PROBLEM 3

Tyagi Mahir

2022-11-16

```
attach(faithful)
require(tidyverse)
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

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

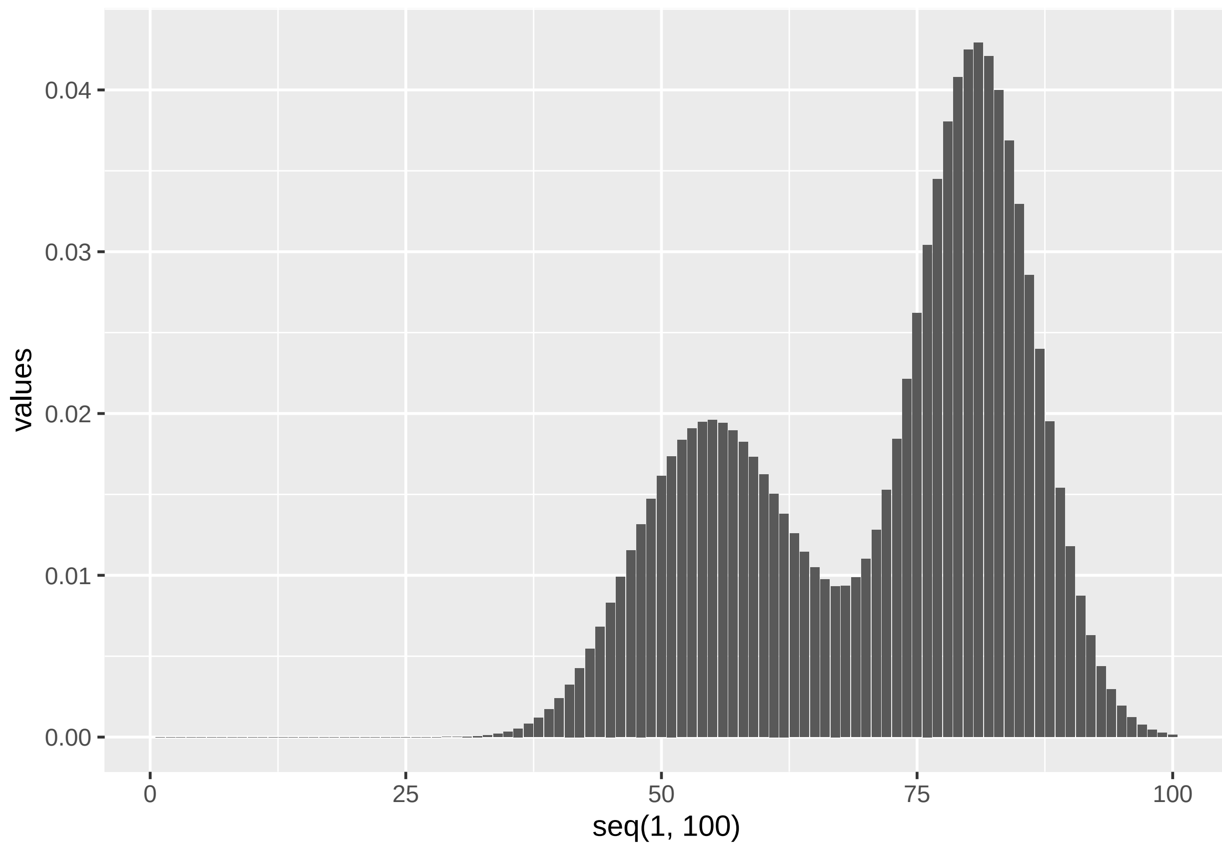
$$\mathbb{P}(60 < \texttt{waiting} < 70)$$

MODEL ONE

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

```
#Code begins here
x=faithful$waiting
llone= function(x,theta){
  p=theta[1]
  if (p<0.2|p>0.8) return(1000000000)
  alpha=theta[2]
  sigma1=theta[3]
  mu=theta[4]
  sigma2=theta[5]
  return((-1)*(sum(log(p*dgamma(x,alpha,sigma1)+(1-p)*dnorm(x,mu,sigma2)))))
}
#Optimizing
opt1=optim(fn=llone,par=c(0.5,50,2,80,2),x=x)
#Visualizing model
modelone=function(x,theta){
  p=theta[1]
  if (p<0|p>1) return(1000000000)
```

```
  alpha=theta[2]
  sigma1=theta[3]
  mu=theta[4]
  sigma2=theta[5]
  return(p*dgamma(x,alpha,sigma1)+(1-p)*dnorm(x,mu,sigma2))
}
modeloneplot=data.frame(values=modelone(seq(1,100),opt1$par))
ggplot(modeloneplot,aes(x=seq(1,100),y=values))+geom_col()
```



Here, the AIC value is:

```
k=5 #5 parameters estimated
aicone=2*k+2*opt1$value
aicone
```

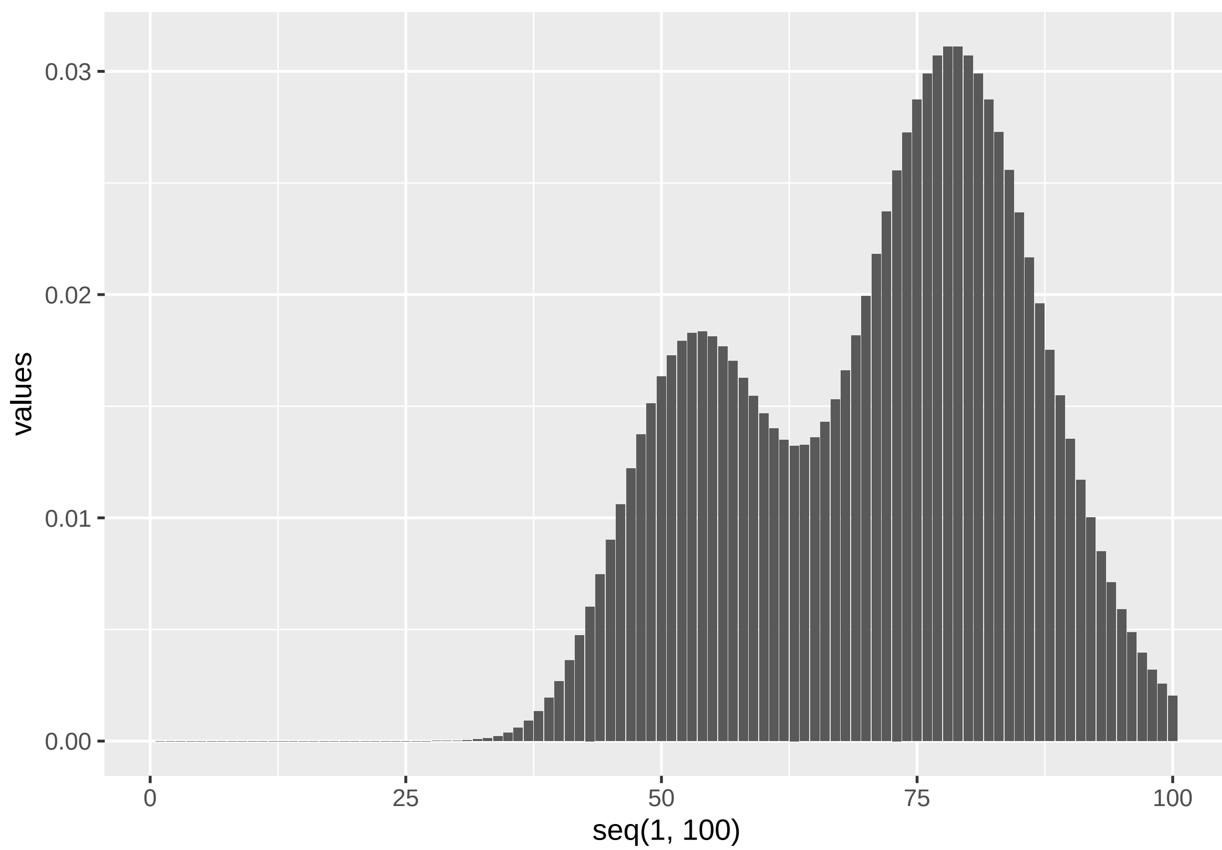
```
## [1] 2084.3
```

MODEL TWO

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

```
lltwo= function(x,theta){
  p=theta[1]
  if (p<0.2|p>0.8) return(1000000000)
  alpha1=theta[2]
```

```
  sigma1=theta[3]
  alpha2=theta[4]
  sigma2=theta[5]
  return((-1)*(sum(log(p*dgamma(x,alpha1,sigma1)+(1-p)*dgamma(x,alpha2,sigma2)))))
}
opt2=optim(fn=lltwo,par=c(0.5,50,2,80,2),x=x)
#Visualising model
modeltwo = function(x,theta){
  p=theta[1]
  alpha1=theta[2]
  sigma1=theta[3]
  alpha2=theta[4]
  sigma2=theta[5]
  return(p*dgamma(x,alpha1,sigma1)+(1-p)*dgamma(x,alpha2,sigma2))
}
modeltwoplot=data.frame(values=modeltwo(seq(1,100),opt2$par))
ggplot(modeltwoplot,aes(x=seq(1,100),y=values))+geom_col()
```



AIC value for model two is:

```
k=5 #5 parameters estimated
aictwo=2*k+2*opt2$value
aictwo
```

```
## [1] 2116.73
```

MODEL THREE

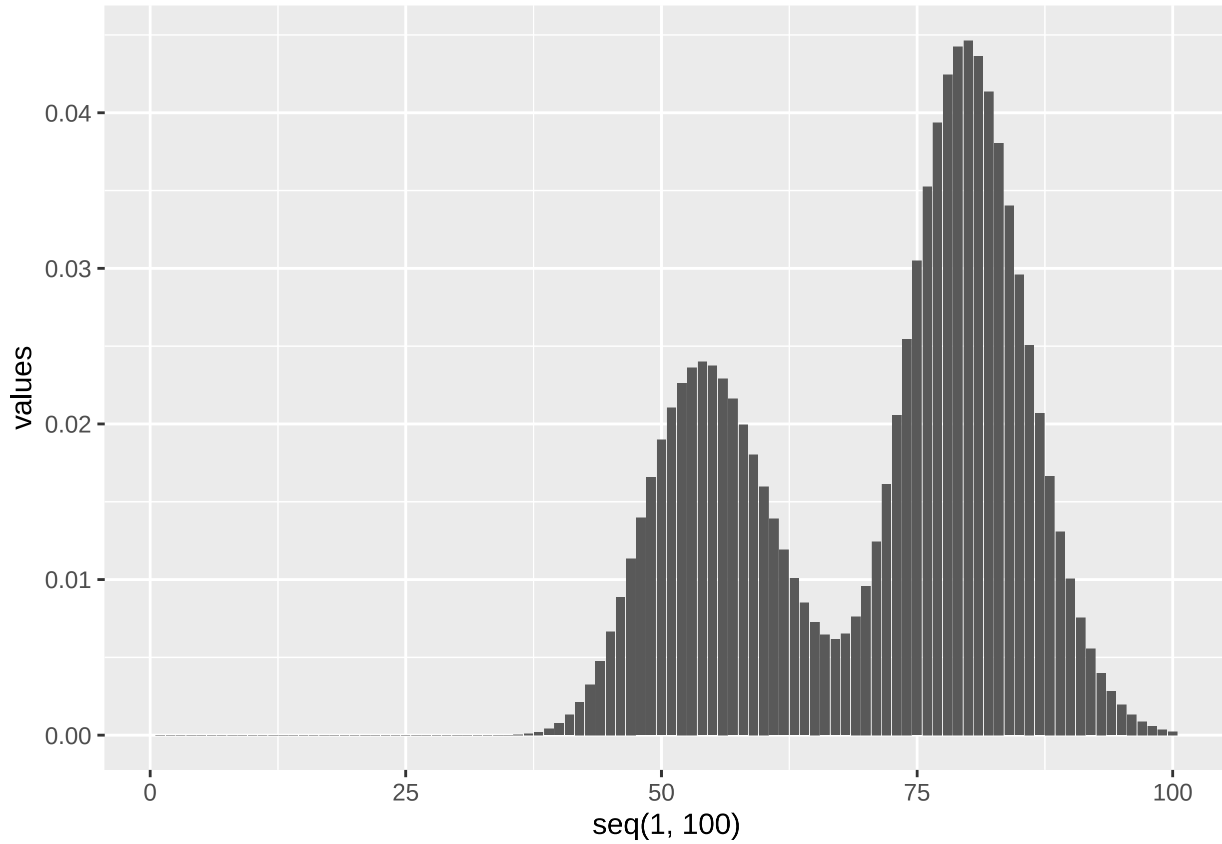
$$f(x) = p * logNormal(x|\mu_1, \sigma_1^2) + (1-p)logNormal(x|\mu_1, \sigma_1^2), \quad 0 < p < 1$$

```
llthree= function(x,theta){
  p=theta[1]
  if (p<0.2|p>0.8) return(1000000000)
  mu1=theta[2]
  sigma1=theta[3]
  mu2=theta[4]
  sigma2=theta[5]
  return((-1)*(sum(log(p*dlnorm(x,mu1,sigma1)+(1-p)*dlnorm(x,mu2,sigma2)))))
}
opt3=optim(fn=llthree,par=c(0.5,log(50),log(2),log(80),log(2)),x=x)
```

```
## Warning in dlnorm(x, mu1, sigma1): NaNs produced

## Warning in dlnorm(x, mu2, sigma2): NaNs produced
```

```
#Visualizing model
modelthree = function(x,theta){
  p=theta[1]
  mu1=theta[2]
  sigma1=theta[3]
  mu2=theta[4]
  sigma2=theta[5]
  return(p*dlnorm(x,mu1,sigma1)+(1-p)*dlnorm(x,mu2,sigma2))
}
modelthreeplot=data.frame(values=modelthree(seq(1,100),opt3$par))
ggplot(modelthreeplot,aes(x=seq(1,100),y=values))+geom_col()
```



And finally the AIC for model three is:

```
k=5 #5 parameters estimated
aicthree=2*k+2*opt3$value
aicthree
```

```
## [1] 2075.42
```

Hence we decide that Model three is the best, as its aic value is the least.

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

```
p=opt3$par[1]
mu1=opt3$par[2]
sigma1=opt3$par[3]
mu2=opt3$par[4]
sigma2=opt3$par[5]
(p*plnorm(70,mu1,sigma1)+(1-p)*plnorm(70,mu2,sigma2))-
  (p*plnorm(60,mu1,sigma1)+(1-p)*plnorm(60,mu2,sigma2))
```

```
## [1] 0.0908389
```