R Assignment 4

Nina Dyatchenko

September 12th, 2018

## Assignment 4: Due Sunday, 16 September at 23:59PM

*For help with Rmarkdown for reports, see this* [*white paper*](http://www.stat.cmu.edu/~cshalizi/rmarkdown/#math-in-r-markdown) *from Carnegie Mellon University’s Department of Statistics and Data Science.*

For each the seven statistical distributions we covered in the last assignment (Normal, Student’s , , , Binomial, Negative Binomial, and Poisson),

1. Generate and store a random vector of 10,000 observations, using the same parameters as the last homework:
   1. ,

set.seed(1)  
xNorm\_n10000 <- rnorm(n=10000, mean=2, sd = sqrt(5))

ii. $t\_{\nu = 4}$,

set.seed(2)  
xrt\_n10000 <- rt(n=10000, df=4)

iii. $\chi^2\_{\nu = 2}$,

set.seed(3)  
xChi\_n10000<- rchisq(n=10000, df=2, ncp=0)

iv. $F\_{n = 90,\ m = 12}$,

set.seed(4)  
xFd\_n10000<- rf(n=10000, df1=90, df2=12)

v. $Bin(n = 9, p = 2/3)$,

set.seed(5)  
xBD\_n10000 <- rbinom(n=10000, size=9, prob=2/3)

vi. $NBin(n = 5, p = 1/2)$, and

set.seed(6)  
xNB\_n10000 <- rnbinom(n = 10000, size = 5, prob = 0.5)

vii. $Pois(\lambda = 3)$.

set.seed(7)  
xP\_n10000 <- rpois(n=10000, 3)

1. Subset the first values from the vector, and of this subset
   1. calculate the 5-Number Summary,

6 number summary?

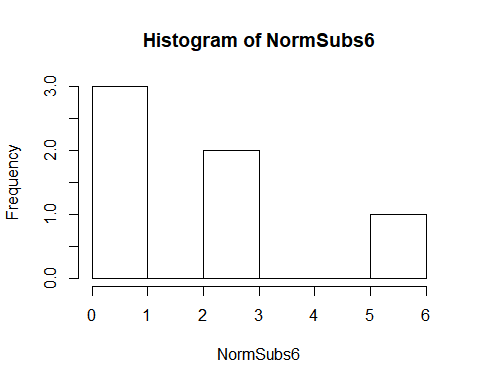
ii. plot the histogram of the subset, and  
iii. plot the estimated density of this subset. (Density assumes as normal).

Normal:

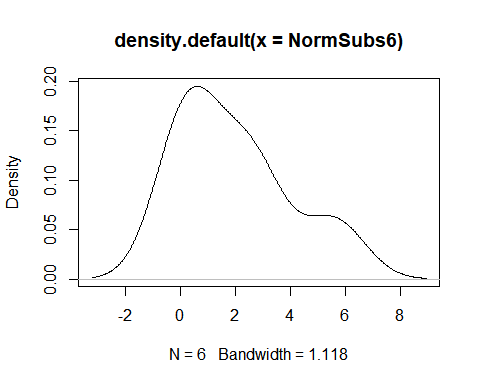
NormSubs6<- xNorm\_n10000[1:6]  
summary(NormSubs6)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1315 0.2738 1.5049 1.9351 2.6553 5.5672

hist(NormSubs6)



plot(density(NormSubs6))

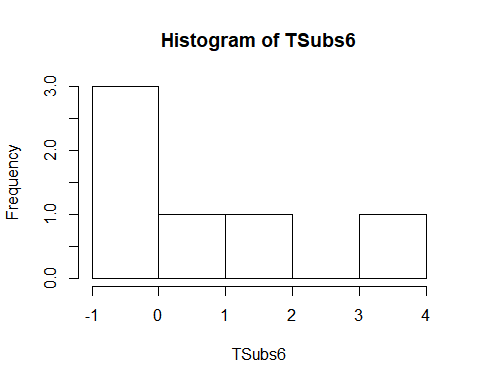


Student’s :

