Tutorial 5 Solutions

STAT 3013/8027

- **Rice Chapter 8, Questions 13, 17 (a, b, c). See handwritten solutions.**
- **Question 25:**
  - A priori I believe the probability of landing up (θ)is between 0.10 and 0.40. A specific value would be 0.25.
  - While you may not have found a thumbtack to throw, there is a data set in R of 100 tosses! Use the first 20 observations (1 = "Up", 0 = "Down")

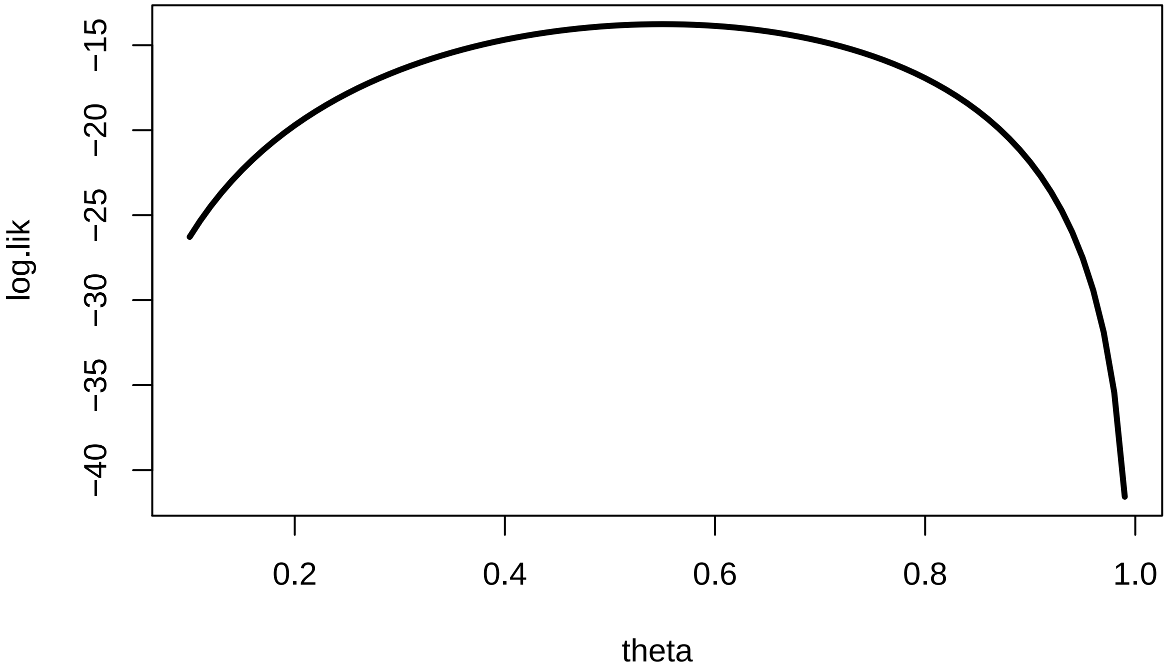
```
library("isdals")
data(thumbtack)

y <- thumbtack[1:20]
y
```

```
##  [1] 1 1 0 0 1 1 0 1 0 0 1 0 1 1 0 0 1 1 1 0
```

The likelihood of the data:

$$
\begin{aligned}
L(\theta|\boldsymbol{y}) &= \prod_{i=1}^{n=20} \theta^{y_i}(1-\theta)^{1-y_i} \\
&= \theta^{\sum_{i=1}^{20} y_i}(1-\theta)^{20-\sum_{i=1}^{20} y_i} \\
\\
\ell(\theta|\boldsymbol{y}) &= \left(\sum_{i=1}^{20} y_i\right) log(\theta) + \left(20 - \sum_{i=1}^{20} y_i\right) log(1-\theta)
\end{aligned}
$$

```
theta <- seq(0.1, 0.99, by=0.01)
log.lik <- NULL
for(i in 1:length(theta)){
 log.lik <- c(log.lik, sum(dbinom(y, 1, theta[i], log=TRUE)))
}

plot(theta, log.lik, type="l", lwd=3)
```



- Now let's run the experiment a bit differently . . . let's flip till we get 5 Ups:

```
check <- 0
c <- 1

while(check!=5){
 z <- thumbtack[21:(21+c)]
 check <- sum(z)
 c <- c+1
}

z
```

```
##  [1] 1 0 0 1 1 0 0 0 0 1 1
```

