Lab #2: Simulating Distributions Based on the Normal

Carley Dziewicki

Due Monday, May 4

Goals

1. Practice simulating and describing the behavior of independent random variables.
2. Explore the properties of statistics and distributions related to the normal distribution.

Problems

For each of the following, submit the required R commands along with your answers below.

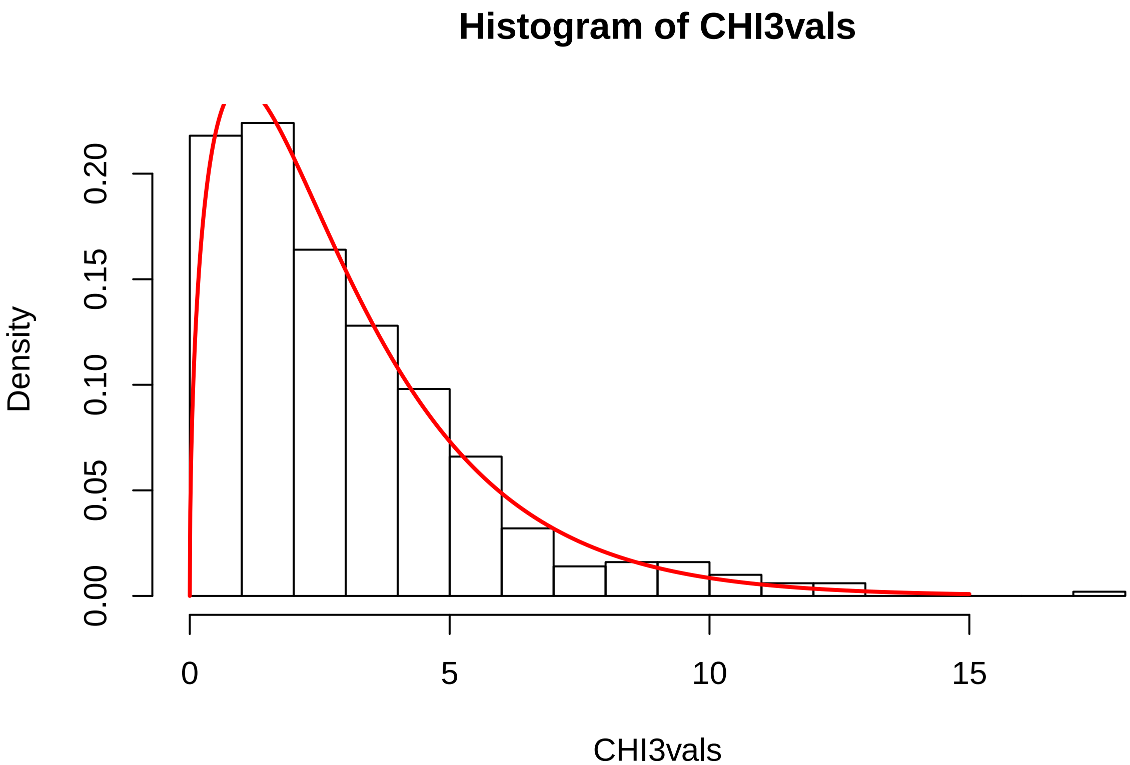
For each of the following, submit the required commands along with your answers. For the first, a script has been provided for you in a separate file, `chap6_lab2script.R`.

1. Simulate a χ^2_3 distribution by summing the squares of three simulated standard normal random variables, each having length 500. Create a density histogram `prob=T` of the simulated χ^2_3 random variable. Superimpose the theoretical χ^2_3 density over the histogram. Comment on how well the simulation fits with what the theory of Section 6.4 would predict. (Note that the command `set.seed(826495)` determines the starting place of the random number generator so that each time you run the script, you will get the same "random" results. Setting the seed is especially useful to ensure that others can reproduce your random simulations.)

