Solution_HW3

Wei-Rou,Xu

2023-11-15

Given the following distributions, use the method of moments (MME) to estimate the model parameters for the two datasets in Homework 2.

- Normal distribution with μ and σ^2.
- Exponential distribution with rate parameter λ.
- Laplace distribution with location parameter μ and scale parameter b.
- Gamma distribution with shape parameter α and rate parameter λ.

Please answer the following questions:

(a) Use the method of moments (MME) to estimate the model parameters.

Distribution	Dataset1	Dataset2
Normal	$(\hat{\mu},\hat{\sigma}^2)=(0.02471,0.71386)$	$(\hat{\mu},\hat{\sigma}^2)=(19.24666,26.80621)$
Exponential	$\hat{\lambda}=40.46803$ or $\hat{\lambda}=2.04496$	$\hat{\lambda}=0.05196$ or $\hat{\lambda}=0.18858$
Laplace	$(\hat{\mu},\hat{b})=(0.02471,0.59744)$	$(\hat{\mu},\hat{b})=(19.24668,3.66103)$
Gamma	$(\hat{\alpha},\hat{\lambda})=(0.00086,0.03461)$	$(\hat{\alpha},\hat{\lambda})=(13.81899,0.71799)$

```r
################################ Set1 ################################
set1=read.csv("C:/Set1.csv")

##### normal #####
data.set <-as.numeric(set1$x)
obj.norm <- function(par){
  mu <- par[1]
  sigma2 <- par[2]
  y <- (mean(data.set)-mu)^2+(var(data.set)-sigma2)^2
  return(y)
}

est.normal.set1 <- optim(c(0.025, 0.7),obj.norm)
est.normal.set1$par
```

```
## [1] 0.02471033 0.71386056
```

```r
##### exponential-(1) #####
obj.exp.1<- function(par){
  lambda <- par
  y <- (mean(data.set)-(1/lambda))^2
  return(y)
}
est.exp.set1.1 <- optim(40, obj.exp.1, method = "Brent", lower = -20, upper = 50)
est.exp.set1.1$par
```

```
## [1] 40.46803
```

```r
## Using this methode to estimate parameters of the exponential distribution will also gain score.

##### exponential-(2) #####
obj.exp.2<- function(par){
  lambda <- par
  y <- (mean(data.set)-(1/lambda))^2+(var(data.set)-(1/(lambda^2)))^2
  return(y)
}
est.exp.set1.2 <- optim(2.04, obj.exp.2, method = "Brent", lower = -20, upper = 50)
est.exp.set1.2$par
```

```
## [1] 2.044962
```

```r
##### Laplace #####
obj.Laplace<- function(par){
  mu <- par[1]
  b <- par[2]
  y <- (mean(data.set)-mu)^2+(var(data.set)-(2*b^2))^2
  return(y)
}

est.Laplace.set1 <- optim(c(0.02,0.6),obj.Laplace)
est.Laplace.set1$par
```

```
## [1] 0.0247106 0.5974367
```

```r
##### Gamma #####
obj.Gamma<- function(par){
  alpha <- par[1]
  lambda <- par[2]
  y <- (mean(data.set)-(alpha/lambda))^2+(var(data.set)-(alpha/(lambda^2)))^2
  return(y)
}
est.Gamma.set1 <- optim(c(0.0008,0.03),obj.Gamma)
est.Gamma.set1$par
```

```
## [1] 0.0008550556 0.0346092452
```

```r
############################### Set2 ###############################
set2=read.csv("C:/Set2.csv")
data.set <-as.numeric(set2$x)

##### normal #####
est.normal.set2 <- optim(c(19, 26),obj.norm)
est.normal.set2$par
```

```
## [1] 19.24666 26.80621
```

```r
##### exponential-(1) #####
est.exp.set2.1 <- optim(0.05, obj.exp.1, method = "Brent", lower = -20, upper = 50)
est.exp.set2.1$par
```

```
## [1] 0.05195693
```

```r
## Using this methode to estimate parameters of the exponential distribution will also gain score.

##### exponential-(2) #####
est.exp.set2.2 <- optim(0.19, obj.exp.2, method = "Brent", lower = -20, upper = 50)
est.exp.set2.2$par
```

```
## [1] 0.1885755
```

```r
##### Laplace #####
est.Laplace.set2 <- optim(c(19,3.6),obj.Laplace)
est.Laplace.set2$par
```

```
## [1] 19.246675  3.661028
```

```r
##### Gamma #####
est.Gamma.set2 <- optim(c(13.8,0.71),obj.Gamma)
est.Gamma.set2$par
```

```
## [1] 13.8189853  0.7179937
```

