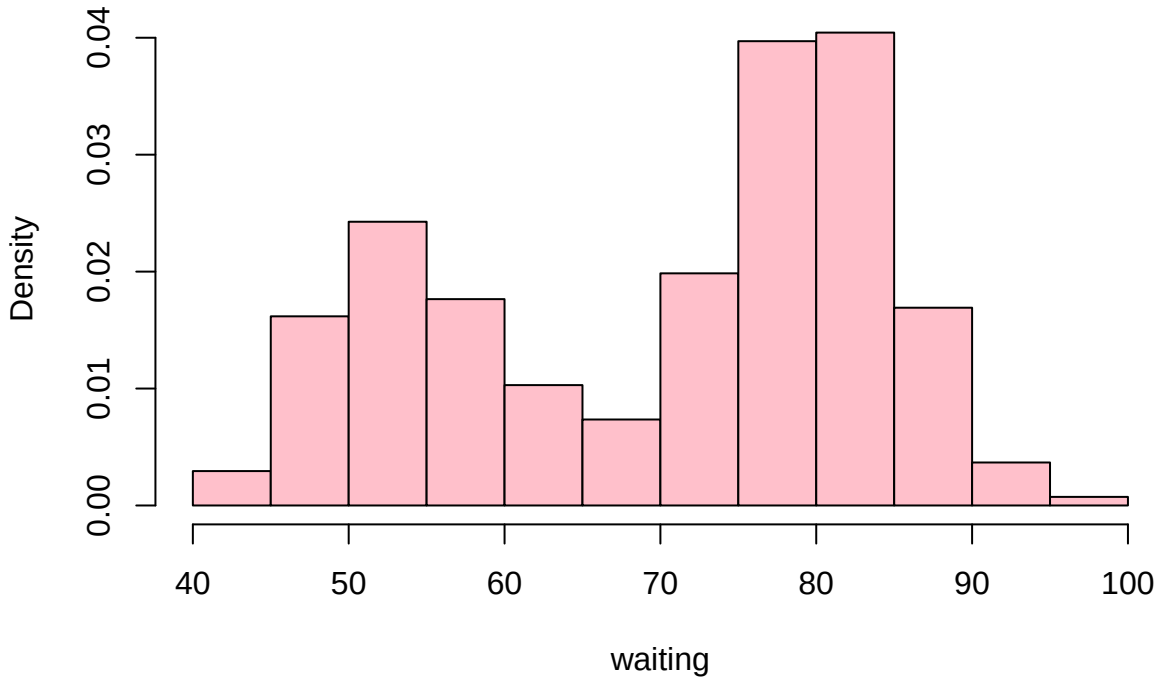
Question3

Varun Agrawal

Problem 3: Analysis of `faithful` datasets.

Consider the `faithful` datasets:

```
attach(faithful)
hist(faithful$waiting,xlab = 'waiting',probability = T,col='pink',main='')
```



Fit following three models using MLE method and calculate **Akaike information criterion** (aka., AIC) for each fitted model. Based on AIC decides which model is the best model? Based on the best model calculate the following probability

$$\mathbb{P}(60 < \texttt{waiting} < 70)$$

(i) **Model 1**:

$$f(x) = p * Gamma(x|\alpha, \sigma_1) + (1-p)N(x|\mu, \sigma_2^2), \quad 0 < p < 1$$

```r
attach(faithful)
data=faithful
waiting=sort(waiting)
Myopt1 <- function(x,p,alfa,sigma1,mu,sigma2){
  est <- function(x1 = x,para){
    est = (-sum(log(para[1]*dgamma(x,para[2],para[3]) + (1-para[1])*dnorm(x,para[4],para[5]))))
    return(est)
    }
  opt <- optim(par = c(p,alfa,sigma1,mu,sigma2),control=list(maxit=10000), fn = est,x1 = x)
  aic_1 <- 10 + 2*est(x1 = x,para = opt$par)
  print(paste("AIC for model 1 = ",aic_1))
  return(c(opt$par,aic_1))
}
p = length(waiting[waiting<70])/length(waiting)
MyFit <- Myopt1(waiting,p,55,1,85,1)
```