Answer: The simulation fits the theory from 6.4 well, that as the degrees of freedom increases the distribution becomes more symmetric and closer to normal, we can see that this distribution with df=3 that the curve is moving towards a normal shape, because it is not as heavily skewed as a chi-sqared distribution with df=1. It is centered around about 2.5 and has a skewed left shape.

```
set.seed(826495)
# Set parameters
degfree <- 3
# Set number of observations
M <- 500
# Initialize vectors of values
# rnorm defaults to Standard Normal
Z1 <- rnorm(M)
Z2 <- rnorm(M)
Z3 <- rnorm(M)
CHI3vals <- Z1^2 + Z2^2 + Z3^2
## density curve below
hist(CHI3vals, prob = T, nclass = "scott")
xvals  <- seq(0,15,.01)
#Compute densities for the chisquare distribution
densities <- dchisq(xvals, degfree)
```

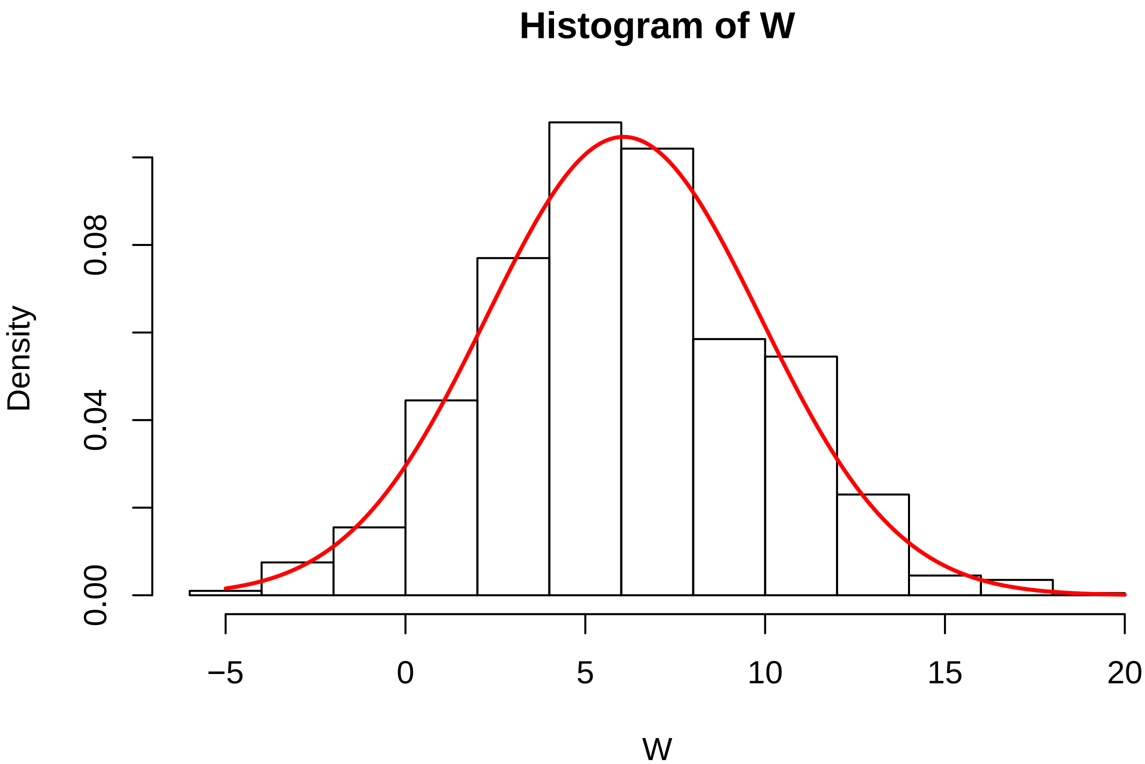
```r
#Superimpose theoretical density over plot
lines(xvals, densities, lwd = 2, col = 2)
```



2. Given $X \sim N(0,1)$, $Y \sim N(2,2)$ and $Z \sim N(4,3)$, what is the distribution of $W = X + Y + Z$? Set the seed equal to 365 and simulate 1000 samples, each of size 1 for X, Y, and Z. Add the values in the three vectors obtain the empirical distribution of W. Create a density histogram of the simulated values of W and superimpose the theoretical density of W.

Answer: The distribution of W looks also to be normal, which is not suprising since we added three normal distribtions together. Then mean and standard deviation also make sense for the addition of the distirbutions. There seems to be no skewness or outliers.

```r
set.seed(365)
# Set number of observations
M <- 1000
n<- 1
# Initialize vectors of values
# rnorm defaults to Standard Normal
X <- rnorm(M,0,1)
Y <- rnorm(M,2,2)
Z <- rnorm(M,4,3)
W <- X + Y + Z
## density curve below
hist(W, prob = T)
# Might need some trial and error to get the endpoints
xvals_w  <- seq(-5,20,.01)
#Compute densities for the chisquare distribution
densities_w <- dnorm(xvals_w,mean(W),sd(W))
#Superimpose theoretical density over plot
lines(xvals_w, densities_w, lwd = 2, col = 2)
```