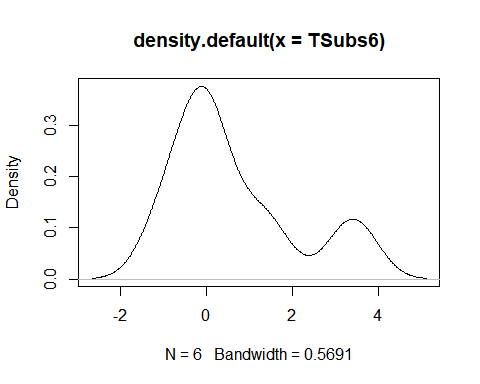
TSubs6<- xrt\_n10000[1:6]  
summary(TSubs6)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.96300 -0.14891 0.03367 0.61769 1.06361 3.40464

hist(TSubs6)



plot(density(TSubs6))

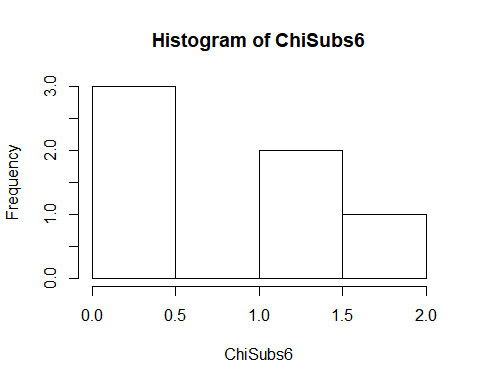


(Not normal)

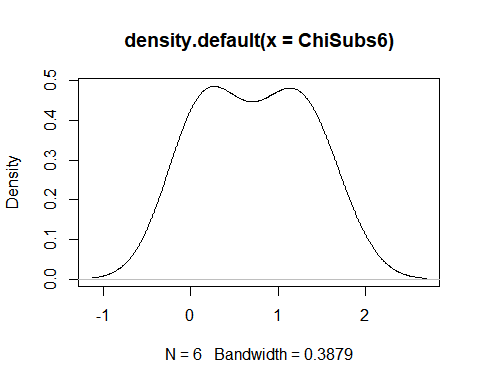
ChiSubs6<- xChi\_n10000[1:6]  
summary(ChiSubs6)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.03434 0.19385 0.74322 0.72766 1.16498 1.52891

hist(ChiSubs6)



plot(density(ChiSubs6))

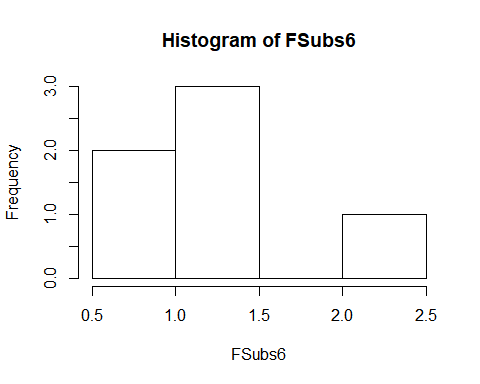


-Distribution

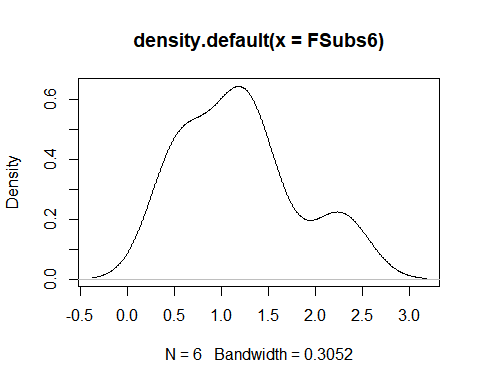
FSubs6<- xFd\_n10000[1:6]  
summary(FSubs6)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.5374 0.7094 1.1642 1.1845 1.3596 2.2585

hist(FSubs6)



plot(density(FSubs6))

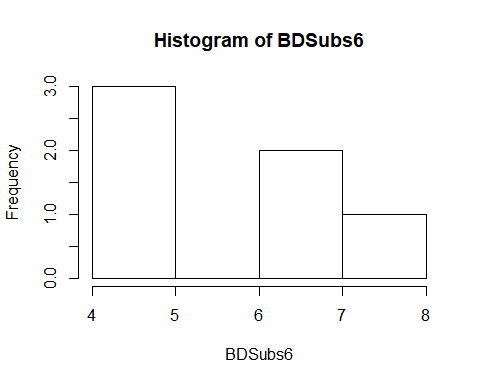


Binomial

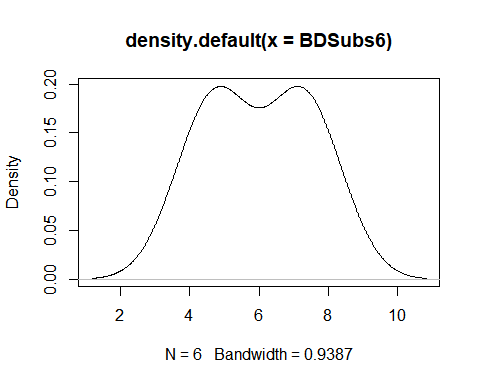
BDSubs6<- xBD\_n10000[1:6]  
summary(BDSubs6)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4 5 6 6 7 8

hist(BDSubs6)



plot(density(BDSubs6))