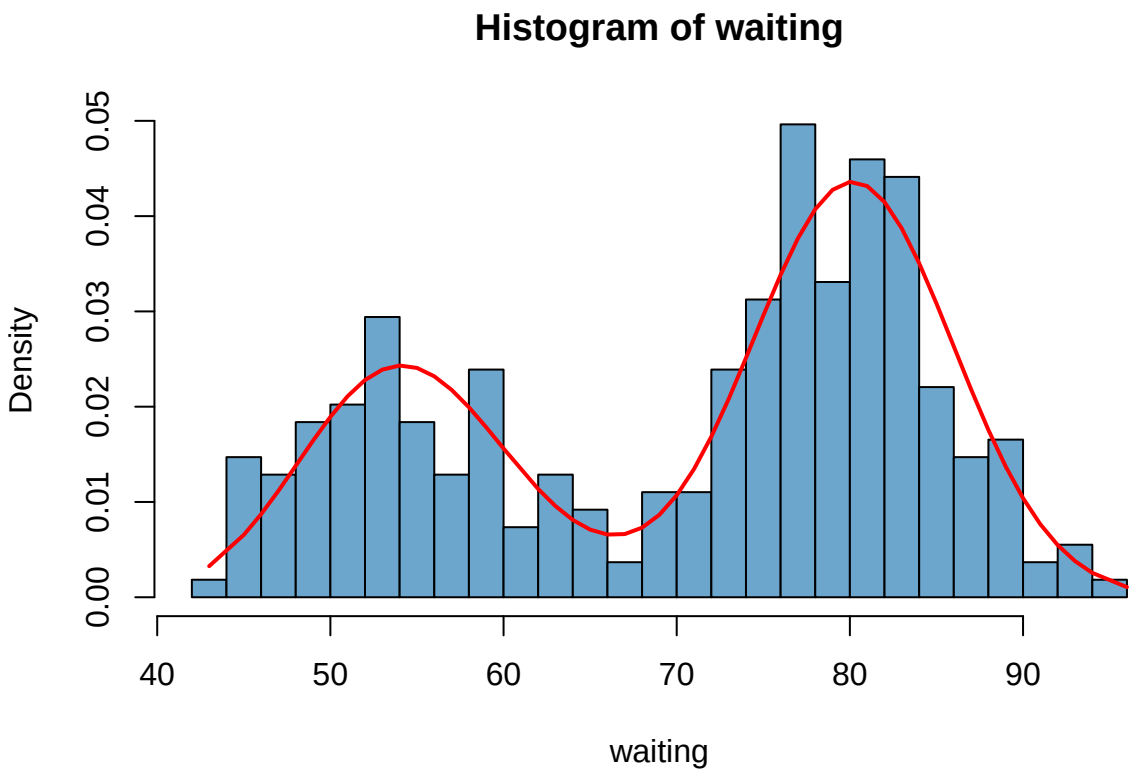
```
## [1] "AIC for model 1 =  2076.18037995149"
```

```r
AIC_1 <- Myopt1(waiting,p,55,1,85,1)[6]
```

```
## [1] "AIC for model 1 =  2076.18037995149"
```

```r
#Estimators
phat <- MyFit[1]
alfahat <- MyFit[2]
sigma1hat <- MyFit[3]
muhat <- MyFit[4]
sigma2hat <- MyFit[5]
dplot=(phat*dgamma(waiting,alfahat,sigma1hat))+((1-phat)*dnorm(waiting,muhat,sigma2hat))

hist(waiting,probability = T,breaks = 25,col = 'skyblue3')
lines(waiting,dplot,lwd=2,col="red")
```



(ii) **Model 2**:

$$f(x) = p * Gamma(x|\alpha_1, \sigma_1) + (1-p)Gamma(x|\alpha_2, \sigma_2), \quad 0 < p < 1$$

```
attach(faithful)
data=faithful
waiting=sort(waiting)
Myopt2 <- function(x,p,alfa,sigma1,mu,sigma2){
  est <- function(x1 = x,para){
    est = (-sum(log(para[1]*dgamma(x,para[2],para[3]) + (1-para[1])*dgamma(x,para[4],para[5]))))
    return(est)
    }
  opt <- optim(par = c(p,alfa,sigma1,mu,sigma2),control=list(maxit=10000), fn = est,x1 = x)
  aic_2 <- 10 + 2*est(x1 = x,para = opt$par)
  print(paste("AIC for model 1 = ",aic_2))
  return(c(opt$par,aic_2))
}
p = length(waiting[waiting<70])/length(waiting)
MyFit2 <- Myopt2(waiting,p,55,1,85,1)
```

