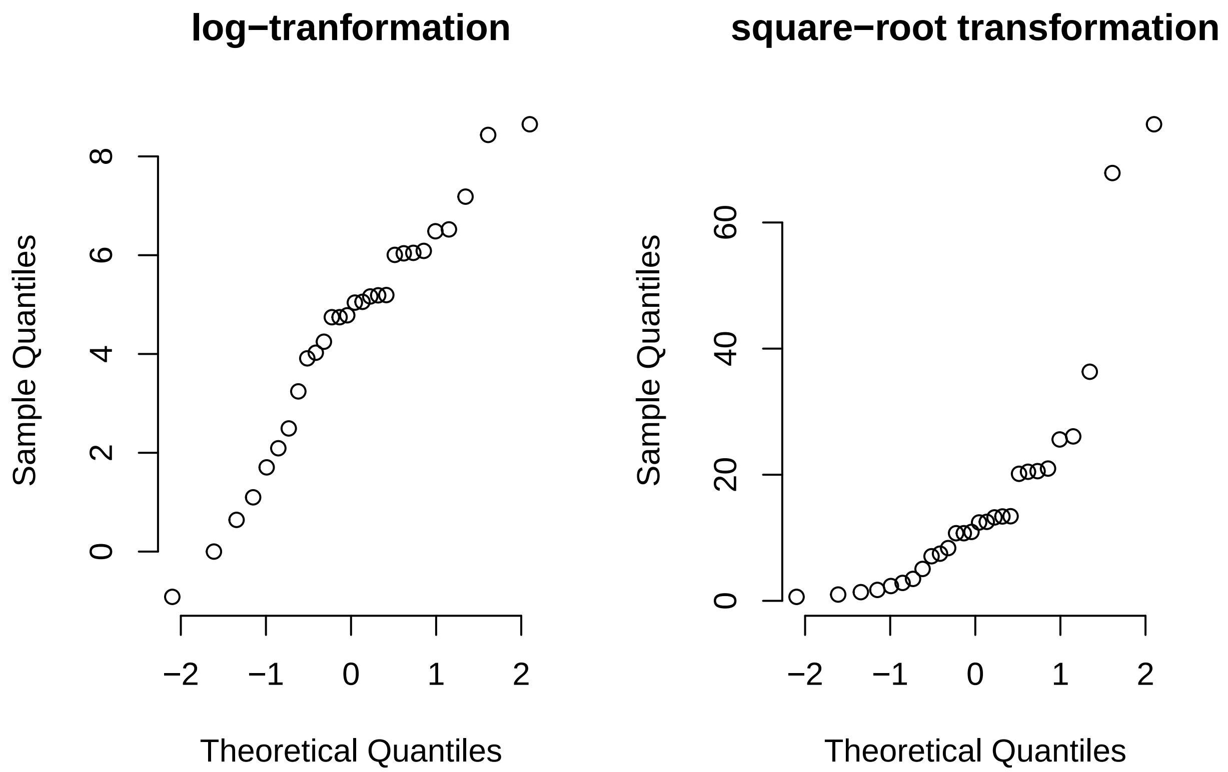
```
b=boxcox(lm(X ~ 1))
```




```
lambda = b$x[which.max(b$y)]
```

The following two QQ plots suggest that log-transformation is the sensible transformation.

```
X.log = log(X)
X.rt = sqrt(X)
par(mfrow=c(1,2))
qqnorm(X.log, pch = 1, frame = FALSE, main ="log-tranformation")
qqnorm(X.rt, pch = 1, frame = FALSE, main ="square-root transformation")
```




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

Question 4.27

```
dat1 = c(0.15, 0.25, 0.10, 0.20, 1.85, 1.97, 0.80, 0.20, 0.10, 0.50, 0.82, 0.40,
      1.80, 0.20, 1.12, 1.83, 0.45, 3.17, 0.89, 0.31, 0.59, 0.10, 0.10, 0.90,
      0.10, 0.25, 0.10, 0.90)

dat2  = c(0.30, 0.22, 0.10, 0.12, 0.20, 0.10, 0.10, 0.10, 0.10, 0.10, 0.10 ,0.17,
      0.20, 2.80, 0.85, 0.10, 0.10, 1.23, 0.45, 0.30, 0.20, 1.20, 0.10, 0.15,
      0.10, 0.20, 0.10, 0.20, 0.35, 0.62, 0.20, 1.22, 0.30, 0.80, 0.15, 1.53,
      0.10, 0.20, 0.30, 0.40, 0.23, 0.20, 0.10, 0.10, 0.60, 0.20, 0.50, 0.15,
      0.60, 0.30, 0.80, 1.10 ,0.20, 0.10, 0.10, 0.10, 0.42, 0.85, 1.60, 0.10,
      0.25, 0.10, 0.20, 0.10)
```