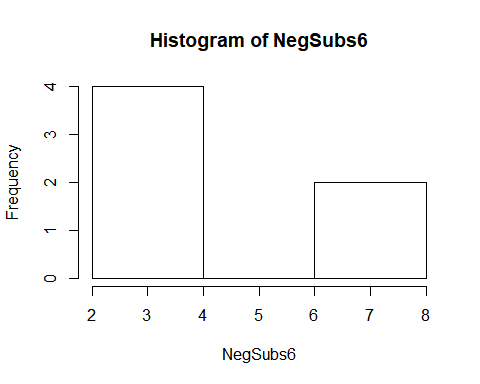


Negative binomial

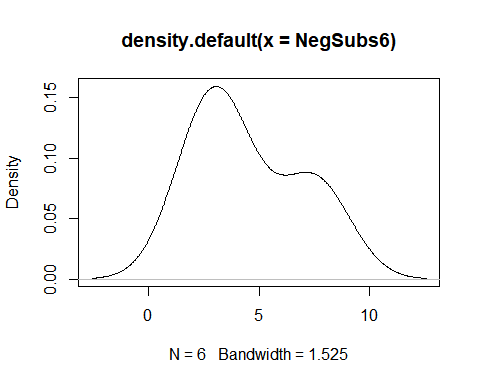
NegSubs6<- xNB\_n10000[1:6]  
summary(NegSubs6)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.00 3.00 3.50 4.50 6.25 8.00

hist(NegSubs6)



plot(density(NegSubs6))

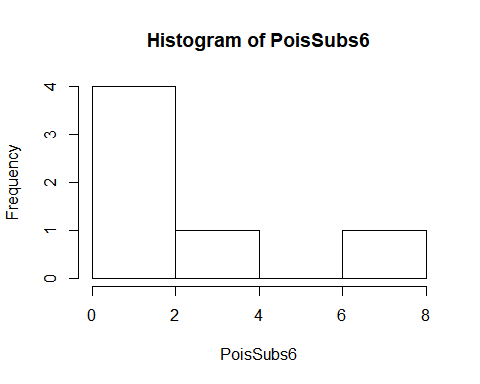


Poisson

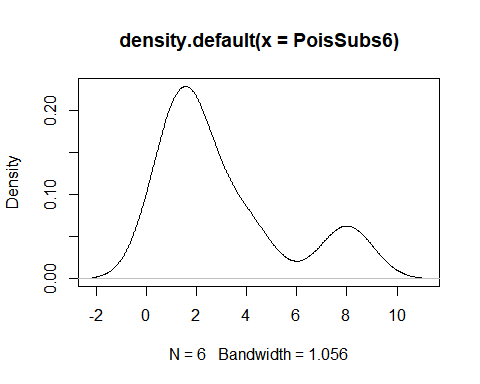
PoisSubs6<- xP\_n10000[1:6]  
summary(PoisSubs6)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.00 1.25 2.00 3.00 3.50 8.00

hist(PoisSubs6)



plot(density(PoisSubs6))



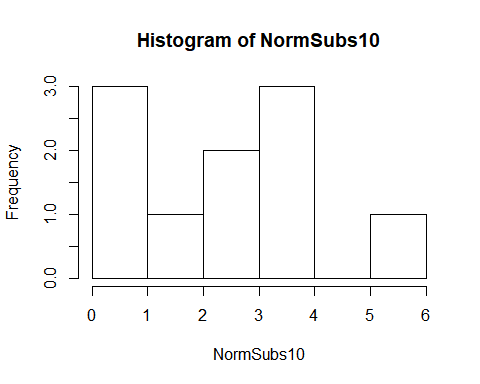
1. Repeat Item 2 for the first values from the random vector you generated in Item 1. Remark on the changing behaviour as the sample size increases.