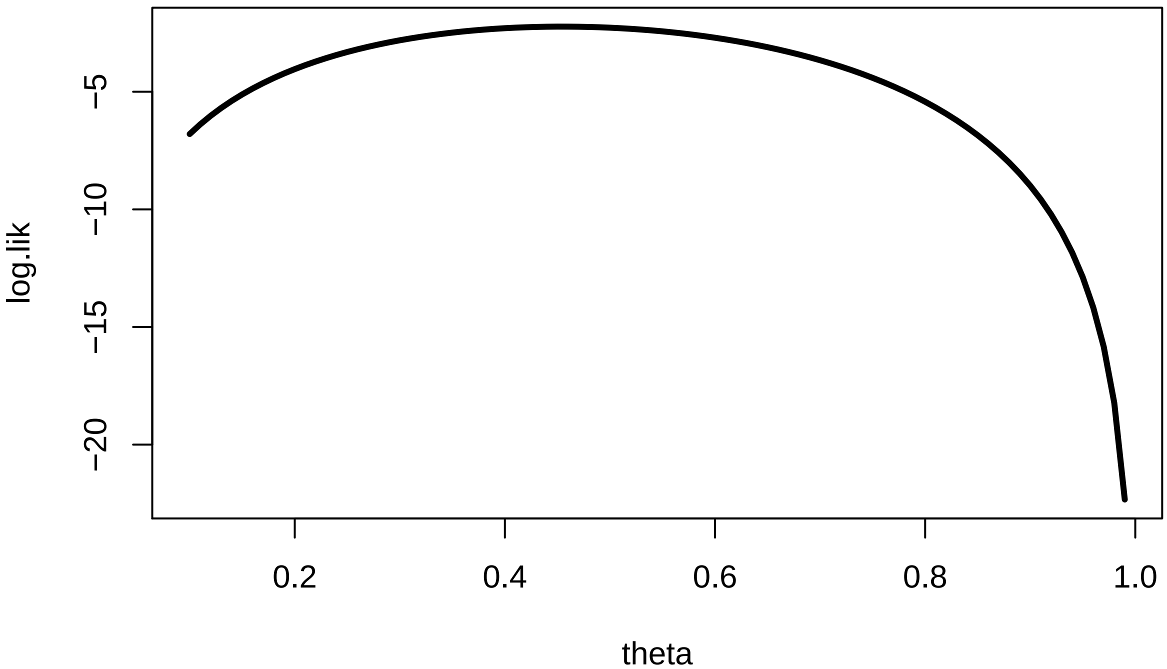
Based on the experiment, the likelihood is based on a negative binomial (flip until we get 5 = r successes):

$$
\begin{aligned}
L(\theta|\boldsymbol{z}) &= \binom{11-1}{5-1}\theta^5(1-\theta)^{14-5} \\
\ell(\theta|\boldsymbol{z}) &= log\binom{11-1}{5-1} + 5log(\theta) + (11-5)log(1-\theta)
\end{aligned}
$$

Notice that this is the same likelihood as above except for a constant in front, which won't change the maximization!

```
theta <- seq(0.1, 0.99, by=0.01)
log.lik <- NULL
for(i in 1:length(theta)){
```

```
  log.lik.i <- log(choose(11-1, 5-1)) + 5*log(theta[i]) + (11-5)*log(1-theta[i])
  log.lik <- c(log.lik, log.lik.i)
}

plot(theta, log.lik, type="l", lwd=3)
```



- Now let's determine the distribution for θ under a uniform prior for θ.

$$
\begin{aligned}
p(\theta|\boldsymbol{y}) &= \frac{p(\boldsymbol{y}|\theta)p(\theta)}{p(\boldsymbol{y})} \\
&\propto p(\boldsymbol{y}|\theta)p(\theta) \\
&= \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i} \times 1 \\
&= \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i}
\end{aligned}
$$

As θ is the random variable, we can see that this is a kernel for a beta(a, b) distribution:

$$
\theta^{\left(\sum_{i=1}^n y_i+1\right)-1}(1-\theta)^{\left(n-\sum_{i=1}^n y_i+1\right)-1}
$$