(a)

As we can see from the following table, the average amount of rainfall per storm in January is generally higher than that of July while they have the same amount of minimum averge rainfall.

```r
sum1 = summary(dat1)
sum2 = summary(dat2)

res = data.frame(cbind(sum1, sum2))
colnames(res) = c("January", "July")
kable(res, caption = "Summary statistics for the two months")
```

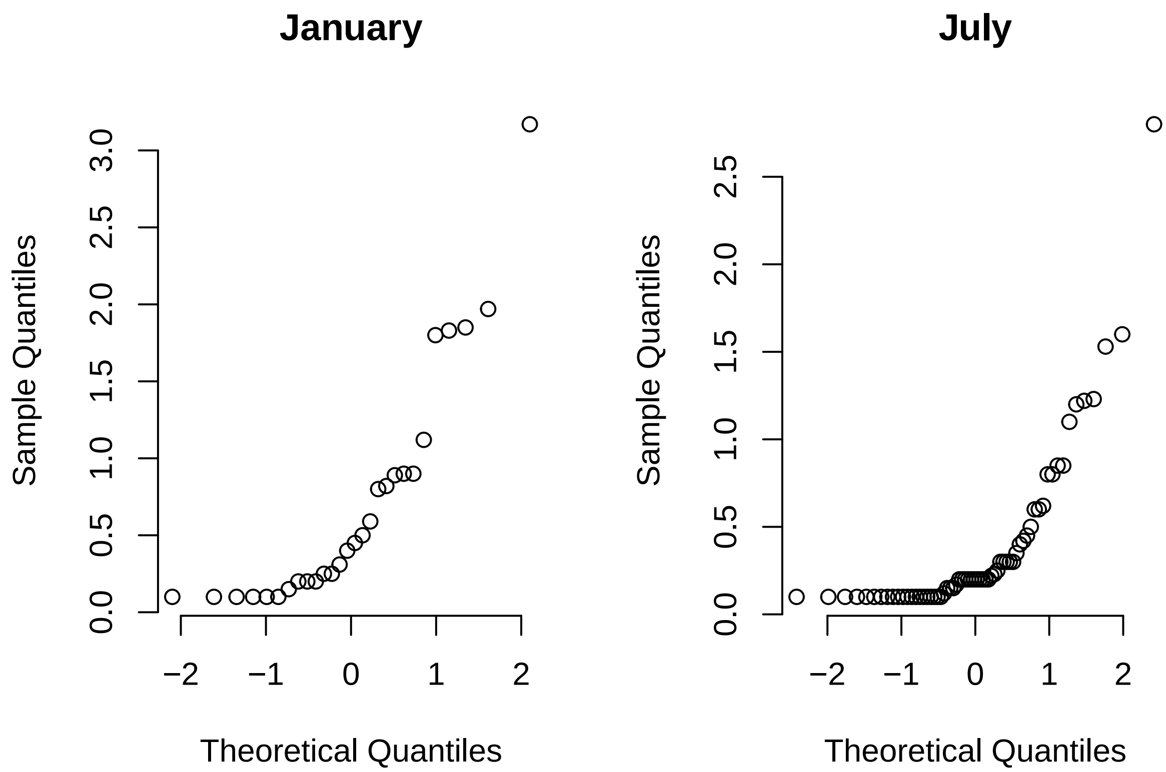
Table 1: Summary statistics for the two months

	January	July
Min.	0.1000000	0.100000
1st Qu.	0.1875000	0.100000
Median	0.4250000	0.200000
Mean	0.7196429	0.393125
3rd Qu.	0.9000000	0.427500
Max.	3.1700000	2.800000

(b)

The QQ-plots suggests that the data are right skewed. Since average amount of rainfall is continuous and positive data, a gamma model could be reasonable.

```r
par(mfrow=c(1,2))
qqnorm(dat1, pch = 1, frame = FALSE, main ="January")
qqnorm(dat2, pch = 1, frame = FALSE, main ="July")
```




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

(c)

