MA 471: Lab Assignment 01

Due on Monday, August 8, 2017

Sai Teja Talasila
140123040

Problem 1

R Code :

```
library(MASS)
library(stats)
options(warn=-1)
A <- read.table("d-csp0108.txt", header=TRUE)
n = nrow(A)

dexp <- function(x,mu,b)
{
    return((1/(2*b))*exp( - (abs(x-mu)/b)) )
}

pdexp <- function(x,mu,b)
{
    z = (x-mu)/b
    return(ifelse(x<mu,0.5*exp(z),1-0.5*exp(-z)))
}

qdexp <- function(x,mu,b)
{
    return(mu+ifelse(x<0.5,b*log(2*x),-b*log(2-2*x)))
}

dmixnormal <- function(x,m1,s1,m2,s2,p)
{
  return(p*dnorm(x,m1,s1)+(1-p)*dnorm(x,m2,s2))
}

dmymixestimate <- function(x,X)
{
  munorm = mean(X)
  std = sd(X)
  return(dmixnormal(x,munorm,std,munorm,std/3,0.4))
}

pmixnormal <- function(x,m1,s1,m2,s2,p)
{
  return(p*pnorm(x,m1,s1)+(1-p)*pnorm(x,m2,s2))
}

pmymixestimate <- function(x,X)
{
  munorm = mean(X)
  std = sd(X)
  return(pmixnormal(x,munorm,std,munorm,std/3,0.4))
}

rmixnormal <- function(n,m1,s1,m2,s2,p)
```

```
{
  x = vector(,n)
  for(i in 1:n)
  {
    u = runif(1,0,1)
    if(u<p)
      x[i] = rnorm(1,m1,s1)
    else
      x[i] = rnorm(1,m2,s2)
  }
  return(x)
}

qmixnormal <- function(x,m1,s1,m2,s2,p)
{
  X = rmixnormal(10000,m1,s1,m2,s2,p)
  return(quantile(X,x))
}

qmymixestimate <- function(x,X)
{
  munorm = mean(X)
  std = sd(X)
  return(qmixnormal(x,munorm,std,munorm,std/3,0.4))
}

fit_dist <- function(X)
{
    hist(X,probability = T,100,main = "Density fits",xlim = c(quantile(X,0.01),
        quantile(X,0.99)))
    munorm = mean(X)
    std = sd(X)
    curve(dnorm(x,munorm,std),add = TRUE,col=1)

    a = fitdistr(X, "t", start = list(m=mean(X),s=sd(X), df=2), lower=c(-1, 0.001,1))
        [1]
    m = a$estimate[1]
    s = a$estimate[2]
    df = a$estimate[3]
    mydt <- function(x, m, s, df) dt((x-m)/s, df)/s
    curve(mydt(x,m,s,df),add=T,col=2)

    muexp = median(X)
    N = length(X)
    b = sum(abs(X-muexp))/N
    curve(dexp(x,muexp,b),add=T,col=3)

    a = fitdistr(X,"cauchy")
    mucauchy = a$estimate[1]
    gamma = a$estimate[2]
    curve(dcauchy(x,location = mucauchy,scale = gamma),add=T,col=4)

    curve(dmymixestimate(x,X),add=T,col=5)
```

```
    legend('topright', legend = c("Normal","t","double exp","cauchy","mixed normal"),
        lty=1, col=c(1,2,3,4,5), bty='n', cex=.75)
}

jpeg("scatter_C.jpeg")
scatter.smooth(A[,2])
dev.off()

jpeg("scatter_SP.jpeg")
scatter.smooth(A[,3])