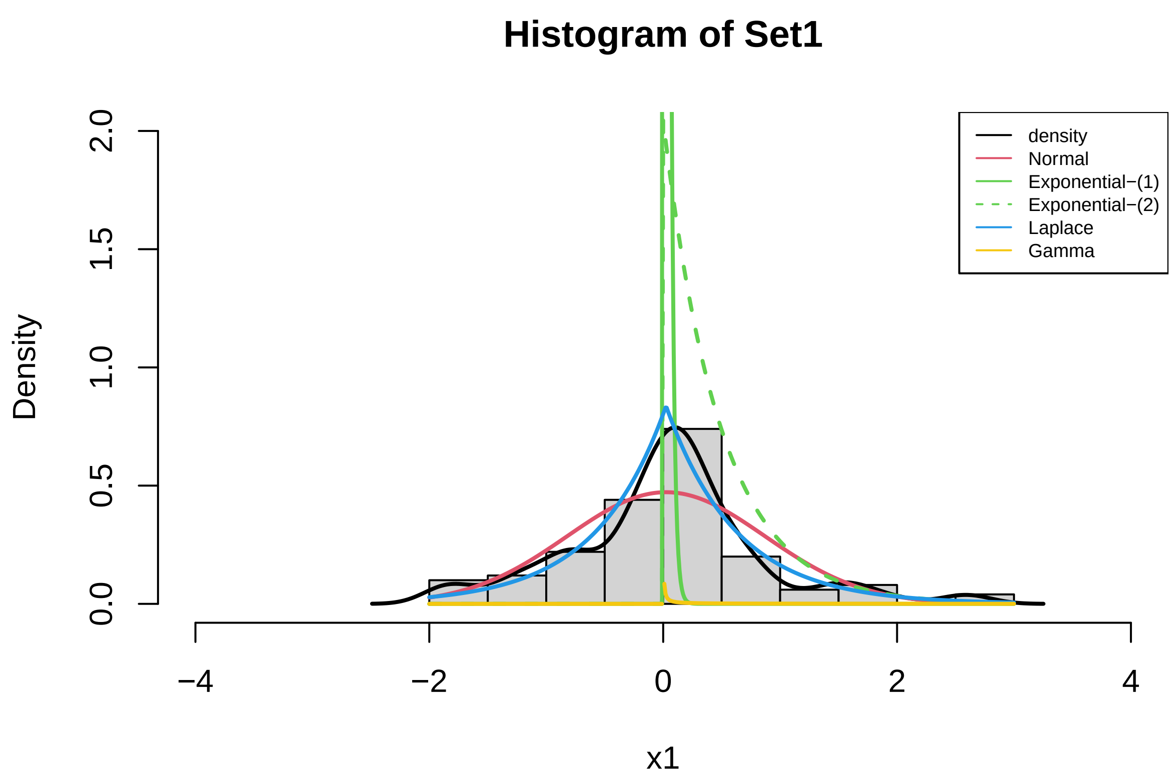
(b) For each dataset, add the "fitted probability density functions" of the given distributions to the histograms of the two datasets. Tying to select more suitable distributions to the data based on your opinion.

```r
############################### Set1 ###############################
x1=set1$x
hist(x1, probability = TRUE, xlim = c(-4, 4), ylim = c(0, 2),main="Histogram of Set1")
lines(density(x1), col = 1, lwd = 2)
xx1 <- seq(-2,3, 0.01)

#Normal
lines(xx1, dnorm(xx1, est.normal.set1$par[1], sqrt(est.normal.set1$par[2])), col = 2, lwd = 2)
```

```
#exponential-(1)
lines(xx1, dexp(xx1, est.exp.set1.1$par), col = 3, lwd = 2)
#exponential-(2)
lines(xx1, dexp(xx1, est.exp.set1.2$par), col = 3, lwd = 2,lty=2)
#Laplace
library(extraDistr)
lines(xx1, dlaplace(xx1,est.Laplace.set1$par[1],est.Laplace.set1$par[2]), col = 4, lwd = 2)
#gamma
lines(xx1, dgamma(xx1,shape=est.Gamma.set1$par[1],rate =est.Gamma.set1$par[2]), col = 7, lwd = 2)

legend("topright", c("density", "Normal", "Exponential-(1)",
                     "Exponential-(2)","Laplace","Gamma"),
       lty = c(1, 1,1,2,1,1), col = c(1,2,3,3,4,7), cex = 0.6)
```