3. Verify empirically that

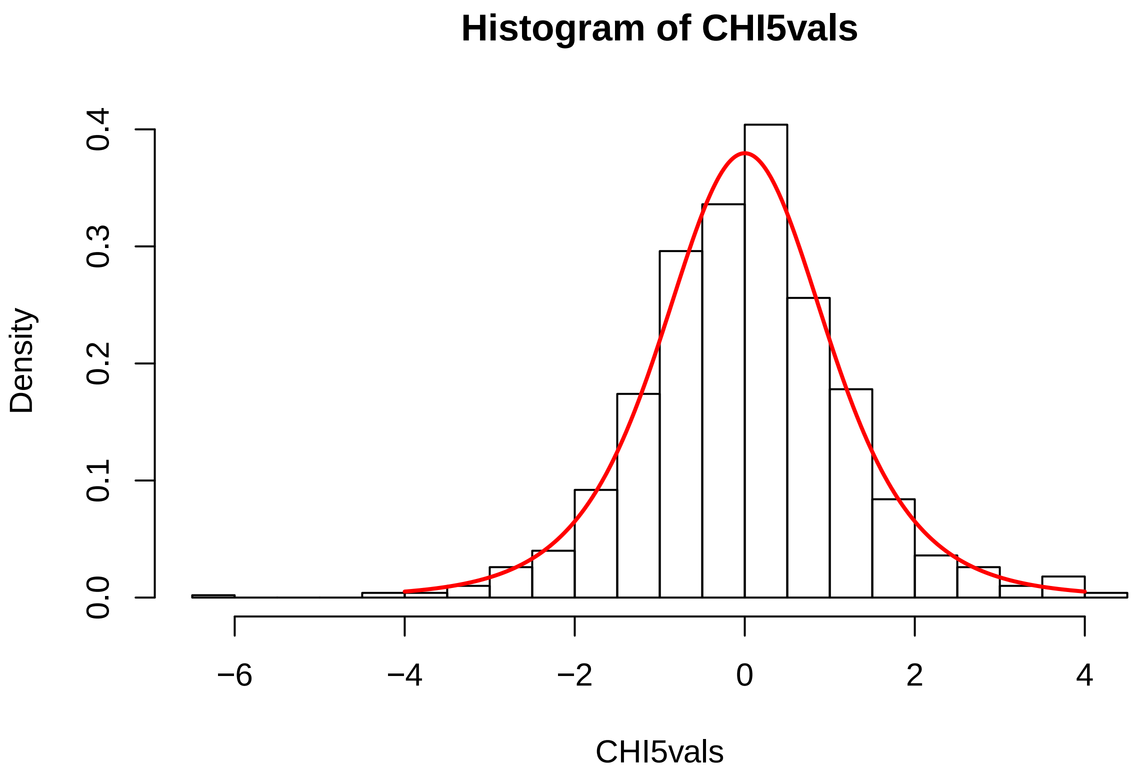
$$\frac{N(0,1)}{\sqrt{\chi_5^2/5}} \sim t_5$$

by setting the seed equal to 36 and generating a sample of size 1000 from a $N(0,1)$ distribution. Generate another sample of size 1000 from a χ_5^2 distribution (you can use the `rchisq()` function). Have R perform the appropriate calculation to arrive at the simulated sampling distribution. Create a density histogram of the results and superimpose a theoretical t_5 density.

Answer: We can see that the distibution looks to be normal shaped with no skewness. It is centered around 0 and sd looks to be less then one. With the t-distribution line with df=5 on top, the distrubtion seemes to verify the statment above.

```
set.seed(36)
# Set parameters
degfree <- 5
# Set number of observations
M <- 1000
# Initialize vectors of values
N1<- rnorm(M)
Z1 <- rchisq(M, degfree)
CHI5vals <- N1/sqrt((Z1)/5)
## Plot histogram
hist(CHI5vals, prob = T, nclass = "scott")
#endpoints
xvals  <- seq(-4,4,.01)
#Compute densities for the distribution
densities <- dt(xvals, degfree)
#Superimpose theoretical density over plot
```

```r
lines(xvals, densities, lwd = 2, col = 2)
```



4. Verify empirically that

$$\chi_3^2 + \chi_2^2 \sim \chi_5^2$$

by setting the seed equal to 79 and generating a sample of size 1000 from a χ_3^2 distribution. Generate another sample of size 1000 from a χ_2^2 distribution. Have R perform the appropriate calculation to arrive at the simulated sampling distribution. Create a density histogram of the results and superimpose a theoretical χ_5^2 density.