dev.off()

jpeg("C.jpeg")
fit_dist(A[,2])
dev.off()

jpeg("SP.jpeg")
fit_dist(A[,3])
dev.off()

survival <- function(X,f = 0,t = 1)
{
    scdf = ecdf(X)
    surv <- function(x) return(1-scdf(x))

    munorm = mean(X)
    std = sd(X)

    a = fitdistr(X, "t", start = list(m=mean(X),s=sd(X), df=2), lower=c(-1, 0.001,1))
        [1]
    m = a$estimate[1]
    s = a$estimate[2]
    df = a$estimate[3]

    muexp = median(X)
    N = length(X)
    b = sum(abs(X-muexp))/N

    a = fitdistr(X,"cauchy")
    mucauchy = a$estimate[1]
    gamma = a$estimate[2]

    mypt <- function(x, m, s, df) pt((x-m)/s, df)
    curve(surv(x),from = quantile(X,f),to = quantile(X,t),lty=1,lwd=1)
    curve(1-pnorm(x,munorm,std),add=T,col=2,lty=2,lwd=2)
    curve(1-mypt(x,m,s,df),add=T,col=3,lty=2,lwd=2)
    curve(1-pdexp(x,muexp,b),add=T,col=4,lty=2,lwd=2)
    curve(1-pcauchy(x,location = mucauchy,scale = gamma),add=T,col=5,lty=2,lwd=2)
    curve(1-pmymixestimate(x,X),add=T,col=6,lty=2,lwd=2)
    legend('topright', legend = c("Normal","t","double exp","cauchy","mix normal"),lty
        =1, col=c(2,3,4,5,6), bty='n', cex=.75)
}
```

```
QQ <- function(X)
{
    munorm = mean(X)
    std = sd(X)*n/(n-1)

    a = fitdistr(X, "t", start = list(m=mean(X),s=sd(X), df=2), lower=c(-1, 0.001,1))
        [1]
    m = a$estimate[1]
    s = a$estimate[2]
    df = a$estimate[3]

    muexp = median(X)
    N = length(X)
    b = sum(abs(X-muexp))/N

    a = fitdistr(X,"cauchy")
    mucauchy = a$estimate[1]
    gamma = a$estimate[2]

    myqt <- function(x,m,s,df) s*qt(x,df)+m
    xseq = seq(0,1,0.01)
    x = quantile(X,xseq)
    y1 = qnorm(xseq,munorm,std)
    y2 = myqt(xseq,m,s,df)
    y3 = qdexp(xseq,muexp,b)
    y4 = qcauchy(xseq,location = mucauchy,scale = gamma)
    y5 = qmymixestimate(xseq,X)
    par(mfrow=c(3,2))
    plot(x,y1,type = "l",col=2,ylab = "Normal Distribution",lwd=2)
    abline(0,1)
    plot(x,y2,type = "l",col=2,ylab = "t Distribution",lwd=2)
    abline(0,1)
    plot(x,y3,type = "l",col=2,ylab = "Laplace Distribution",lwd=2)
    abline(0,1)
    plot(x,y4,type = "l",col=2,ylab = "Cauchy Distribution",lwd=2)
    abline(0,1)
    plot(x,y5,type = "l",col=2,ylab = "Mix Normal Distribution",lwd=2)
    abline(0,1)
    par(mfrow=c(1,1))
}

jpeg("C_survival.jpeg")
survival(A[,2])
dev.off()

jpeg("SP_survival.jpeg")
survival(A[,3])
dev.off()
```