Normal:

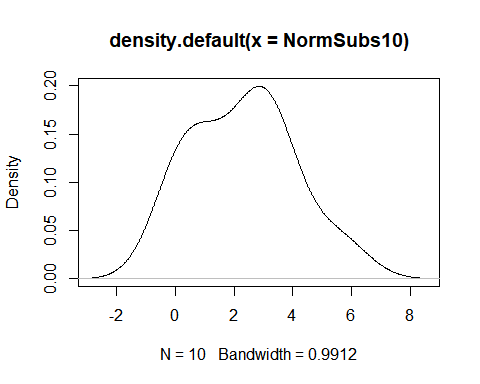
NormSubs10<- xNorm\_n10000[1:10]  
summary(NormSubs10)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1315 0.7787 2.5737 2.2956 3.2381 5.5672

hist(NormSubs10)



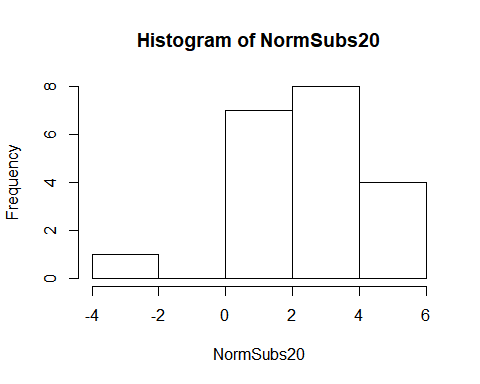
plot(density(NormSubs10))



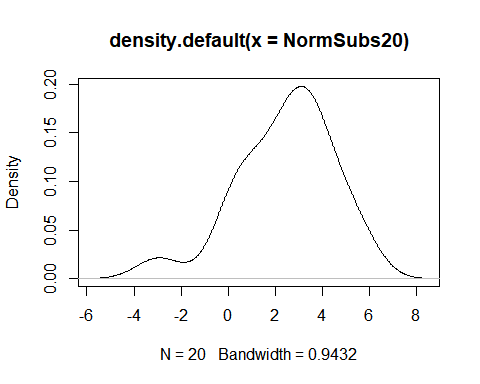
NormSubs20<- xNorm\_n10000[1:20]  
summary(NormSubs20)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.952 1.141 2.804 2.426 3.697 5.567

hist(NormSubs20)



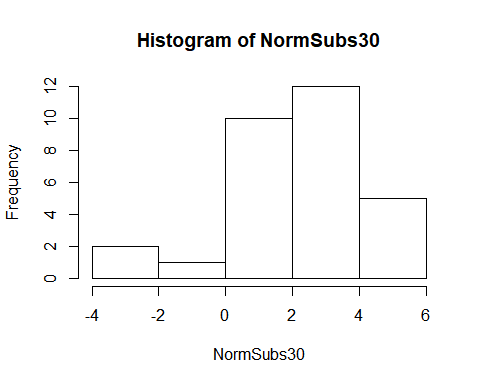
plot(density(NormSubs20))



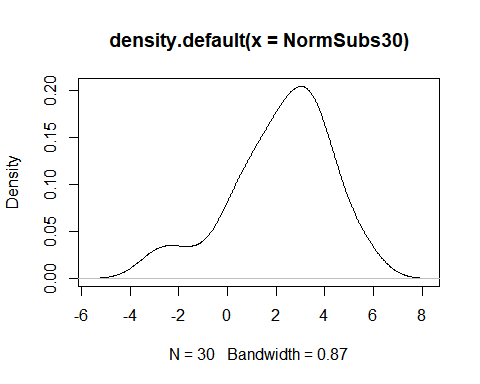
NormSubs30<- xNorm\_n10000[1:30]  
summary(NormSubs30)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.952 1.027 2.574 2.184 3.585 5.567

hist(NormSubs30)



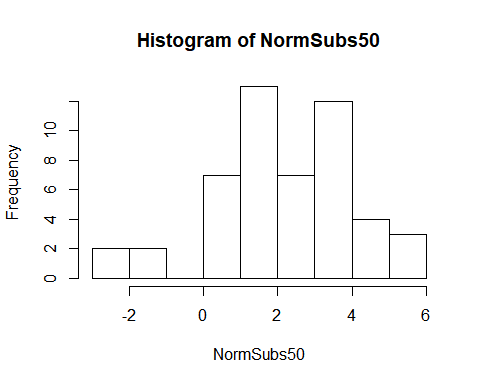
plot(density(NormSubs30))