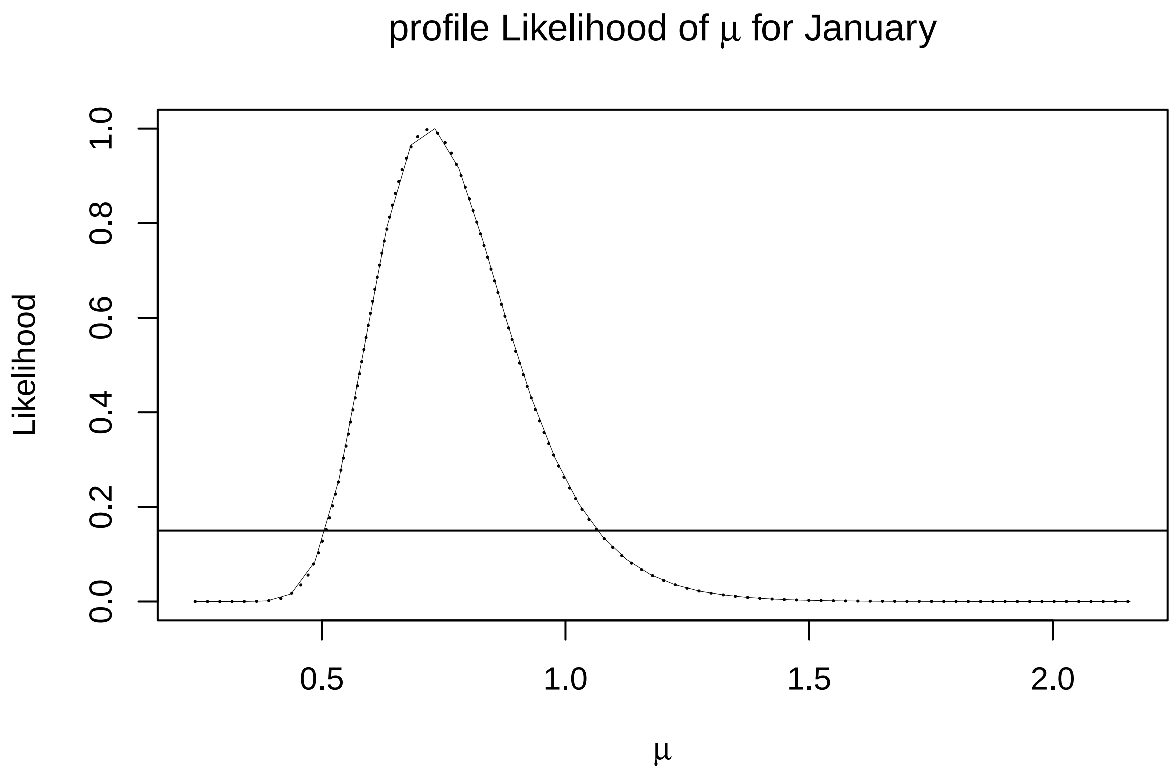
The density of the $gamma(\alpha, \lambda)$ model is given by

$$
\begin{aligned}
f(x) &= \frac{1}{\Gamma(\alpha)}\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}, \quad x > 0 \\
&= \frac{1}{x\Gamma(1/\phi)}\left(\frac{x}{\phi\mu}\right)^{1/\phi}\exp\left(-\frac{x}{\phi\mu}\right), \quad x > 0
\end{aligned}
$$

Where μ is the population mean and variance is $\phi\mu^2$.