Explanation :

Instead of trial and error method, the help of inbuilt MLE estimate functions were taken to get good estimators of hyperparameters for some distributions.
For Normal Distribution

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}(X_i)$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

Inbuilt function was used to find estimates of hyper-parameters of t distribution.

For laplace distribution:

$$\hat{\mu} = \underset{i=1,\ldots,n}{\text{median}} X_i$$

$$\hat{b} = \frac{1}{N}\sum_{i=1}^{n}|X_i - \hat{\mu}|$$

Inbuilt function was used to find estimates of hyper-parameters of Cauchy distribution.

For Mixed Normal Distribution, by looking at the histogram fit, the parametsrs were estimated. $\hat{\mu}_1$ and $\hat{\mu}_2$ were kept equal to the emperical mean. $\hat{\sigma}_1$ was kept same as emperical standard deviation and $\hat{\sigma}_2$ was kept equal to $\frac{\hat{\sigma}_1}{3}$. p was kept 0.4.

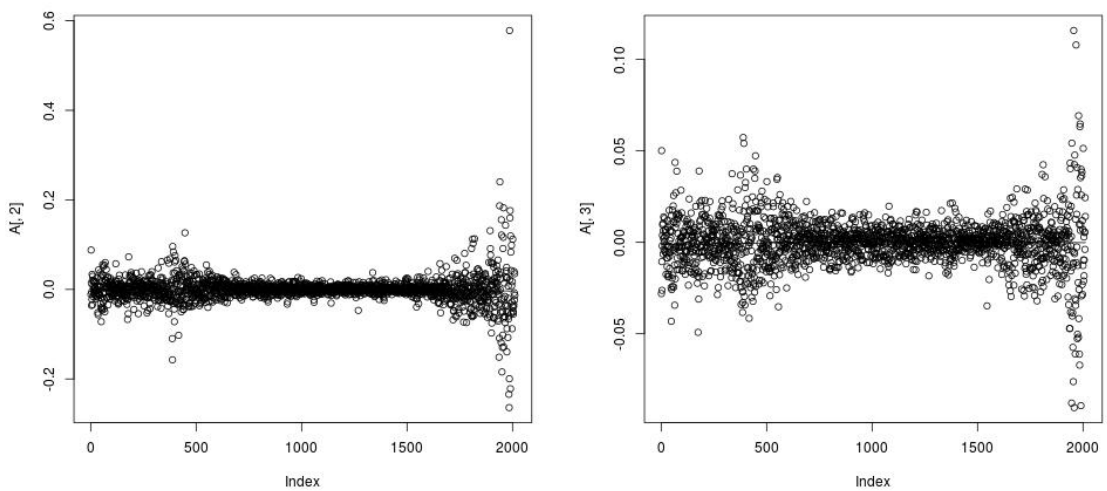
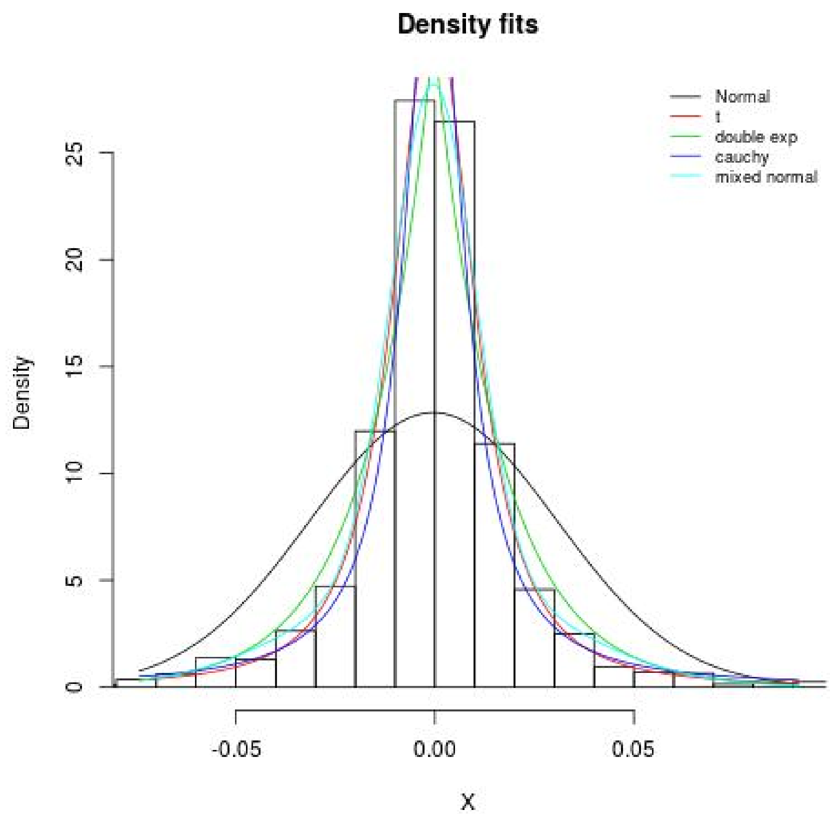