Where $a = (\sum_{i=1}^n y_i + 1)$ and $b = (n - \sum_{i=1}^n y_i + 1)$. Based on beta distribution the mean and variance are:

$$
\begin{aligned}
E[\theta|\boldsymbol{y}] &= \frac{\sum_{i=1}^n y_i + 1}{\sum_{i=1}^n y_i + 1 + n - \sum_{i=1}^n y_i + 1} \\
&= \frac{\sum_{i=1}^n y_i + 1}{n+2}
\end{aligned}
$$

```r
n <- length(y)
a <- sum(y)+1
b <- n - sum(y)+1

m <- a/(a+b)
v <- (a*b)/( (a+b)^2 * (a + b + 1))

m
```

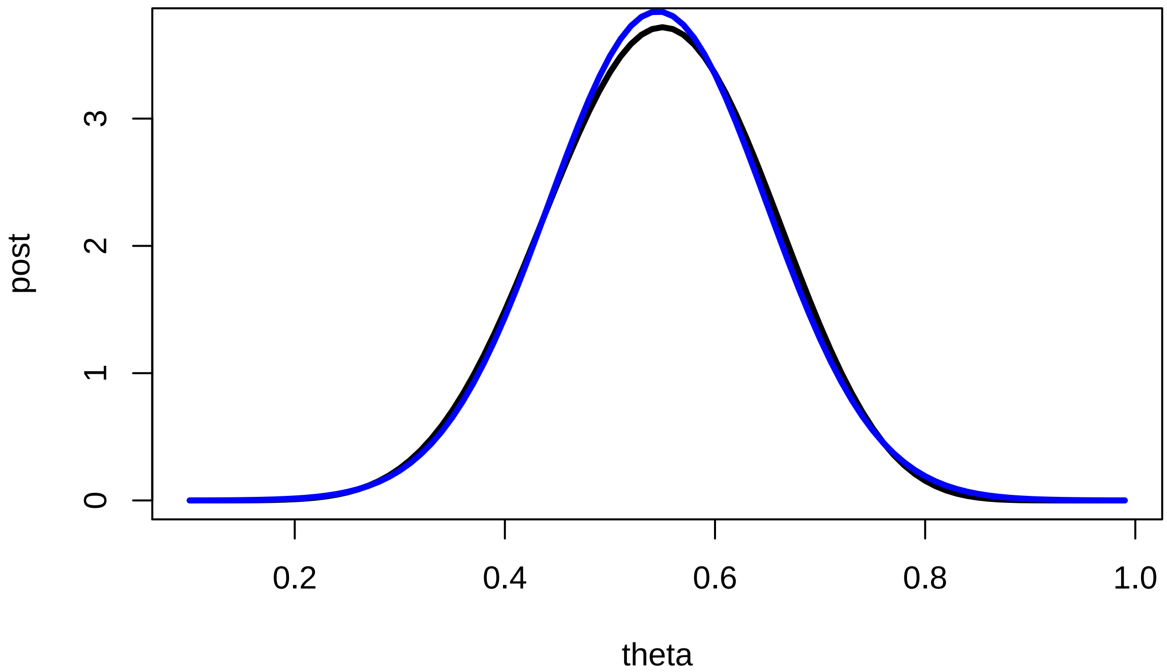
```
## [1] 0.5454545
```

```r
v
```

```
## [1] 0.01077973
```

- Let's plot the posterior based on observing $\boldsymbol{y}$ along with a normal approximation based on the mean and variance above:

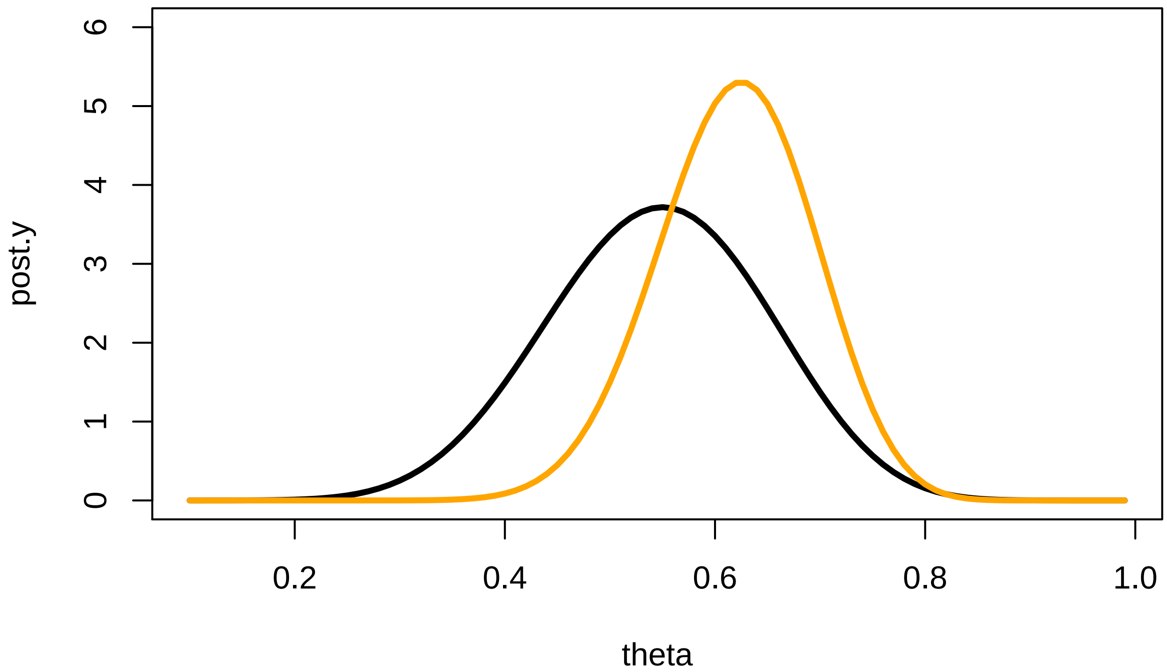
```r
theta <- seq(0.1, 0.99, by=0.01)
post <- NULL
norm.approx <- NULL

for(i in 1:length(theta)){
post.i <- dbeta(theta[i], sum(y)+1, 20-sum(y)+1)
post <- c(post, post.i)

norm.approx.i <- dnorm(theta[i], m, sqrt(v))
norm.approx <- c(norm.approx, norm.approx.i)
}

plot(theta, post, type="l", lwd=3)
lines(theta, norm.approx, lwd=3, col="blue")
```



The normal approximation (blue) is very similar to the posterior (black).

- Now lets throw the tack 20 more times (label these $\boldsymbol{x}$) and examine the two posteriors.

```
n.y <- length(y)

x <- thumbtack[40:59]
w <- c(y, x)
n.w <- length(w)


theta <- seq(0.1, 0.99, by=0.01)
post.y <- NULL
post.w <- NULL

for(i in 1:length(theta)){
post.y.i <- dbeta(theta[i], sum(y)+1, n.y-sum(y)+1)
post.y <- c(post.y, post.y.i)

post.w.i <- dbeta(theta[i], sum(w)+1, n.w-sum(w)+1)
post.w <- c(post.w, post.w.i)
    }



plot(theta, post.y, type="l", lwd=3, ylim=c(0, 6))
lines(theta, post.w, lwd=3, col="orange")
```

