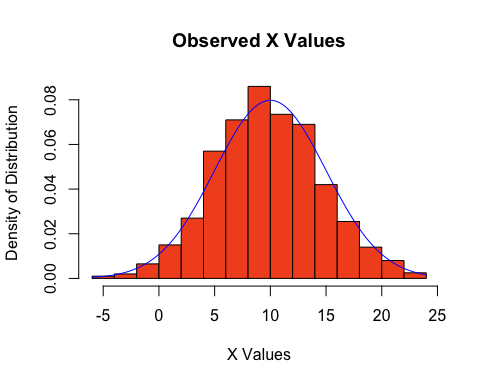
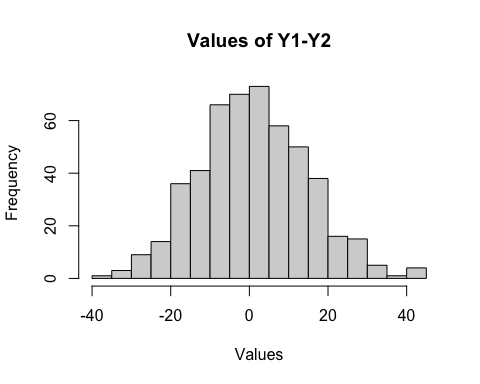
23603-lab-2.R

2023-03-05

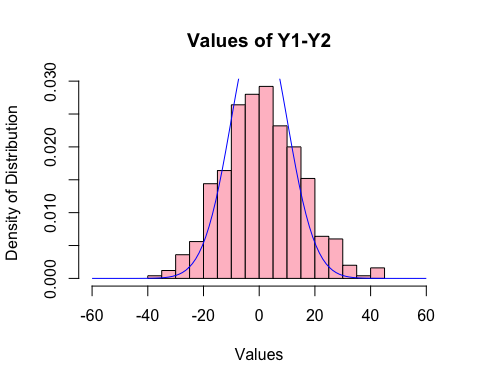
*#1*  
*#a*  
x = rnorm(1000,10,5)  
hist(x, prob = TRUE, breaks = 20, col = '#F35423', main = 'Observed X Values', xlab = 'X Values', ylab = 'Density of Distribution')  
curve(dnorm(x,10,5), add = TRUE, col = 'blue')



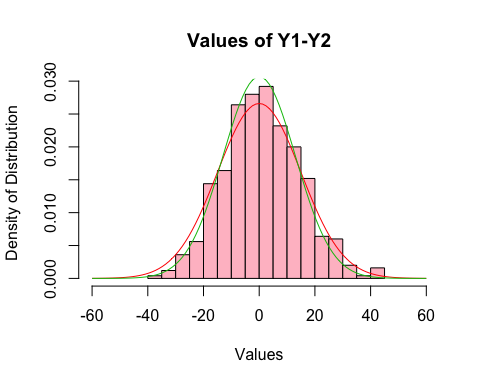
*#b*  
y1 = rnorm(500,10,5)  
y2 = rnorm(500,10,12)  
d = y2-y1  
hist(d, breaks = 20, main = 'Values of Y1-Y2', xlab = 'Values', ylab = 'Frequency')



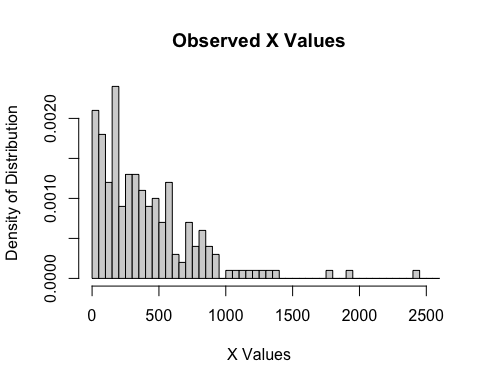
*#c*  
*#Histogram from part b is centered around 0, with a symmetric shape*  
*#As such, I would predict that d is also a normal distribution,*  
*#with mean 0, and standard deviation 10 as most of the distribution*  
*#seems to be within the range (-20,20), i.e. two standard deviations from the mean*  
  
hist(d, breaks = 20, prob = TRUE, col = 'pink', main = 'Values of Y1-Y2', xlab = 'Values', ylab = 'Density of Distribution', xlim = range(-60,60))  
curve(dnorm(x,0,10), add = TRUE, col = 'blue')



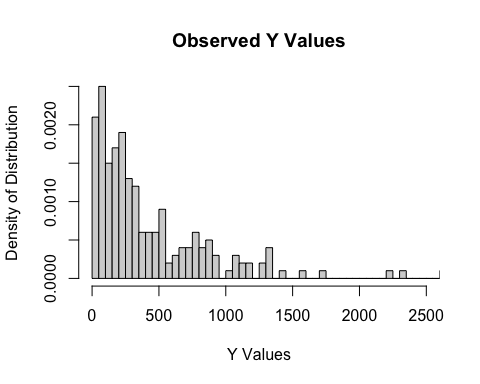
*#d*  
hist(d, breaks = 20, prob = TRUE, col = 'pink', main = 'Values of Y1-Y2', xlab = 'Values', ylab = 'Density of Distribution', xlim = range(-60,60))  
*#The curve of the PDF is too narrow, so try standard deviation of 15*  
curve(dnorm(x,0,15), add = TRUE, col = 'red')  
*#The curve is closer, but may not be tall enough*  
*#so try standard deviation 13*  
curve(dnorm(x,0,13), add = TRUE, col = '#00BF0F')