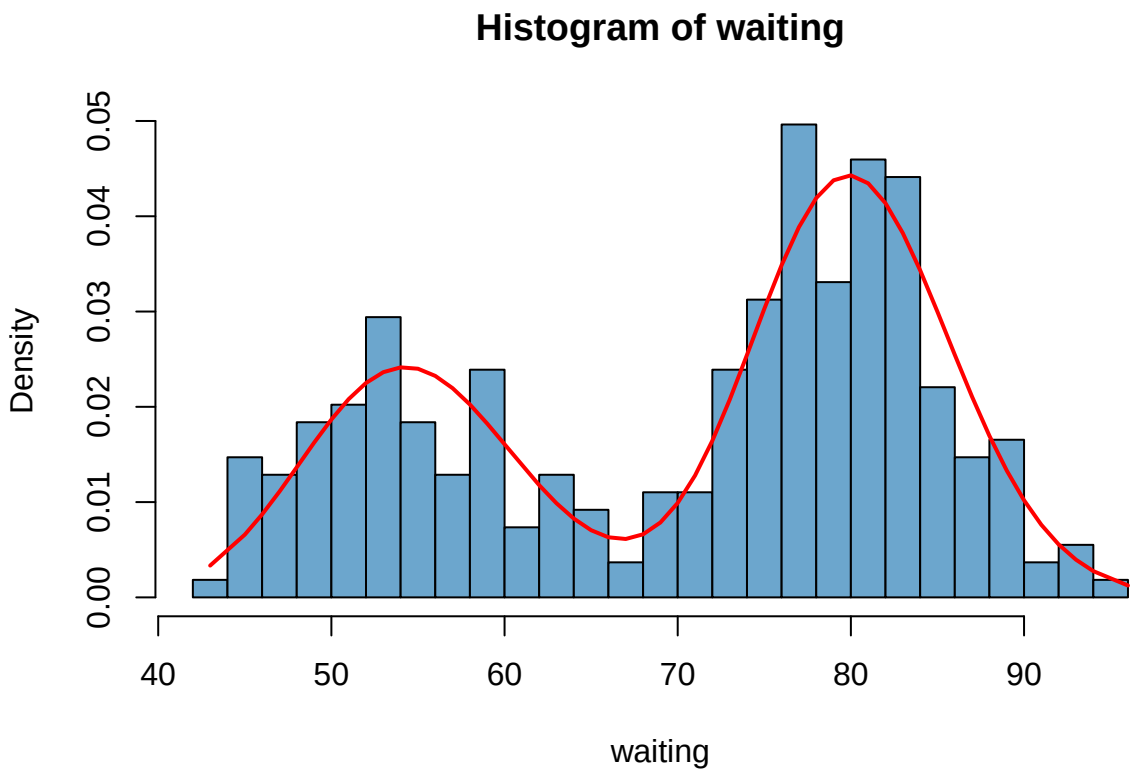
```
## [1] "AIC for model 1 =  2076.11652389509"
```

```
AIC_2 <- Myopt2(waiting,p,55,1,85,1)[6]
```

```
## [1] "AIC for model 1 =  2076.11652389509"
```

```r
#Estimators
phat2 <- MyFit2[1]
alfahat2 <- MyFit2[2]
sigma1hat2 <- MyFit2[3]
muhat2 <- MyFit2[4]
sigma2hat2 <- MyFit2[5]
dplot2=(phat2*dgamma(waiting,alfahat2,sigma1hat2))+((1-phat2)*dgamma(waiting,muhat2,sigma2hat2))

hist(waiting,probability = T,breaks = 25,col = 'skyblue3')
lines(waiting,dplot2,lwd=2,col="red")
```



(iii) **Model 3**:

$$f(x) = p * logNormal(x|\mu_1, \sigma_1^2) + (1-p) logNormal(x|\mu_1, \sigma_1^2), \quad 0 < p < 1$$

```r
attach(faithful)
data=faithful
waiting=sort(waiting)
Myopt3 <- function(x,p,alfa,sigma1,mu,sigma2){
  est <- function(x1 = x,para){
    est = (-sum(log(para[1]*dlnorm(x,para[2],para[3]) + (1-para[1])*dlnorm(x,para[4],para[5]))))
    return(est)
    }
  opt <- optim(par = c(p,alfa,sigma1,mu,sigma2),control=list(maxit=10000), fn = est,x1 = x)
  aic_3 <- 10 + 2*est(x1 = x,para = opt$par)
  print(paste("AIC for model 1 = ",aic_3))
```

```
  return(c(opt$par,aic_3))
}
p = length(waiting[waiting<70])/length(waiting)
MyFit3 <- Myopt3(waiting,p,3.9,0.015,4.4,0.012)
```