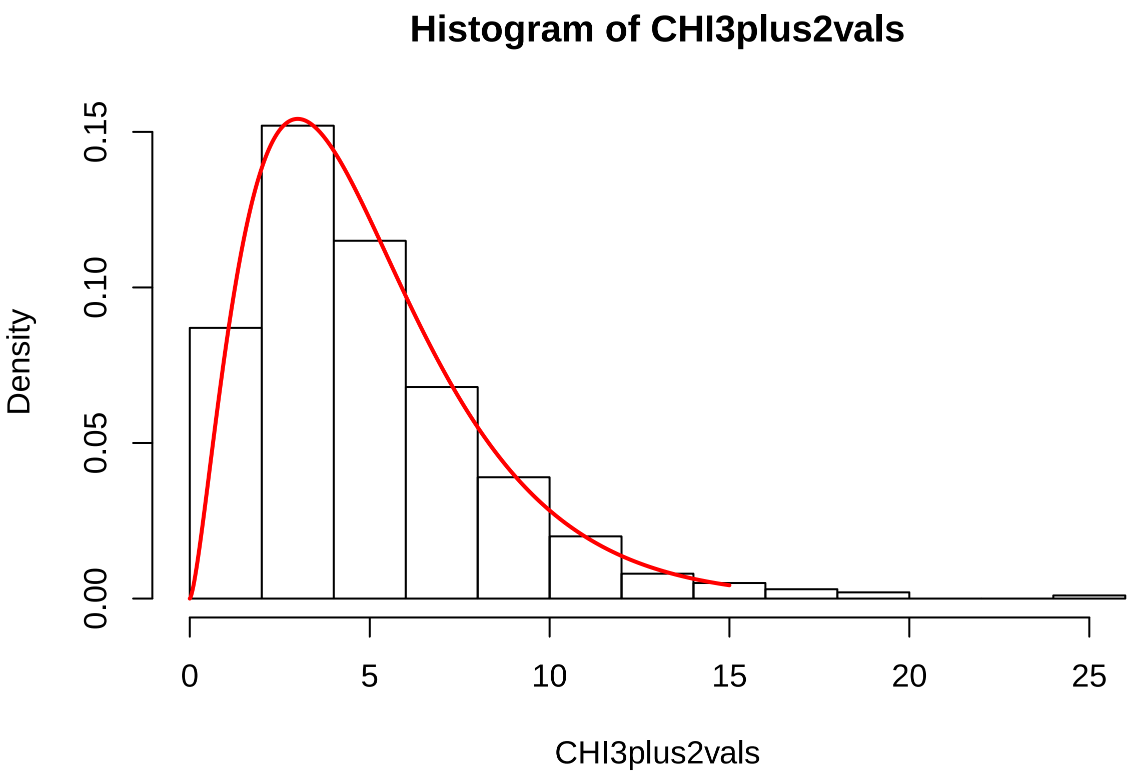
Answer: We can see the with the line on top that the statment above does seem to be true from our theoretical distribution. The center is around 5 and the shape is skewed left.

```r
set.seed(79)
# Set parameters
degfree <- 3
# Set number of observations
M <- 500
# Initialize vectors of values
Z1 <- rnorm(M)
Z2 <- rnorm(M)
Z3 <- rnorm(M)
CHI3vals <- Z1^2 + Z2^2 + Z3^2

X1 <- rnorm(M)
X2<- rnorm(M)
CHI2vals <- X1^2 + X2^2

CHI3plus2vals<- CHI3vals + CHI2vals
```

```r
## Plot histogram of simulated sampling distribution- be sure "prob=T" to allow superimposing
## density curve below
hist(CHI3plus2vals, prob = T, nclass = "scott")
# Might need some trial and error to get the endpoints
xvals  <- seq(0,15,.01)
#Compute densities for the chisquare distribution
densities <- dchisq(xvals, 5)
#Superimpose theoretical density over plot
lines(xvals, densities, lwd = 2, col = 2)
```



5. **Extra Credit Problem:** Set the seed equal to 10 and simulate 1000 random samples of size $n_X = 65$ from a $N(4, \sqrt{2})$ distribution and 1000 random samples of size $n_y = 90$ from a $N(5, \sqrt{3})$ distribution. Verify that the simulated statistic $\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$ actually follows an $F_{64,89}$ distribution by plotting a histogram with the appropriate density curve superimposed. (Hint: You'll need to learn about the `df()` command.)