STAT 805 Computational Mathematics and Statistics

Submission By:

Name: Afiq Abdul Hamid

ID: 18015518

Email: zrj7023@autuni.ac.nz

Assignment 2

**Question 1** Given null Hypothesis = 3.1 and Obs = 7, for one sided tests

**Question 1 a):** State the precise value of p-value for observation = 7

P value is the Probability that given the Null is true (Ho: λ = 3.1) of seeing the Observation or an Observation more extreme. If this value is less than the significance 0.05 then we reject the Ho. If it is Greater, then we fail to reject Ho

The Precise Definition of P value: P ( X = 7| λ = 3)

**Question 1 b):** Calculate the p-value for this observation and interpret

Cat(“p value is: ”, ppois(7,3.1, lower.tail=FALSE))

[1] p value is: 0.01421251

This means that for one sided poisson distribution Obs x = 7 for lambda = 3.1 is significant. We reject the Null Hypothesis.

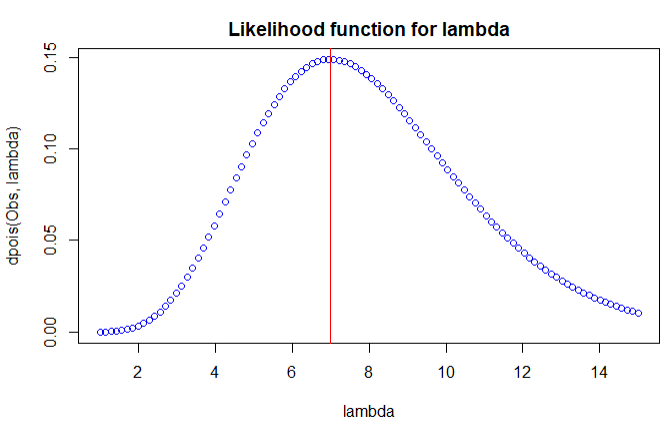
**Question 1 c):** For X = 7 plot a likelihood function of λ. Choose a sensible range

Obs <- 7

lambda <- seq(from=1,to=15,len=100)

plot(lambda,dpois(Obs,lambda),main='Likelihood function for lambda', col = 'blue')

abline(v=7, col = "red" )



Sensible value for lambda is between 2 and 14

**Question 1 d):** For Observation of X = 7 plot a log likelihood function for λ. Estimate a credible interval for λ. Rougly.

supp <- dpois(Obs, lambda, log = TRUE)

plot(lambda, supp-max(supp), main='Log Likelihood function of Lambda')

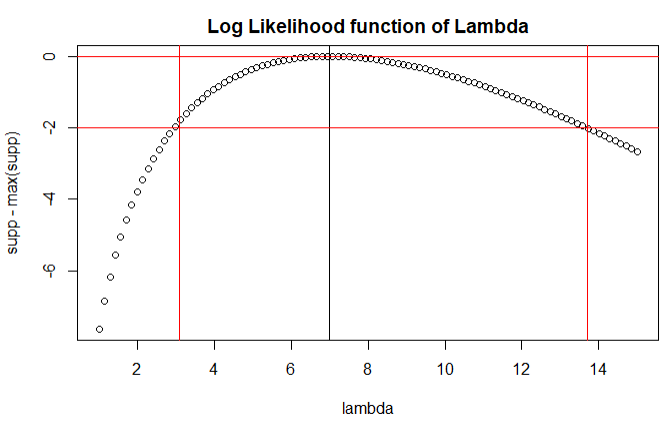
abline(v=Obs)

abline(v=3.1, col = "red")

abline(v=13.7, col = "red")

abline(h = 0, col = "red")

abline(h = -2, col = "red")



Credible Interval of Lambda is between 3.1 and 13.8 for 2 Units of Support which is sufficient for this case.

**Question 1 e):** A Bayesian comes with a prior λ as exponential with rate 0.6. Plot Posterior likelihood function with observation 7. Density of the exponential distribution is dexp(x, rate=0.1)

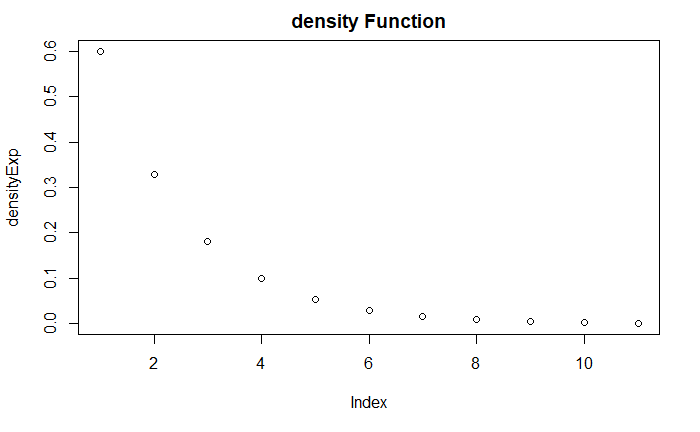
#First plot density

rangeX <- 0:10

rate <- 0.6

densityExp <- dexp(rangeX, rate)

plot(densityExp, main = "density Function")



lambdaSeq2 <- seq(from = 0, to = 16, len = 100)