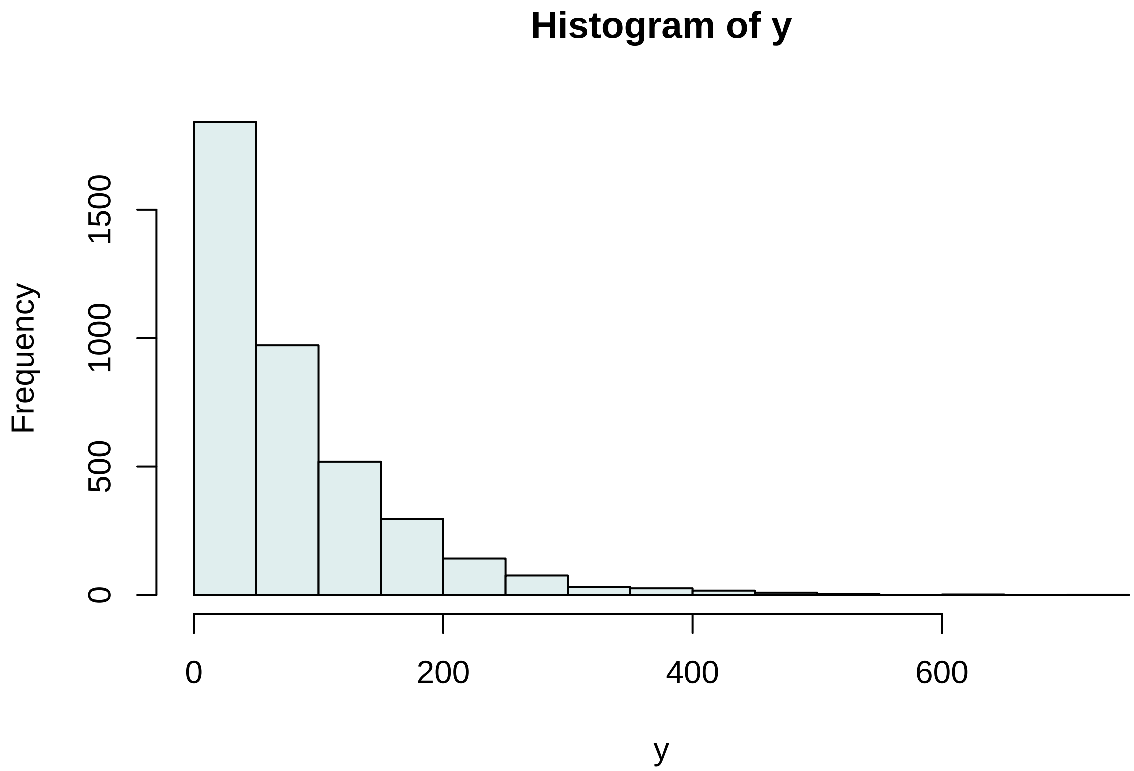


From the figure, with the full 40 data points, we see that are beliefs have changed (orange posterior). Also the variability is smaller.

- **A question for you all: What would the posterior look like if I used the first 20 data points (y) and clauclated a posterior. Now used the posterior as the prior for θ and observed the next 20 data points (x). What would that posterior look like?**
- **Question 43:**

a. Let's load in the data and examine a histogram of the data.

```
data <- read.table("gamma-arrivals.txt")
y <- data$V1
hist(y, col="azure2")
```



Based on the histogram, an gamma(a,b) distribution does not seem like an unreasonable model for the data.

$$f(y) = \frac{1}{\gamma(a)b^a} y^{a-1} exp(-y/b)$$

b. Let's first determine the **Method of Moments** for a and b. Our system of equations is:

$$\begin{aligned} E[Y] = ab &= \frac{1}{n}\sum_{i=1}^{n} y_i = m_1 \\ E[Y^2] = ab^2 + a^2b^2 &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 = m_2 \end{aligned}$$

Solving this system of equations we have:

$$\tilde{a} = \frac{m_1^2}{m_2 - m_1^2} \quad \tilde{b} = \frac{m_2 - m_1^2}{m_1}$$

```
n <- length(y)
m1 <- mean(y)
m2 <- sum(y^2)/n

a.mom <- m1^2/(m2-m1^2)
```

```r
    b.mom <- (m2-m1^2)/m1

    a.mom
```

```
## [1] 1.012352
```

```r
    b.mom
```

```
## [1] 78.95989
```

Now let's consider the **Maximum Likelihood** estimators.

$$L(a,b) = \prod_{i=1}^{n} \frac{1}{\Gamma(a)\ b^a} y_i^{a-1} exp\left(-y_i/b\right)$$

Here we have the log-likelihood:

$$\ell(a,b) = -n\ log\left(\Gamma(a)\right) - n\ a\ log(b) + (a-1)\sum_{i=1}^{n} log(y_i) - \sum_{i=1}^{n} y_i/b$$

$$\frac{\partial \ell(a,b)}{\partial a} = -n\psi(a) - nlog(b) + \sum_{i=1}^{n} log(y_i) \quad (1)$$
$$\frac{\partial^2 \ell(a,b)}{\partial a^2} = -n\psi'(a) \quad (2)$$

In Eqn (1), let's substitute in for $b = \sum_{i=1}^{n} y_i/(na)$. So we have an equation which only has a:

$$\frac{\partial \ell(a,b)}{\partial a} = -n\psi(a) - nlog\left(\sum_{i=1}^{n} y_i/(na)\right) + \sum_{i=1}^{n} log(y_i) \quad (3)$$
$$\frac{\partial^2 \ell(a,b)}{\partial a^2} = -n\psi'(a) \quad (4)$$