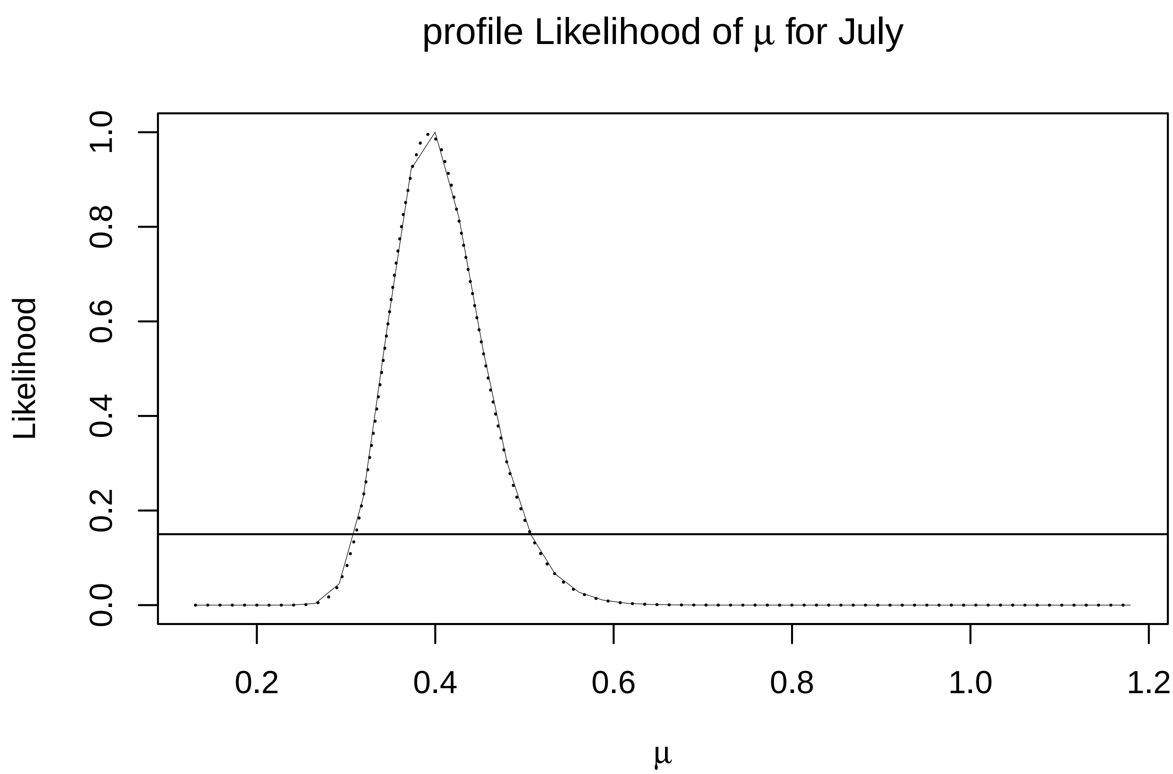
```
fun0= function(mu,phi){
      alpha=1/phi
      lam= 1/(phi*mu)
      a= -(n)*lgamma(alpha)+ n*alpha*log(lam) +
         (alpha-1)*sum(log(dat1))- lam*sum(dat1)
      -a
      }
n= length(dat1)
np= 40
mu1 = mean(dat1)/3
mu2 = mean(dat1)*3
```

```r
mu=  seq(mu1,mu2,len=np)
phi= seq(1,4,np)
alp= 1 / phi
lam= 1 / (phi*mu)
ll2= outer(mu,phi,'fun0')
like2= exp(min(ll2)-ll2)

# par(mfrow=c(2,2))
# contour(mu,phi,like2,
#         xlab=expression(mu),ylab=expression(phi),
#         level=c(.1,.3,.5,.7,.9))
# title(expression('(a) Likelihood contour'))

phi1.mle = phi[which(like2==max(like2),arr.ind = TRUE)[2]]
mu1.mle = mu[which(like2==max(like2),arr.ind = TRUE)[1]]

## se by delta method ##
phi1.se = var(dat1)/(phi1.mle/mean(dat1))^2
mu1.se = var(dat1)/(mu1.mle/mean(dat1))^2

# profile likelihood
like= apply(like2,1,max)
plot(mu,like,xlab=expression(mu),
     ylab='Likelihood',type='n')
lines(mu,like,lwd=.3)
abline(h=.15)
title(expression(paste('profile Likelihood of ',mu," for January")))

np=100
mu= seq(mu1,mu2,len=np)
ll= fun0(phi=phi1.mle,mu)
like= exp(min(ll)-ll)
lines(mu,like,lty='dotted',lwd=1.5)
```




```
n = length(dat2)
fun0= function(mu,phi){
      alpha=1/phi
      lam= 1/(phi*mu)
      a= -(n)*lgamma(alpha)+ n*alpha*log(lam) +
         (alpha-1)*sum(log(dat2))- lam*sum(dat2)
      -a
      }
ll=NULL
np=40
mu1= mean(dat2)/3
mu2 = mean(dat2)*3
mu=  seq(mu1,mu2,len=np)
phi= seq(1,4,len=np)
alp= 1/phi
lam= 1/(phi*mu)
ll2= outer(mu,phi,'fun0')
like2= exp(min(ll2)-ll2)

# par(mfrow=c(2,2))
# contour(mu,phi,like2,
#         xlab=expression(mu),ylab=expression(phi),
#         level=c(.1,.3,.5,.7,.9))
# title(expression('(a) Likelihood contour'))

## MLE
```

```r
phi2.mle = phi[which(like2==max(like2),arr.ind = TRUE)[2]]
mu2.mle = mu[which(like2==max(like2),arr.ind = TRUE)[1]]

## se by delta method
phi2.se = var(dat2)/(phi2.mle/mean(dat2))^2
mu2.se = var(dat2)/(mu1.mle/mean(dat2))^2

# profile likelihood
like= apply(like2,1,max)
plot(mu,like,xlab=expression(mu),
     ylab='Likelihood',type='n')
  lines(mu,like,lwd=.3)
  abline(h=.15)
  title(expression(paste('profile Likelihood of ',mu, " for July")))

np=100
mu= seq(mu1,mu2,len=np)
ll= fun0(phi=phi2.mle,mu)
like= exp(min(ll)-ll)
lines(mu,like,lty='dotted',lwd=1.5)
```




```r
res = data.frame(phi.MLE=c(phi1.mle, phi2.mle), phi.SE = c(phi1.se, phi2.se),
                 mu.MLE=c(mu1.mle, mu2.mle), mu.SE = c(mu1.se, mu2.se))
rownames(res)=c("January", "July")
kable(res, caption = "MLEs and Standard errors for each data")
```