Figure 1: Scatter Plot of C and SP

After fitting the distribution on the histogram, the results on the 2 data were :



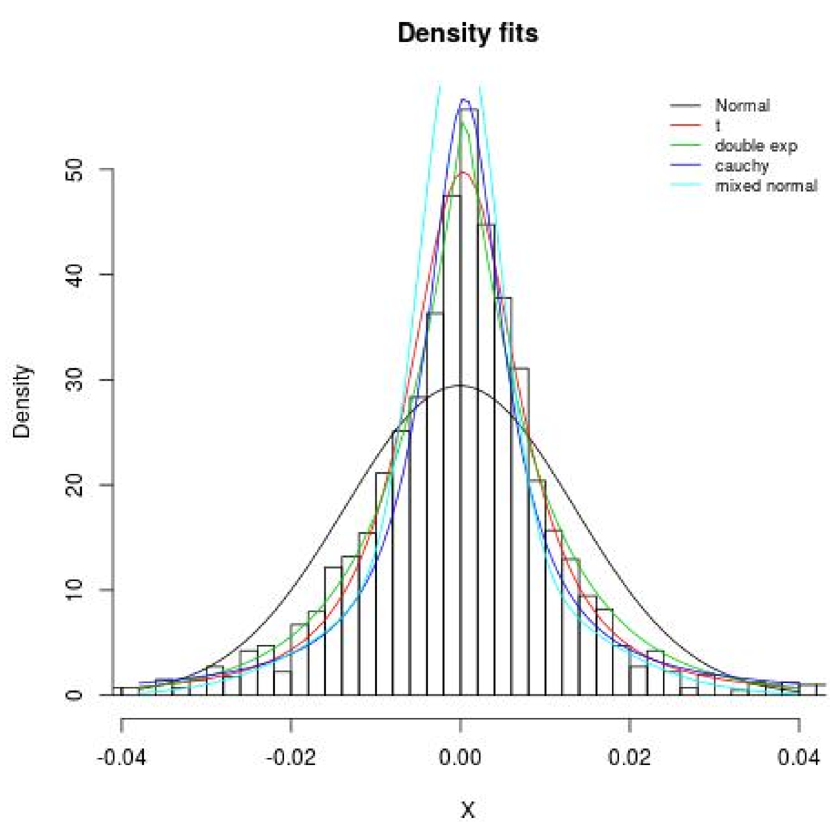


Figure 2: Density fits of C and SP

From the plot, it appears that t, double exponential and Cauchy fit the histogram better than Normal. Still I woulkd choose double exponential as it appears that it is able to fit the histogram on a wider interval.