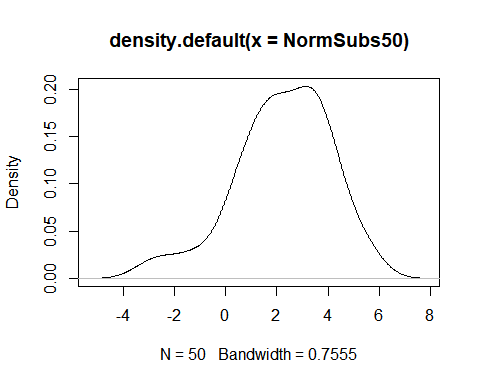
NormSubs50<- xNorm\_n10000[1:50]  
summary(NormSubs50)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.952 1.168 2.289 2.225 3.628 5.567

hist(NormSubs50)



plot(density(NormSubs50))

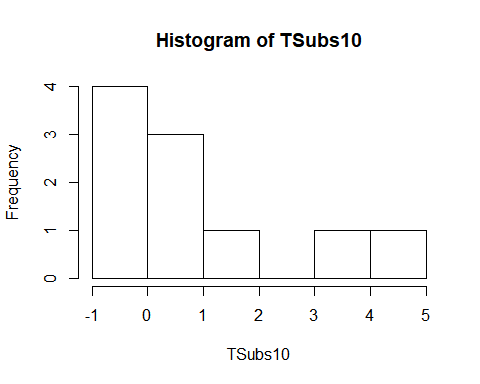


Student’s :

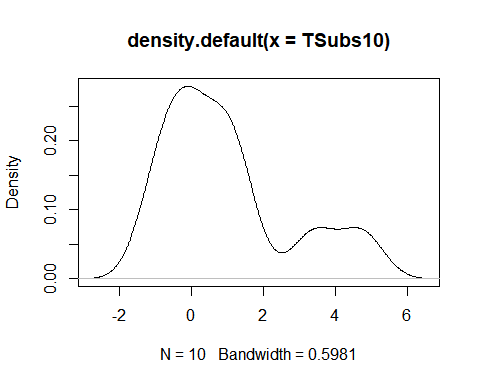
TSubs10<- xrt\_n10000[1:10]  
summary(TSubs10)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.9630 -0.1489 0.5992 0.9656 1.2625 4.7326

hist(TSubs10)



plot(density(TSubs10))



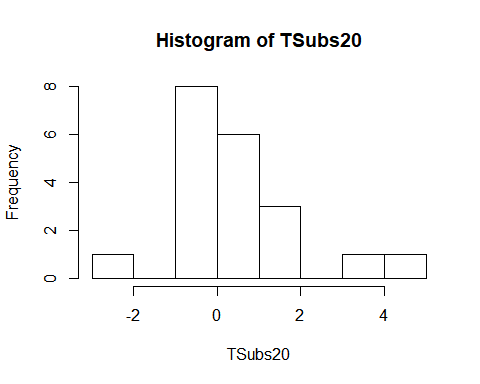
var(TSubs10)

## [1] 3.343344

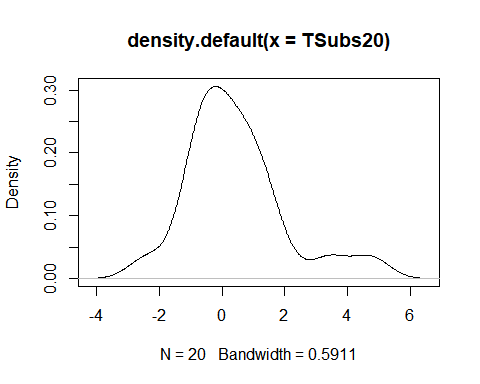
TSubs20<- xrt\_n10000[1:20]  
summary(TSubs20)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.3694 -0.5698 0.2301 0.4317 1.0323 4.7326

hist(TSubs20)



plot(density(TSubs20))



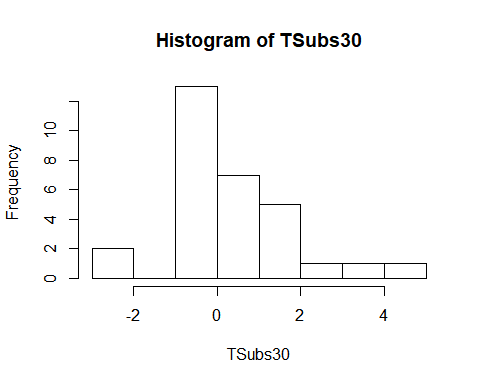
var(TSubs20)