#multiply prior to get pdf

u <- exp(-1.6\*lambdaSeq2)\*((lambdaSeq2^7)/(factorial(7)))

plot(lambdaSeq2, u ,

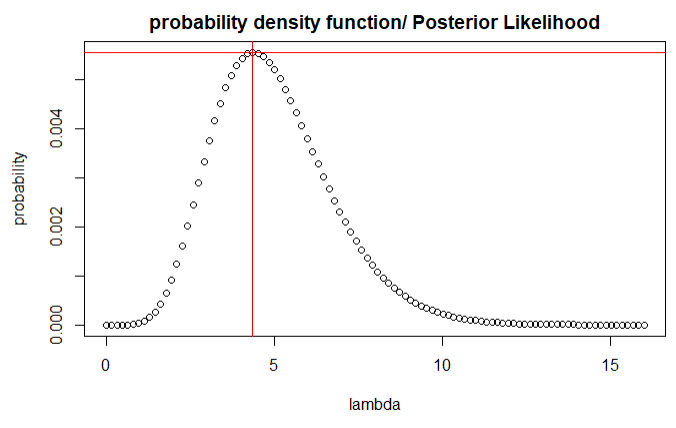
xlab = "lambda",

ylab = "p",

main = "probability density function/ Posterior Likelihood")

abline(h=max(u), col = “red”)

abline(v=4.35, col = “red”)



**Question 2** Given a Beta distribution of a = 1.3 b = 1.5 performed 5 times with 3 success and 2 failure

**Question 2 a):** Plot the prior and posterior PDFs.

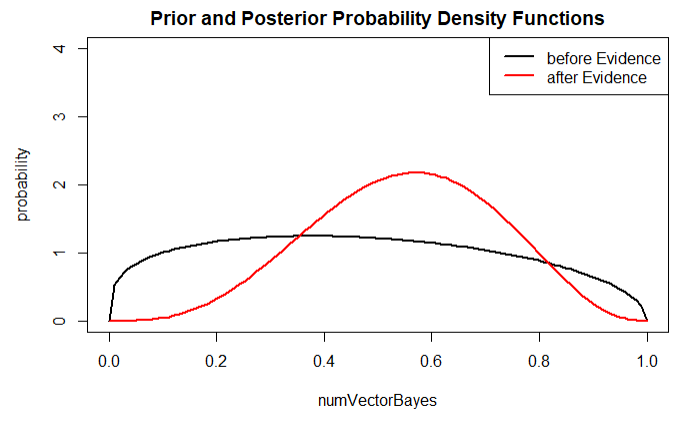
numVectorBayes <- seq(from = 0 , 1, len = 100)

plot(numVectorBayes, numVectorBayes\*0 , ylim=c(0,4),ylab="probability", main="Prior and Posterior Probability Density Functions", type="n")

points(numVectorBayes,dbeta(numVectorBayes,1.3,1.5),lwd=2,type="l",col="black")

points(numVectorBayes,dbeta(numVectorBayes,(1.3+3),(1.5+2)),lwd=2,type="l",col="red")

legend("topright", lwd = 2,col=c("black","red"), legend=c("before Evidence", "after Evidence"))



**Question 2 b):** Calculate the maximum likelihood estimate for p, and give the prior mode

aprior <- 1.3

bprior <- 1.5

apost <- aprior + 3

bpost <- bprior + 2

pbeta(1/2 , apost ,bpost)

priodMode <- (aprior-1)/(aprior + bprior -2)

priodMode

[1] 0.3810817 #MLE for P

[1] 0.375 #mode of Beta Distribution

**Question 2 c):** calculate posterior mean

postMean = apost/(apost + bpost)

postMean

[1] 0.5512821

**Question 2 d):** provide a Bayesian credible interval: Two values pL and pU such that pPost = 0.95

x<-seq(from=0,to=1,by=0.01) #set a sequence from 0 to 1

y<-dbeta(x,apost, bpost) #using dbeta to calculate density with alpha = 4.3 and beta =

plot(x,y, main = "Beta Distribution") #draw the plot

significance <- 0.05

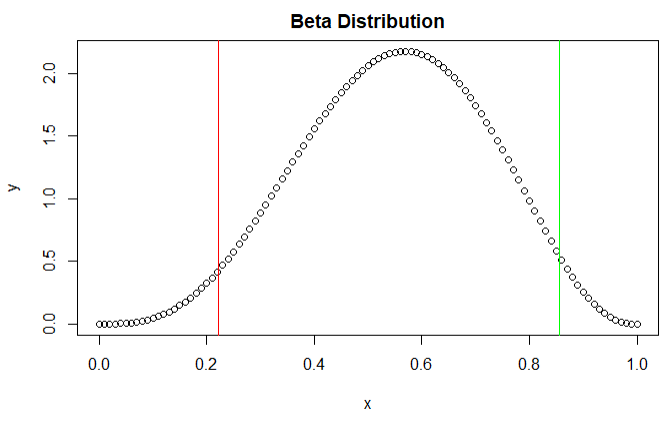
rightSide <- qbeta( 1 - significance/2, apost, bpost)

leftSide<- qbeta(significance/2, apost, bpost)

abline(v = rightSide , col = "green")

abline(v = leftSide , col = "red")

cat(leftSide, " < x < ", rightSide)