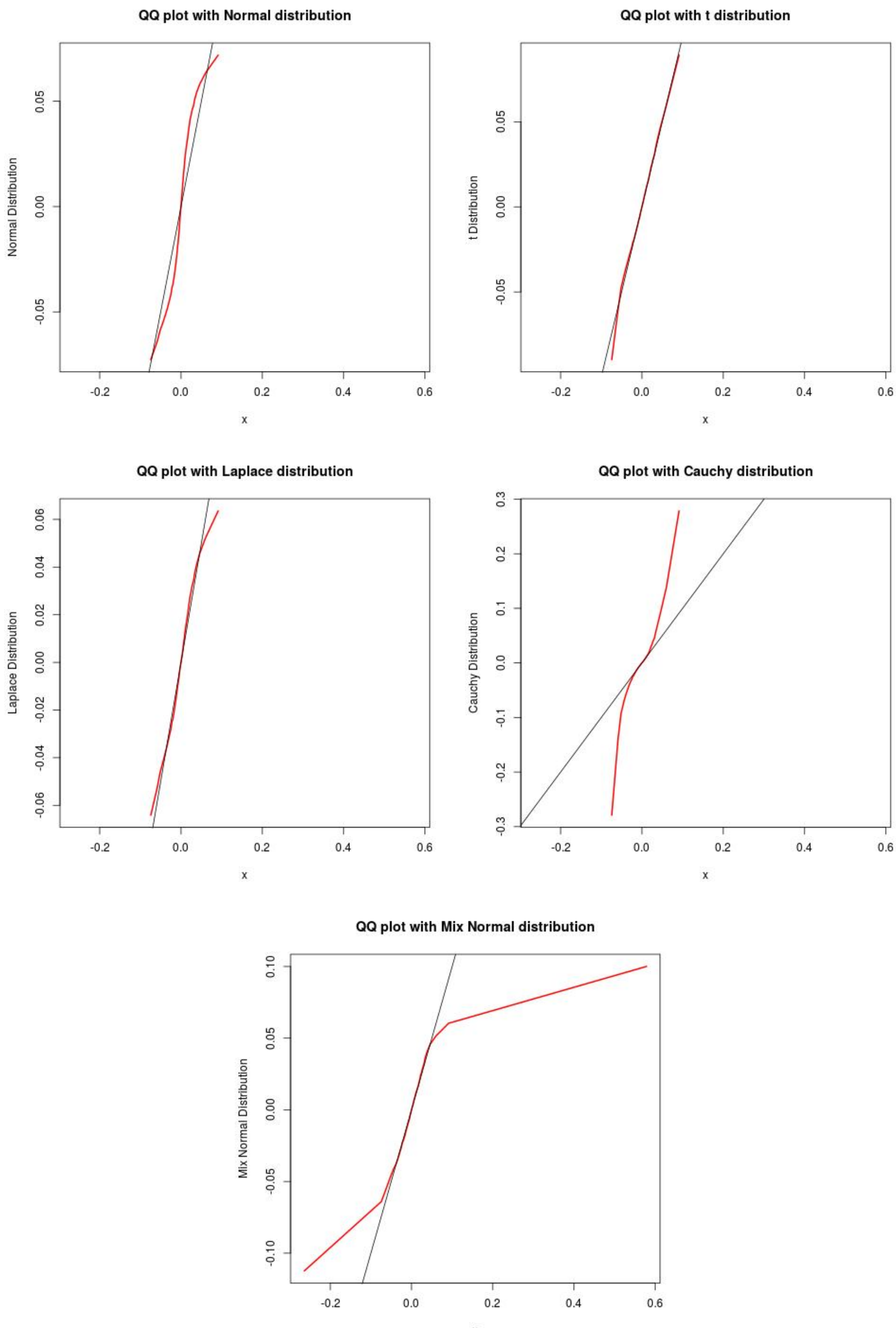


Figure 3: Quantile plots of C