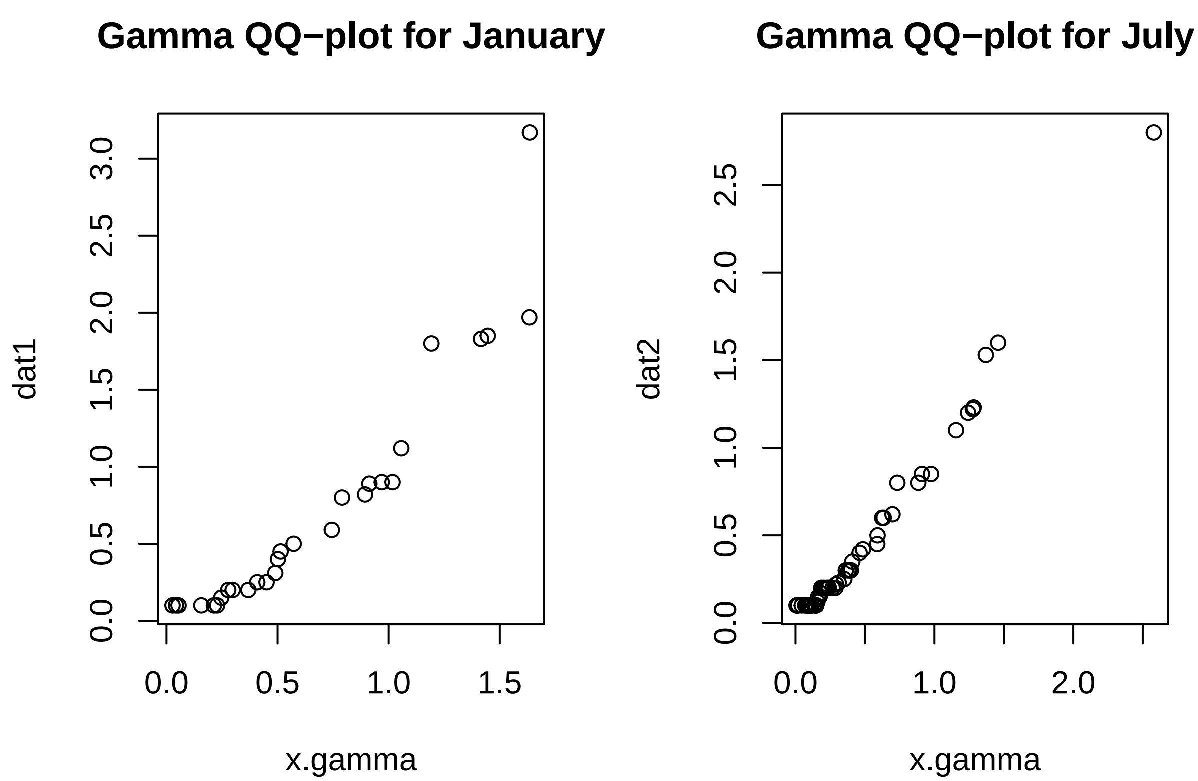
Table 2: MLEs and Standard errors for each data

	phi.MLE	phi.SE	mu.MLE	mu.SE
January	1	0.3031262	0.7319444	0.5658056
July	1	0.0358564	0.3998451	0.0669284

(d)

The gamma QQ-plot suggests that the gamma model is reasonable.

```
par(mfrow=c(1,2))
alp1= 1/phi1.mle
lam1= 1/(phi1.mle*mu1.mle)
x.gamma = rgamma(ppoints(length(dat1)), shape =alp1 , rate = lam1)
dat1 = sort(dat1)
qqplot(x.gamma, dat1, main="Gamma QQ-plot for January")
alp2= 1/phi2.mle
lam2= 1/(phi2.mle*mu2.mle)
x.gamma = rgamma(ppoints(length(dat2)), shape =alp2 , rate = lam2)
dat1 = sort(dat2)
qqplot(x.gamma, dat2, main="Gamma QQ-plot for July")
```