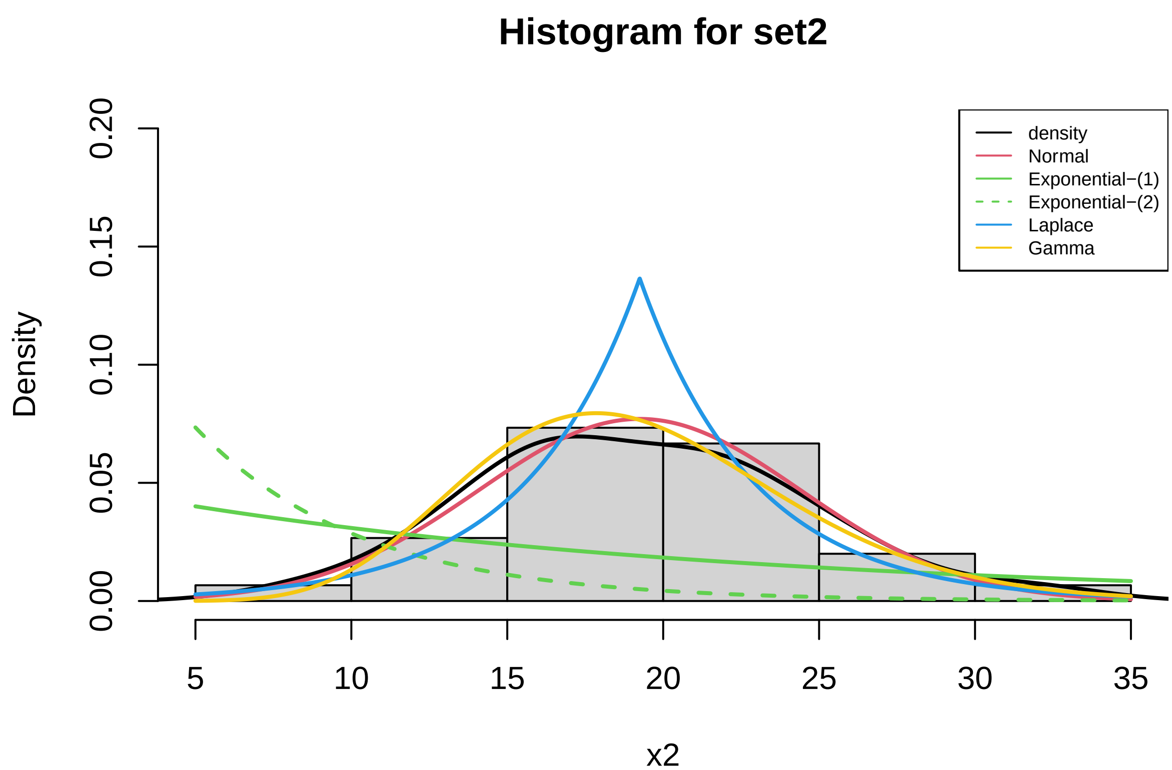



```
################################ Set2 ################################
x2=set2$x
hist(x2, probability = TRUE, xlim = c(5,35), ylim = c(0, 0.2),main="Histogram for set2")
lines(density(x2), col = 1, lwd = 2)
xx2<-  seq(5,35, 0.01)

#Normal
lines(xx2, dnorm(xx2, est.normal.set2$par[1], sqrt(est.normal.set2$par[2])), col = 2, lwd = 2)
#exponential-(1)
lines(xx2, dexp(xx2, est.exp.set2.1$par), col = 3, lwd = 2)
#exponential-(2)
```

```
lines(xx2, dexp(xx2, est.exp.set2.2$par), col = 3, lwd = 2,lty=2)
#Laplace
library(extraDistr)
lines(xx2, dlaplace(xx2,est.Laplace.set2$par[1],est.Laplace.set2$par[2]), col = 4, lwd = 2)
#gamma
lines(xx2, dgamma(xx2,shape=est.Gamma.set2$par[1],rate =est.Gamma.set2$par[2]), col = 7, lwd = 2)

legend("topright", c("density", "Normal", "Exponential-(1)",
                     "Exponential-(2)","Laplace","Gamma"),
       lty = c(1, 1,1,2,1,1), col = c(1,2,3,3,4,7), cex = 0.6)
```