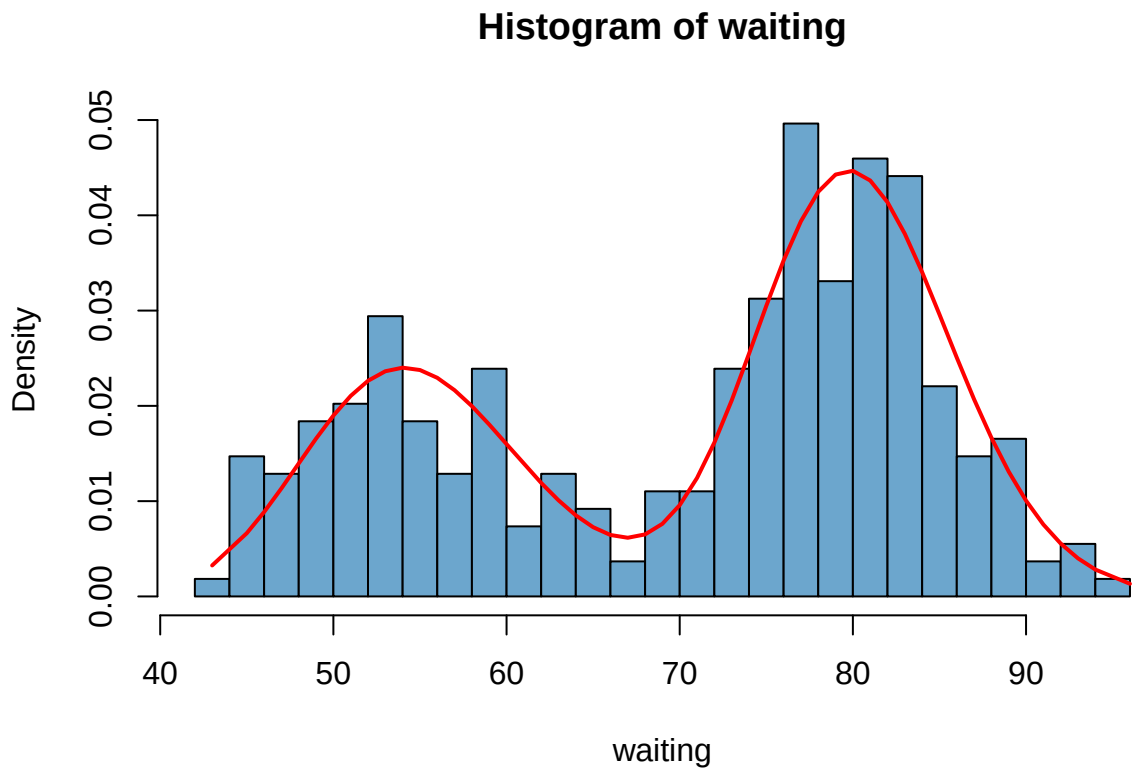
```
## [1] "AIC for model 1 =  2075.42004355982"
```

```
AIC_3 <- Myopt3(waiting,p,3.9,0.015,4.4,0.012)[6]
```

```
## [1] "AIC for model 1 =  2075.42004355982"
```

```
#Estimators
phat3 <- MyFit3[1]
alfahat3 <- MyFit3[2]
sigma1hat3 <- MyFit3[3]
muhat3 <- MyFit3[4]
sigma2hat3 <- MyFit3[5]
dplot3=(phat3*dlnorm(waiting,alfahat3,sigma1hat3))+((1-phat3)*dlnorm(waiting,muhat3,sigma2hat3))

hist(waiting,probability = T,breaks = 25,col = 'skyblue3')
lines(waiting,dplot3,lwd=2,col="red")
```



```r
listAIC <- c('AIC1' = AIC_1,'AIC2' = AIC_2,'AIC3' = AIC_3)
minAIC <- which.min(listAIC)
minAIC
```

```
## AIC3
##    3
```

As we can see AIC for 3^{rd} model is lowest So we will use it for getting our required probability,

$$\mathbb{P}(60 < \texttt{waiting} < 70)$$

```r
reqd_prob <- (phat3*plnorm(70, meanlog = alfahat3, sdlog = sigma1hat3) + (1-phat3)*plnorm(70, meanlog =
print(paste('Require Prob is = ', reqd_prob))
```

```
## [1] "Require Prob is =  0.0908335053749842"
```