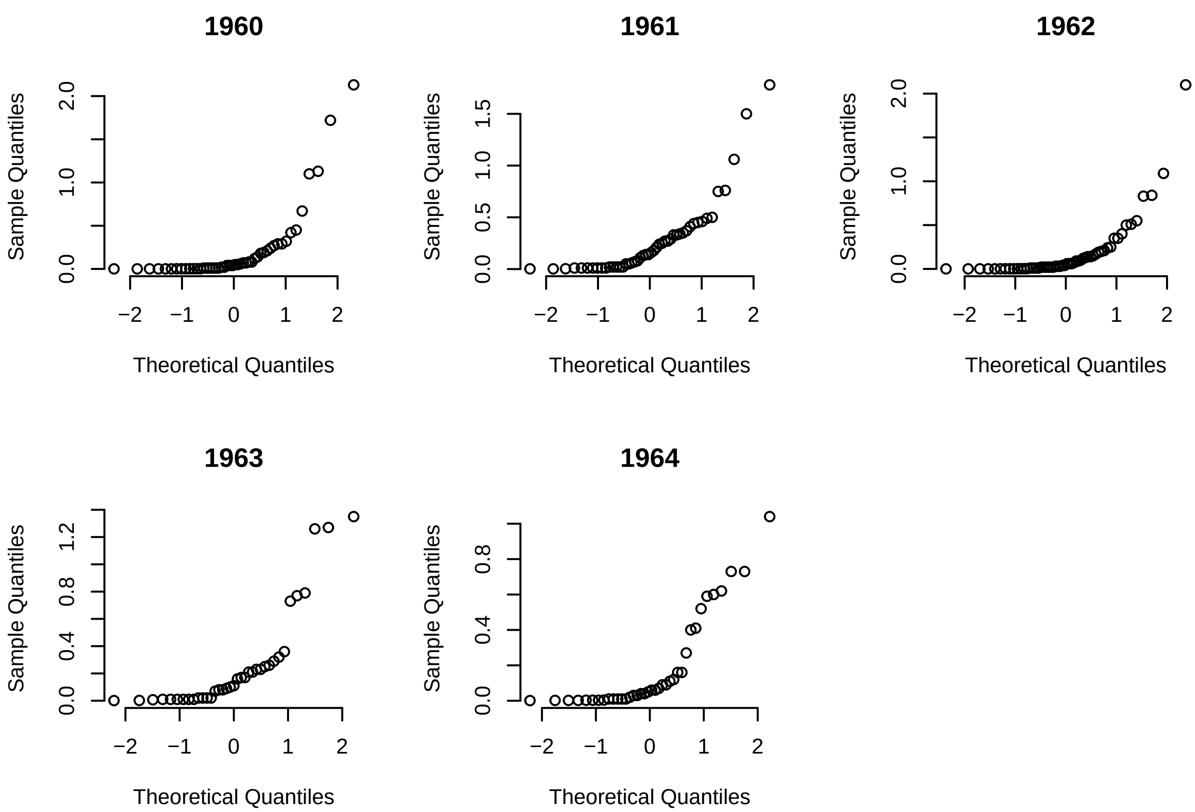
```r
par(mfrow=c(1,1))
```

Question 4

Identify the distribution of rainfall

From the QQ-plot we can see that the data were right skewed. Since the average amount of rainfall is continuous and positive data, a gamma model could be reasonable.

```r
par(mfrow=c(2,3))
qqnorm(data[,1], pch = 1, frame = FALSE, main ="1960")
qqnorm(data[,2], pch = 1, frame = FALSE, main ="1961")
qqnorm(data[,3], pch = 1, frame = FALSE, main ="1962")
qqnorm(data[,4], pch = 1, frame = FALSE, main ="1963")
qqnorm(data[,5], pch = 1, frame = FALSE, main ="1964")
par(mfrow=c(1,1))
```