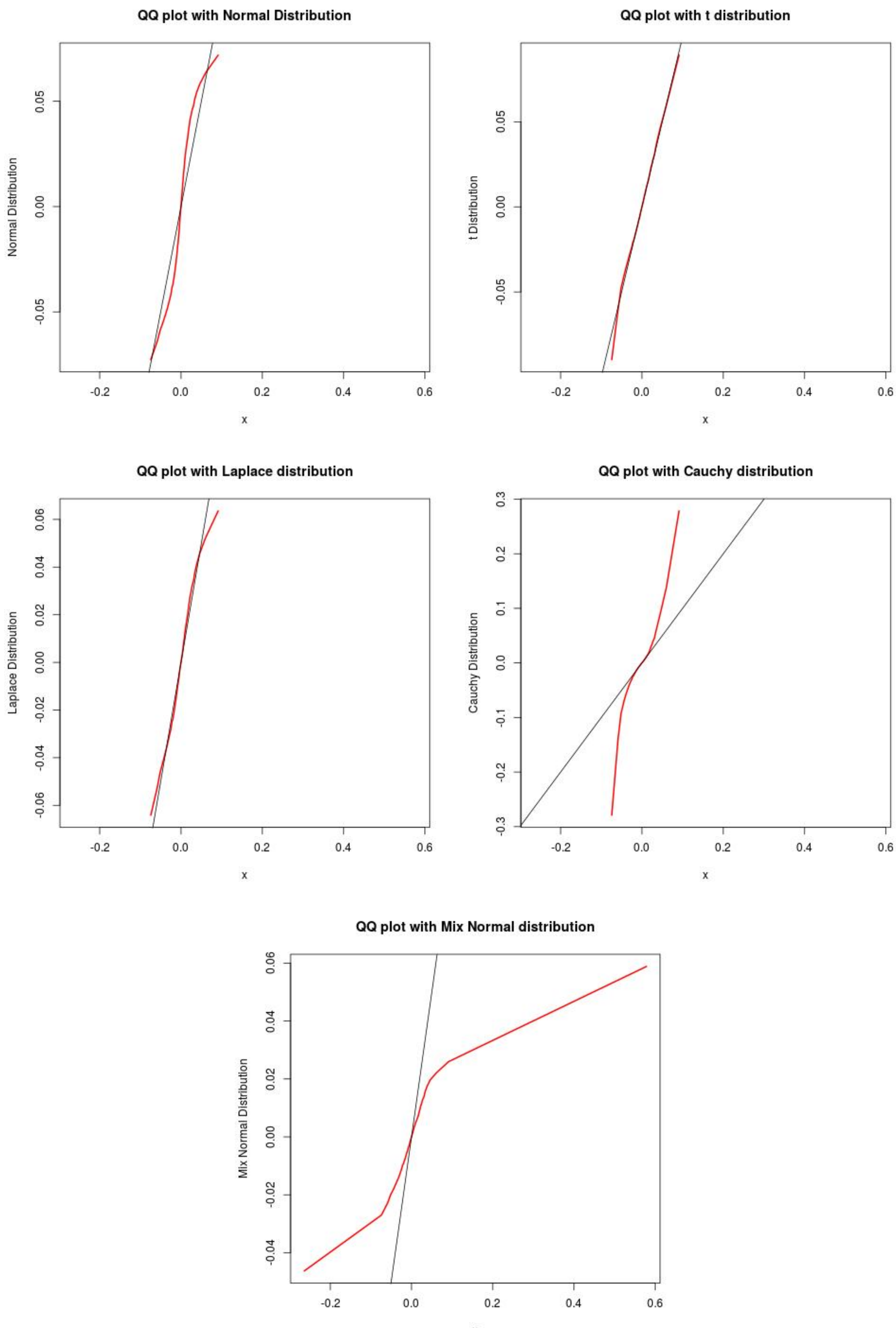
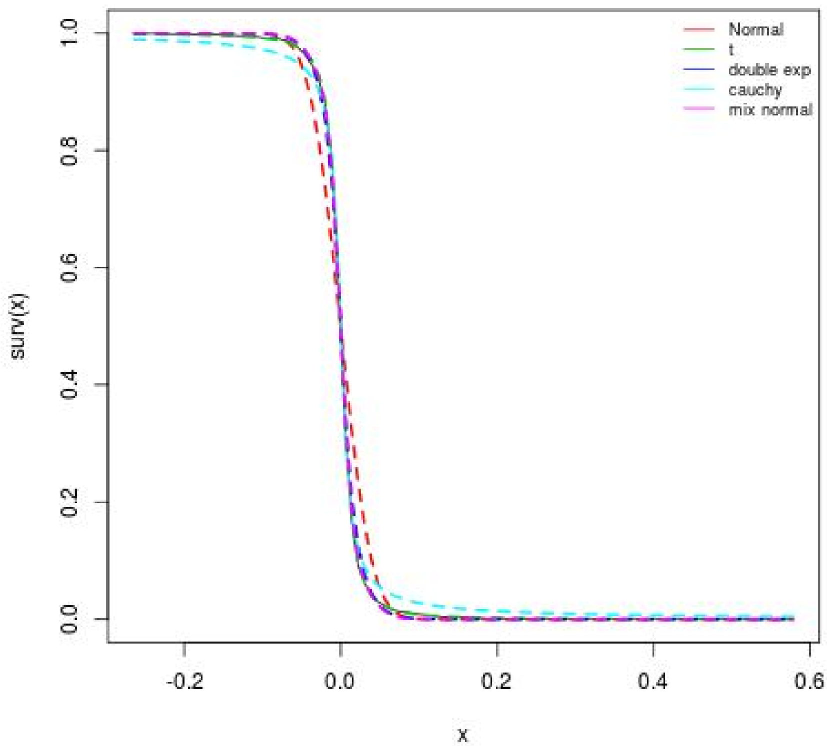