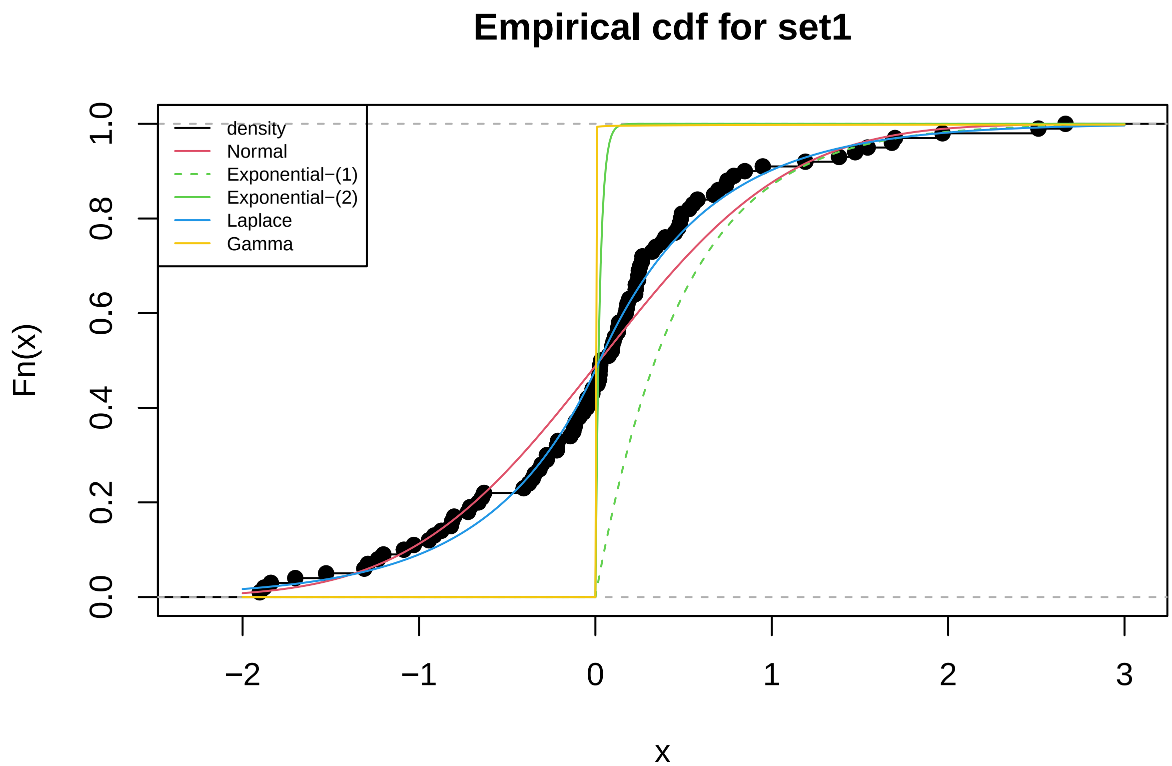


From the figures, I suggest that the laplace and normal distribution is for set 1 and the normal and gamma distribution for set 2.