The density of the $gamma(\alpha, \lambda)$ model is given by

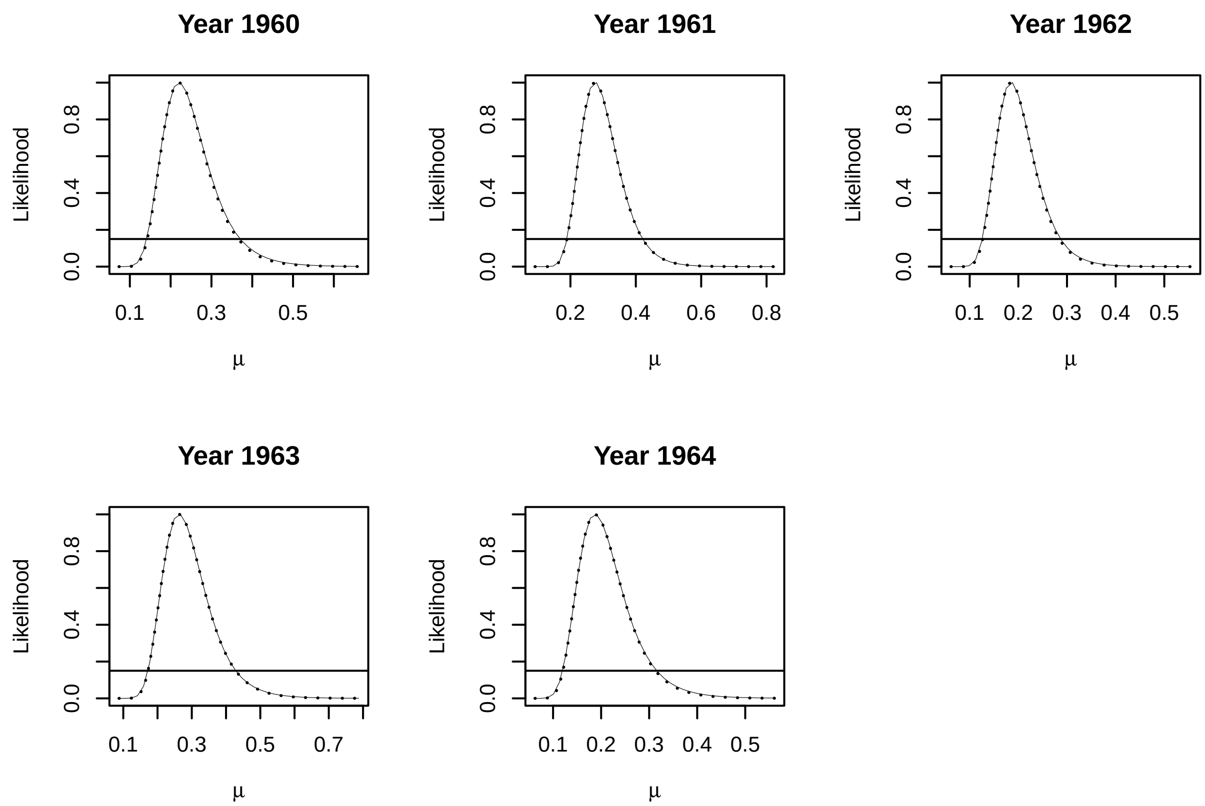
$$
\begin{aligned}
f(x) &= \frac{1}{\Gamma(\alpha)}\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}, \quad x > 0 \\
&= \frac{1}{x\Gamma(1/\phi)}\left(\frac{x}{\phi\mu}\right)^{1/\phi}\exp\left(-\frac{x}{\phi\mu}\right), \quad x > 0
\end{aligned}
$$

Where μ is the population mean and variance is $\phi\mu^2$.