Figure 4: Quantile plots of SP

Interesting plots!
In both the cases, we observe that t and Laplace distribution give better QQ plots than Normal and Cauchy. Cauchy has very heavy tails and Normal is not able to fit good with the data. Mix normal is the worst :/



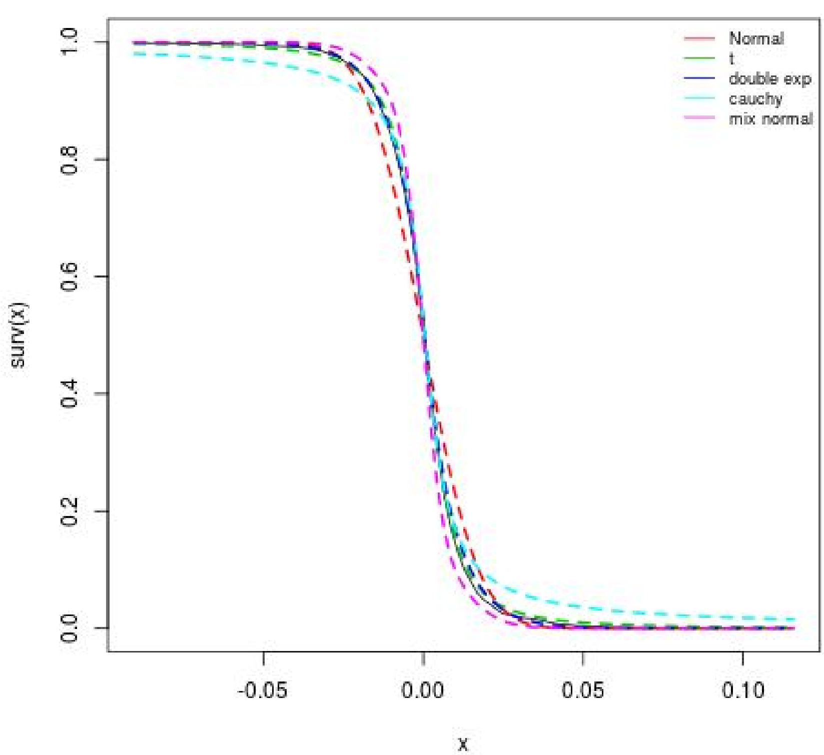


Figure 5: Density fits of C and SP

Problem 1 continued on next page. . .

Similar observations from Survival plots, Cauchy's being a very heavy tailed distribution is not fitting good whereas t and Laplace have best fittings to emperical survival function. Normal and Mixed normal also does not fit good.