## [1] 2.523075

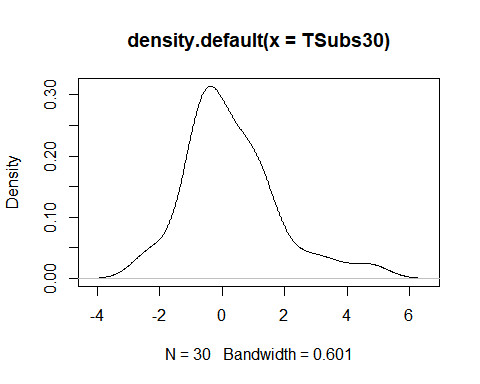
TSubs30<- xrt\_n10000[1:30]  
summary(TSubs30)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.36936 -0.66960 0.03367 0.29728 1.09720 4.73264

hist(TSubs30)



plot(density(TSubs30))



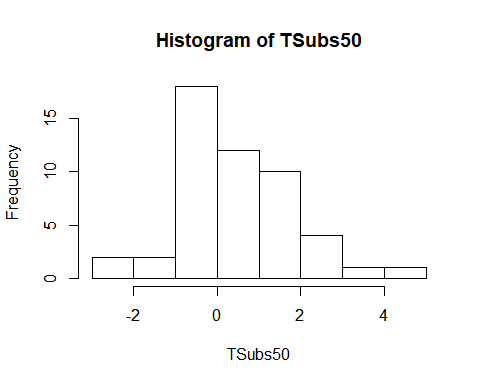
var(TSubs30)

## [1] 2.245901

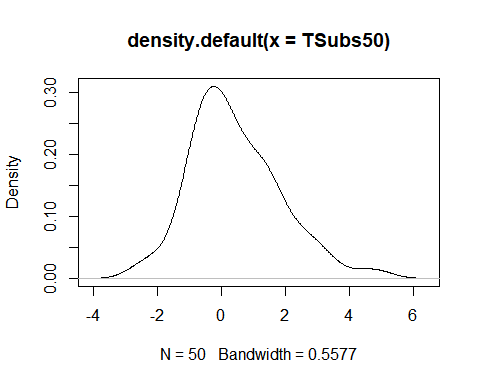
TSubs50<- xrt\_n10000[1:50]  
summary(TSubs50)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.3694 -0.5108 0.2383 0.4470 1.3048 4.7326

hist(TSubs50)



plot(density(TSubs50))



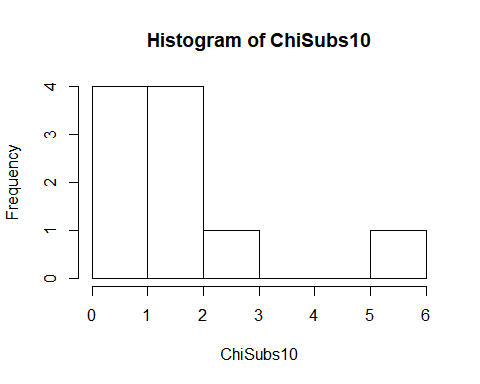
var(TSubs50)

## [1] 1.95557

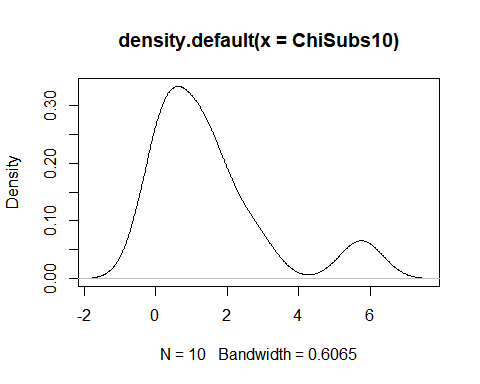
ChiSubs10<- xChi\_n10000[1:10]  
summary(ChiSubs10)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.03434 0.37583 1.11595 1.51670 1.80707 5.75437

hist(ChiSubs10)



plot(density(ChiSubs10))