Here are the profile likelihood plot of μ for each year

```
year = c(1960,1961,1962,1963,1964)
## store results
phi.mle = c()
mu.mle = c()
phi.se = c()
mu.se = c()
N = c()

par(mfrow=c(2,3))
for(i in 1:5){
dat = na.omit(data[,i]) ## rainfall each year without NA values
N[i] = length(dat) ## sample size
n = length(dat) ## sample size
fun0= function(mu,phi){
      alpha=1/phi
      lam= 1/(phi*mu)
      a= (-n)*lgamma(alpha)+ n*alpha*log(lam) +
         (alpha-1)*sum(log(dat))- lam*sum(dat)
      -a
      }
ll=NULL
np=40
x.bar = mean(dat, na.rm =T)
mu=  seq(x.bar/3,x.bar*3,len=np)
phi= seq(2,7,len=np)
alp= 1/phi
lam= 1/(phi*mu)
ll2= outer(mu,phi,'fun0')
like2= exp(min(ll2)-ll2)

## MLE
phi.mle[i] = phi[which(like2==max(like2),arr.ind = TRUE)[2]]
mu.mle[i] = mu[which(like2==max(like2),arr.ind = TRUE)[1]]

## se by delta method ##
phi.se[i] = var(dat)/(phi.mle[i]/x.bar)^2
mu.se[i] = var(dat)/(mu.mle[i]/x.bar)^2

# profile likelihood
like= apply(like2,1,max)
plot(mu,like,xlab=expression(mu),
     ylab='Likelihood',type='n')
  lines(mu,like,lwd=.3)
  abline(h=.15)
  title(paste0("Year " ,year[i]))

np=100
mu= seq(x.bar/3, x.bar*3,len=np)
ll= fun0(phi=phi.mle[i],mu)
like= exp(min(ll)-ll)
```

```
lines(mu,like,lty='dotted',lwd=1.5)
}

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



The MLEs and SEs of the gamma distribution parameters for each year are shown in the following table

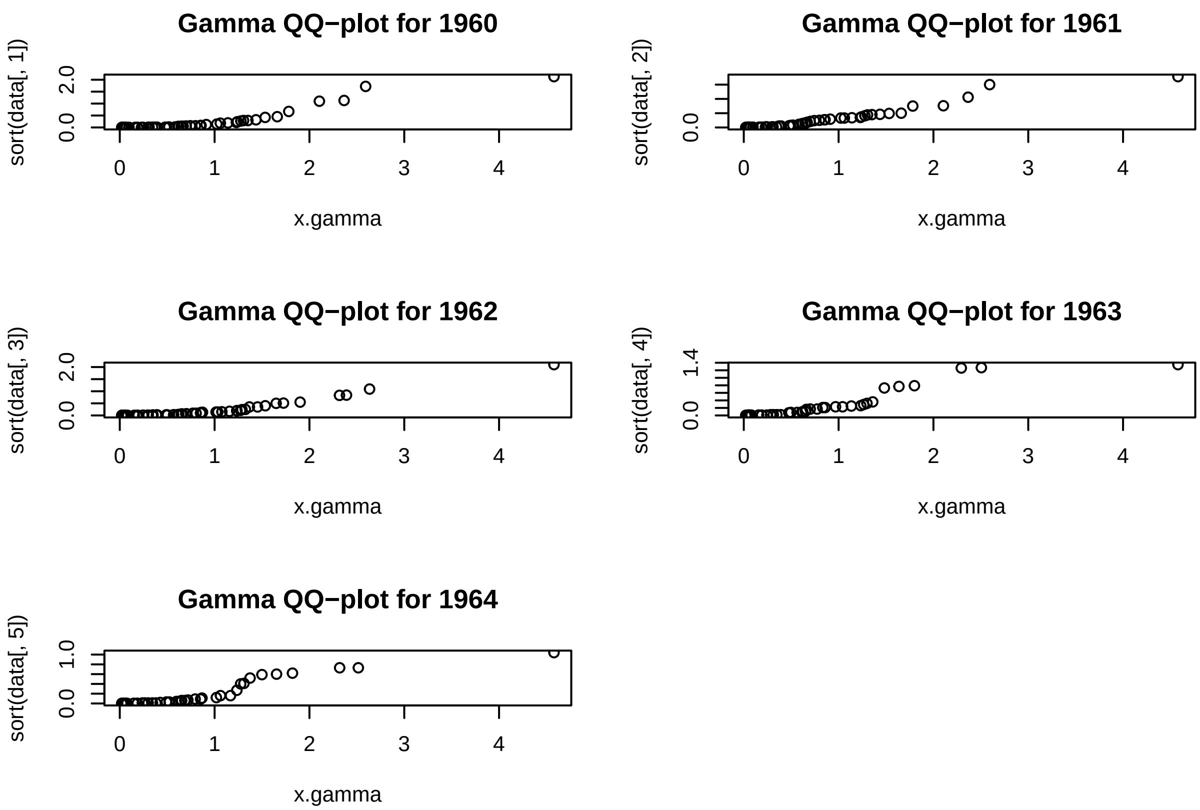
```
res = data.frame(mu.MLE = mu.mle, mu.SE = mu.se, phi.MLE = phi.mle, phi.SE = phi.se , Storms = N)
rownames(res) = c(1960,1961,1962,1963,1964)
kable(res, caption= "MLEs and SEs of the gamma distribution parameters for each year")
```

Table 3: MLEs and SEs of the gamma distribution parameters for each year

	mu.MLE	mu.SE	phi.MLE	phi.SE	Storms
1960	0.2240573	0.1864600	2.769231	0.0012206	48
1961	0.2796373	0.1327703	2.000000	0.0025956	48
1962	0.1879081	0.1179625	2.384615	0.0007325	56
1963	0.2669185	0.1342176	2.000000	0.0023906	37
1964	0.1903036	0.0704215	2.256410	0.0005009	38

The gamma QQ-plot suggests that the gamma model is reasonable.

```
par(mfrow=c(3,2))
alp1= 1/phi1.mle
lam1= 1/(phi1.mle*mu1.mle)
x.gamma = rgamma(ppoints(length(data[,1])), shape =alp1 , rate = lam1)
qqplot(x.gamma, sort(data[,1]), main="Gamma QQ-plot for 1960")
qqplot(x.gamma, sort(data[,2]), main="Gamma QQ-plot for 1961")
qqplot(x.gamma, sort(data[,3]), main="Gamma QQ-plot for 1962")
qqplot(x.gamma, sort(data[,4]), main="Gamma QQ-plot for 1963")
qqplot(x.gamma, sort(data[,5]), main="Gamma QQ-plot for 1964")
par(mfrow=c(1,1))
```



Dry year or Wet year

Based on the distribution of mu, which is the expected average rainfall per storm, it is evident that year 1960, 1961, and 1963 can be viewed as Wet years and the rest of years can be viewed as fry years. They are wet year because there were individual storm produced more rain.

Discussion

Since the values of average amount of rainfall share the same features of being positive, right skewed, and continuous, I think the results of my analysis can be generalized to the any study related to average amount of rainfall. The next steps after this analysis would be using the feature of gamma distribution while analyzing problems related to the average amount of rainfall