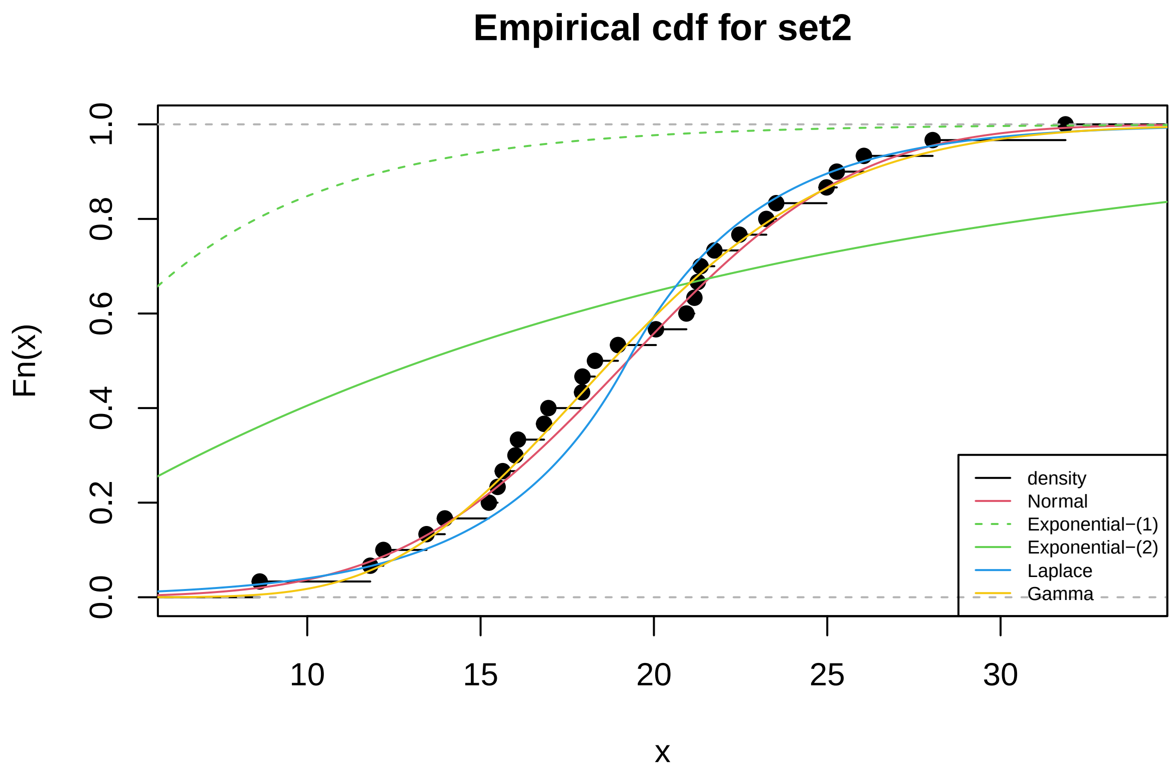
(c) Plot the "fitted cumulative distribution functions (cdf)" with the empirical cdf of the two datasets.

```
################################ Set1 ################################
data.set <-as.numeric(set1$x)
plot(ecdf(data.set),main="Empirical cdf for set1")
lines(xx1, pnorm(xx1, est.normal.set1$par[1], sqrt(est.normal.set1$par[2])), col = 2)
lines(xx1, pexp(xx1, est.exp.set1.1$par), col = 3)
lines(xx1, pexp(xx1, est.exp.set1.2$par), col = 3,lty=2)
lines(xx1, plaplace(xx1, est.Laplace.set1$par[1],est.Laplace.set1$par[2]), col = 4)
```

```r
lines(xx1, pgamma(xx1, est.Gamma.set1$par[1],est.Gamma.set1$par[2]), col = 7)
legend("tople", c("density", "Normal", "Exponential-(1)",
                     "Exponential-(2)","Laplace","Gamma"),
       lty = c(1,1,2,1,1), col = c(1,2,3,3,4,7), cex = 0.6)
```




```r
############################### Set2 ###############################
data.set <-as.numeric(set2$x)
plot(ecdf(data.set),main="Empirical cdf for set2")
lines(xx2, pnorm(xx2, est.normal.set2$par[1], sqrt(est.normal.set2$par[2])), col = 2)
lines(xx2, pexp(xx2, est.exp.set2.1$par), col = 3)
lines(xx2, pexp(xx2, est.exp.set2.2$par), col = 3,lty=2)
lines(xx2, plaplace(xx2, est.Laplace.set2$par[1],est.Laplace.set2$par[2]), col = 4)
lines(xx2, pgamma(xx2, est.Gamma.set2$par[1],est.Gamma.set2$par[2]), col = 7)
legend("bottomri", c("density", "Normal", "Exponential-(1)",
                     "Exponential-(2)","Laplace","Gamma"),
       lty = c(1,1,2,1,1), col = c(1,2,3,3,4,7), cex = 0.6)
```



From the figures, I suggest that the laplace distribution is better than the Normal distribution for set 1 and the normal and gamma distribution for set 2.