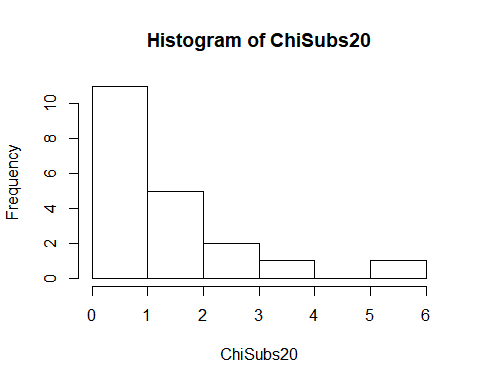
var(ChiSubs10)

## [1] 2.974062

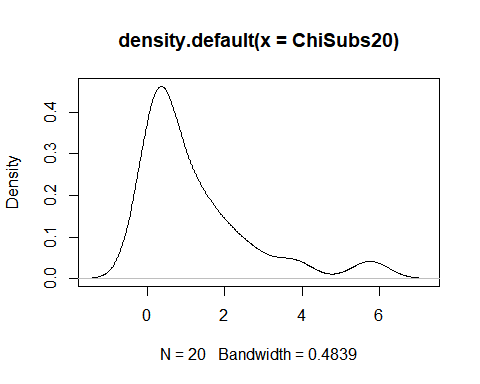
ChiSubs20<- xChi\_n10000[1:20]  
summary(ChiSubs20)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.03434 0.30987 0.53186 1.23980 1.62163 5.75437

hist(ChiSubs20)



plot(density(ChiSubs20))



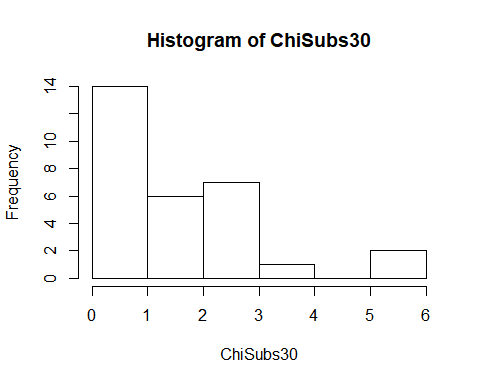
var(ChiSubs20)

## [1] 2.152357

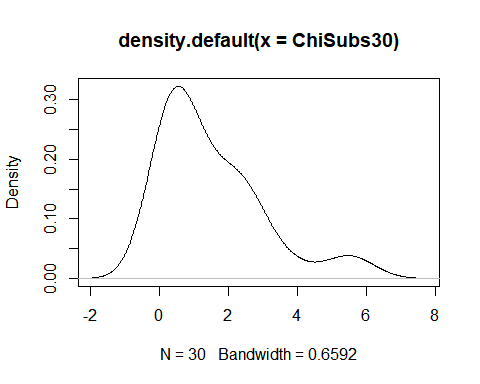
ChiSubs30<- xChi\_n10000[1:30]  
summary(ChiSubs30)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.03434 0.34600 1.10074 1.51396 2.28392 5.75437

hist(ChiSubs30)



plot(density(ChiSubs30))



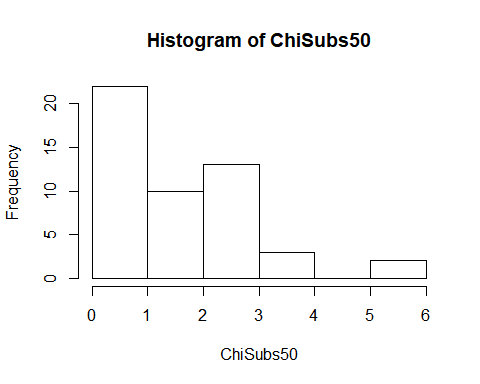
var(ChiSubs30)

## [1] 2.211067

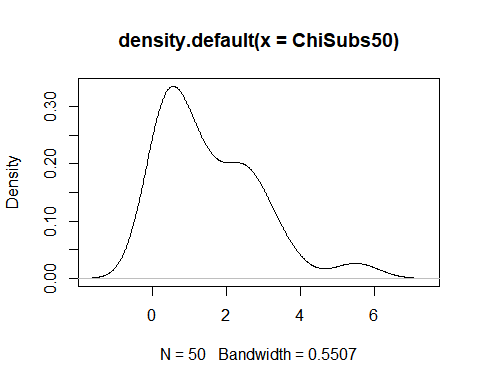
ChiSubs50<- xChi\_n10000[1:50]  
summary(ChiSubs50)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.03434 0.43565 1.19880 1.55241 2.33072 5.75437

hist(ChiSubs50)



plot(density(ChiSubs50))



var(ChiSubs50)