Credible Bayesian Interval: 0.2215257 < x < 0.8563363

Pl = 0.2215257

Pu = 0.8563363

**Question 3:** Faculty Office with 11 staff, 5 professors, 6 non professors, 7 offices with windows, 6 offices no windows

**Question 3 a):** State a sensible null hypothesis

Null Hypoyhesis (Ho): There is no difference in the probability of getting an office with a window regardless if you are professor or non-professor.

Getting an office with a window is independent of being a professor.

**Question 3 b):** Is it a one-sided test or 2 sided test?

One sided because the probability of professors not getting an office with a window may only depart in the direction of a non-professor getting an office with a window and that reason alone, but otherwise the professors tend to have their way.

There is an asymmetric leaning towards the professors getting an office (one sided)

**Question 3 c):** use fiser.test(x) and interpret result.

a<-matrix(c(5,0,2,4),2,2)

dimnames(a)<-list(office=c("window", "no window"), professor=c("yes","no"))

print(a)

fisher.test(a, alternative = "greater")

professor

office yes no

window 5 2

no window 0 4

Fisher's Exact Test for Count Data

data: a

p-value = 0.04545

alternative hypothesis: true odds ratio is greater than 1

95 percent confidence interval:

1.054377 Inf

sample estimates:

odds ratio

Inf

Since the exact P value is less than 0.05 that is enough evidence for us to reject our null hypothesis. That means that Professor are more likely to get an office with a window. There is a significant association between Professors that are staff and those professors getting assigned an office with a window. Getting an office with a window is dependent on being a professor.

**Question 3 d):** Give an example of a two-by-two contingency table that you encounter in your personal day-to-day life. State what your observation is, what your null is and what it means, and

carry out a Fisher's exact test. State whether a one-sided or two-sided test is used, and

justify this. Interpret your findings in non-statistical language.

I went around and surveyed Students from different Faculties on What Type of Laptops (macbook/ Windows) they used. I want to see whether students in STEM field (Science, Technology, Engineering, Mathematics) prefer Macbook over windows compared to students in non-STEM (business/ art/ design) field. Looking for P (Mac | STEM)

Here is my data

|  |  |  |
| --- | --- | --- |
| Name | Field of Study | Laptop |
| Afiq | STEM | Win |
| Darion | STEM | Win |
| Alfred | Non STEM | Win |
| Jake | Non STEM | Mac |
| Colleen | STEM | Mac |
| Jon Snow | Non STEM | Win |
| Sally | STEM | Win |
| Marius | STEM | Mac |
| Lucy | Non STEM | Mac |
| Adriana | STEM | Mac |
| Daenarys | Non STEM | Win |
| Tyrion | STEM | Mac |
| Jaime | STEM | Win |
| Lyon | Non STEM | Mac |
| Aegon | Non STEM | Mac |
| Margret | STEM | Win |
| Desmond | Non STEM | Win |

17 people interview

9 STEM

8 Non STEM

8 Mac

9 Win

Null Hypeothesis:

Ho: There is no preference for STEM students to use Mac or Windows laptop. Students taking STEM field are not more likely to own a macbook.

Make contingency table

|  |  |  |
| --- | --- | --- |
|  | STEM | Non STEM |
| Mac | 4 | 4 |
| Win | 5 | 4 |

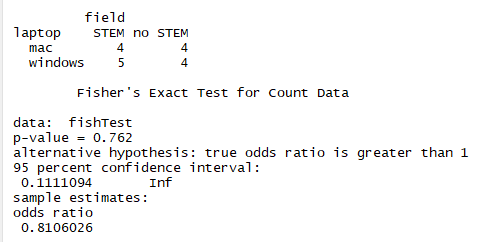
We put in R:

fishTest<-matrix(c(4,5,4,4),2,2)

dimnames(fishTest)<-list(laptop=c("mac", "windows"), field=c("STEM","no STEM"))

print(fishTest)

fisher.test(fishTest, alternative = "greater")



Based of the fishers test there is no significance. We fail to reject the null.

The test is a two-sided test because there is no asymmetrical leaning in STEM field. You could own either a mac or windows machine. There is no leaning towards one side.

Findings elaborated:

In the age-old struggle of windows vs mac there is no dependence on whether that student is taking STEM Subjects or not Non-STEM subject that determines their choice of laptop. No correlation.

Code assignment 2 can be found on Repo: <https://github.com/coderXmachina2/computationalMathematics/tree/master/ass2>