(d) Provide the necessary evidence for selecting suitable models for the two datasets via suitable hypothesis testings. (could be choose one or more.)

```
############################### Set1 ###############################
data.set <-as.numeric(set1$x)
###### ks.test ######
ks.test(data.set, rnorm(1000,est.normal.set1$par[1], sqrt(est.normal.set1$par[2])))$p
```

```
## [1] 0.165906
```

```
ks.test(data.set, rexp(1000,est.exp.set1.1$par))$p
```

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

```
ks.test(data.set, rlaplace(1000,est.Laplace.set1$par[1],est.Laplace.set1$par[2]))$p
```

```
## [1] 0.9197635
```

```r
ks.test(data.set, rgamma(1000,est.Gamma.set1$par[1],est.Gamma.set1$par[2]))$p
```

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

```r
###### cvm.test ######
library("goftest")
cvm.test(data.set, "pnorm",est.normal.set1$par[1], sqrt(est.normal.set1$par[2]))$p
```

```
## [1] 0.1340311
```

```r
cvm.test(data.set, "pexp",est.exp.set1.1$par)$p
```

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

```r
cvm.test(data.set, "plaplace", est.Laplace.set1$par[1],est.Laplace.set1$par[2])$p
```

```
## [1] 0.7716481
```

```r
cvm.test(data.set, "pgamma",est.Gamma.set1$par[1],est.Gamma.set1$par[2])$p
```

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

	Normal	Exponential	Laplace	Gamma
Kolmogorov–Smirnov test	p-value>0.5	p-value<0.00001	p-value>0.5	p-value<0.00001
Cramer-Von Mises Test	p-value>0.5	p-value<0.00001	p-value>0.5	p-value<0.00001

If p-value<0.05, the null hypothesis is rejected, indicating that we have enough evidence to say that the data is not from the distribution. According to the test results, Set 1 from the Normal distribution and Laplace distribution.

```r
############################### Set2 ###############################
data.set <-as.numeric(set2$x)
###### ks.test ######
ks.test(data.set, rnorm(1000,est.normal.set2$par[1], sqrt(est.normal.set2$par[2])))$p
```

```
## [1] 0.9837968
```

```r
ks.test(data.set, rexp(1000,est.exp.set2.1$par))$p
```

```
## [1] 7.35576e-05
```

```r
ks.test(data.set, rlaplace(1000,est.Laplace.set2$par[1],est.Laplace.set2$par[2]))$p
```

```
## [1] 0.7412729
```

```r
ks.test(data.set, rgamma(1000,est.Gamma.set2$par[1],est.Gamma.set2$par[2]))$p
```

```
## [1] 0.976918
```

```r
###### cvm.test ######
library("goftest")
cvm.test(data.set, "pnorm",est.normal.set2$par[1], sqrt(est.normal.set2$par[2]))$p
```

```
## [1] 0.9909949
```

```r
cvm.test(data.set, "pexp",est.exp.set2.1$par)$p
```

```
## [1] 8.917683e-05
```

```r
cvm.test(data.set, "plaplace", est.Laplace.set2$par[1],est.Laplace.set2$par[2])$p
```

```
## [1] 0.5402988
```

```r
cvm.test(data.set, "pgamma",est.Gamma.set2$par[1],est.Gamma.set2$par[2])$p
```

```
## [1] 0.9961664
```

	Normal	Exponential	Laplace	Gamma
Kolmogorov–Smirnov test	p-value>0.05	p-value<0.00001	p-value>0.05	p-value>0.05
Cramer-Von Mises Test	p-value>0.05	p-value<0.00001	p-value>0.05	p-value>0.05

If p-value<0.05, the null hypothesis is rejected, indicating that we have enough evidence to say that the data is not from the distribution. According to the test results, Set 2 from the Normal distribution, Laplace distribution, and Gamma distribution.