## [1] 1.790028

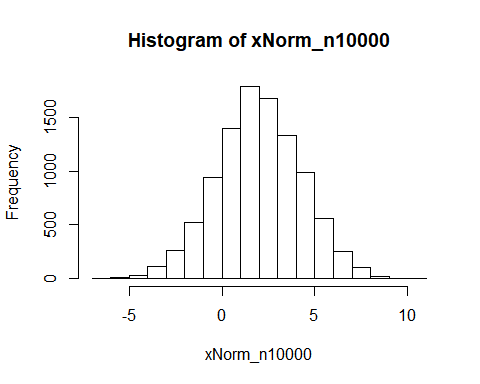
1. Repeat Item 2 for the entire vector (). For smaller values of from continuous distributions, which tool do you think gave a better representation of the full data: histogram or density plot? Did this change when you inspected the discrete distributions?

Normal:

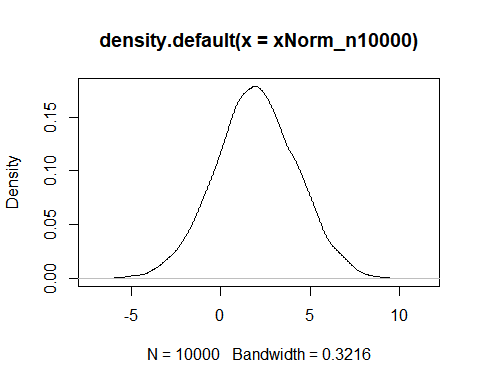
summary(xNorm\_n10000)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -6.2093 0.4942 1.9644 1.9854 3.5153 10.5200

hist(xNorm\_n10000)



plot(density(xNorm\_n10000))



var(xNorm\_n10000)

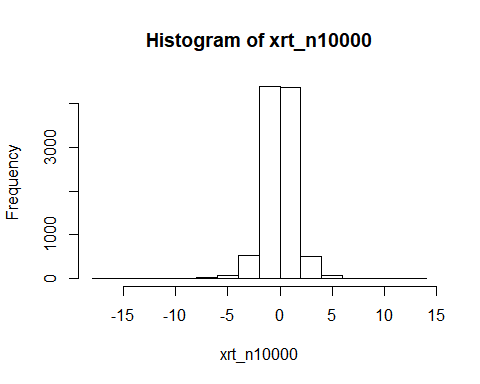
## [1] 5.124328

Student’s :

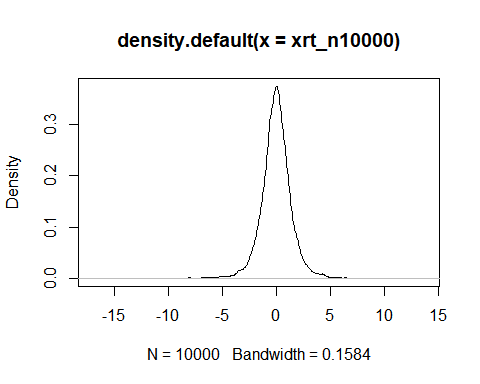
summary(xrt\_n10000)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -16.581973 -0.744999 -0.004989 -0.014118 0.742671 13.393131

hist(xrt\_n10000)



plot(density(xrt\_n10000))

 5. Contrast the 5-Number Summaries at each of the sample sizes (6, 10, 20, 30, 50, and 10000) for the skewed distributions (Negative binomial)vs. the symmetric distributions (Binomial, T, Normal).

summary(NormSubs6)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1315 0.2738 1.5049 1.9351 2.6553 5.5672

summary(NormSubs10)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1315 0.7787 2.5737 2.2956 3.2381 5.5672

summary(NormSubs20)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.952 1.141 2.804 2.426 3.697 5.567

summary(NormSubs30)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.952 1.027 2.574 2.184 3.585 5.567

summary(NormSubs50)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.952 1.168 2.289 2.225 3.628 5.567

summary(xNorm\_n10000)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -6.2093 0.4942 1.9644 1.9854 3.5153 10.5200

var(NormSubs6)

## [1] 4.446056

var(NormSubs10)

## [1] 3.046572

var(NormSubs30)

## [1] 4.269997

var(NormSubs50)

## [1] 3.456079

var(xNorm\_n10000)

## [1] 5.124328