- Where $\psi(a) =$ `digamma`(a) and $\psi'(a) =$ `trigamma`(a).

```r
## Let's find the MLE of a using the N-R Approach.
## Then we can solve for b analytically
## Write some functions for U and H
U <- function(a){
  n <- length(y)
  out <-  -n* digamma(a) - n*log(sum(y)/(n*a)) + sum(log(y))
  return(out)
  }

H <- function(a){
    n <- length(y)
```

```r
    out <- -n*trigamma(a)
    return(out)
    }

## Starting values - use MoM estimator
a <- mean(y)^2/( (n-1)*var(y)/n)

## set a stopping point
eps <- 1e-07
check <- 10

## Save the results.
out <- a

## Run the algorithm
while(check > eps){
a.new <- a - U(a)/H(a)
check <- sum(abs(a-a.new))
a <- a.new
out <- rbind(out, t(a))
    }

a.mle <- a
b.mle <- sum(y)/(n*a.mle)

##
a.mle
```

```
## [1] 1.026332
```

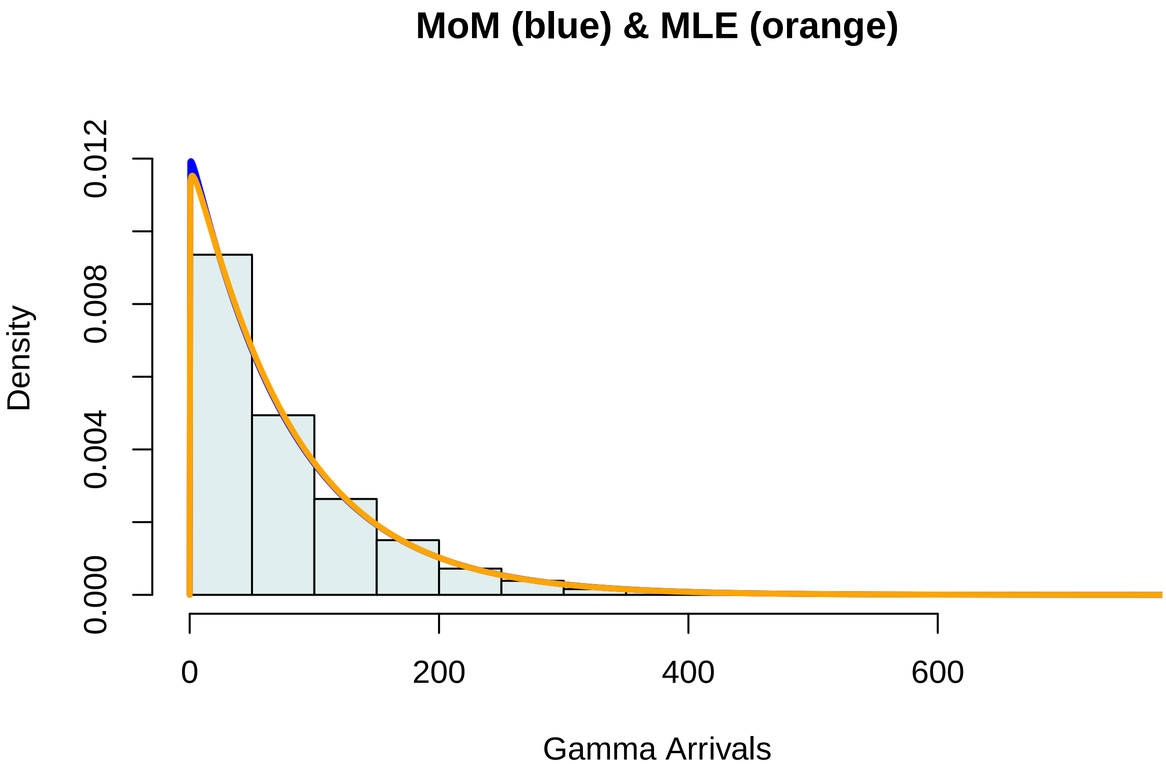
```r
b.mle
```

```
## [1] 77.8844
```

The ML and MoM estimates are very similar.

c. Let's plot the two fitted densities on top of the histogram.

```r
hist(y, col="azure2", freq=FALSE, ylim=c(0, 0.013), xlab="Gamma Arrivals", main="MoM (bl
x <- seq(0, 800, by=0.5)
lines(x, dgamma(x, shape=a.mom, scale=b.mom), lwd=3, col="blue")
lines(x, dgamma(x, shape=a.mle, scale=b.mle), lwd=3, col="orange")
```



We can see the results are very similar as the two lines essentiall overlap, except for a slight difference near 0.