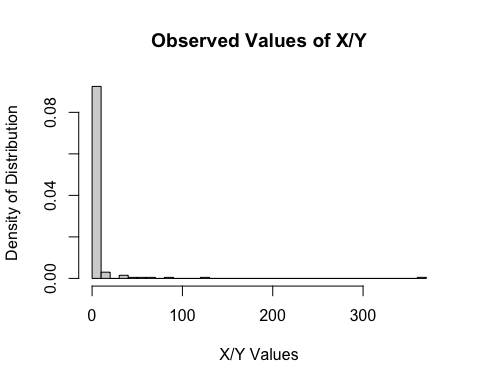
*#This curve seems to follow our simulation closely,*   
*#so predict that d follows a normal distribution*  
*#with mean 0 and S.D 13*  
  
*#2*  
*#a*  
n=200  
Xsamples = rep(0,n)  
Ysamples = rep(0,n)  
Rsamples = rep(0,n)  
**for**(i **in** 1:n)  
{  
 Xsamples[i] = rexp(1, rate = 1/400)  
 Ysamples[i] = rexp(1, rate = 1/400)  
 **if** (Ysamples[i]==0){  
 Rsamples[i] = 0  
 } **else**{  
 Rsamples[i] = Xsamples[i]/Ysamples[i]  
 }  
}  
hist(Xsamples, prob = TRUE, breaks = 50, xlim = range(0,2500), main = 'Observed X Values', xlab = 'X Values', ylab = 'Density of Distribution')



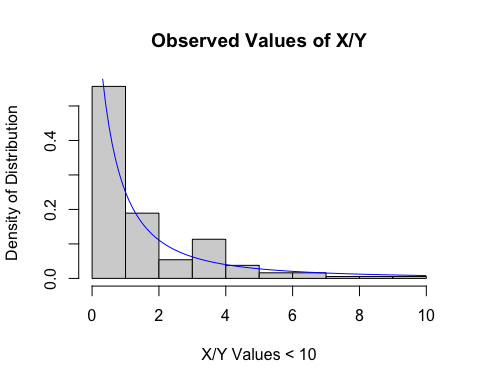
hist(Ysamples, prob = TRUE, breaks = 50, xlim = range(0,2500),main = 'Observed Y Values', xlab = 'Y Values', ylab = 'Density of Distribution')



hist(Rsamples, prob = TRUE, breaks = 50, main = 'Observed Values of X/Y', xlab = 'X/Y Values', ylab = 'Density of Distribution')



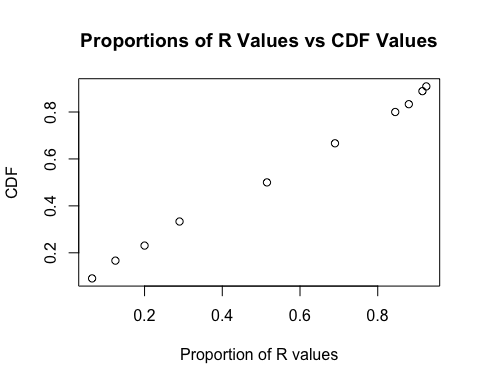
*#b*  
hist(Rsamples[Rsamples <10], prob = TRUE, breaks = 10, main = 'Observed Values of X/Y', xlab = 'X/Y Values < 10', ylab = 'Density of Distribution')  
curve(1/((1+x)^2), add = TRUE, col = 'blue')



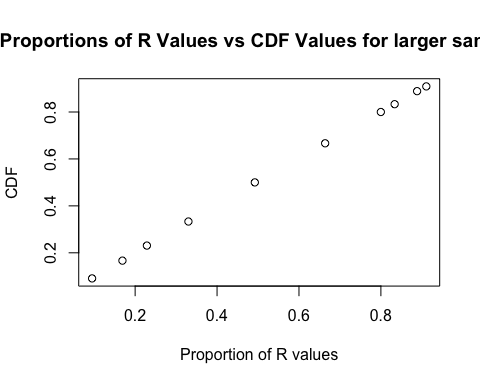
*#My plot does support the professor's claim*   
*#The plot of the proposed PDF would correspond to high densities*  
*#For small values of r, and low densities as r becomes larger*  
*#This can be seen on the plot, as the curve of the PDF*  
*#follows the height of the bars closely, with high density at low r values*  
*#And low density at high r values*  
  
*#c*  
*#CDF F(r) is the integral of the pdf from the lower bound of the*   
*#possible values up to r*  
*#So F(r) = integral from 0 to r (1/(1+r)^2 dr)*  
*#Giving F(r) = 1-1/(1+r) for r>= 0*  
Fr = rep(0,10)  
r = c(.1,.2,.3,.5,1,2,4,5,8,10)  
**for** (i **in** 1:10)  
{  
 Fr[i] = 1 - 1/(1+r[i])  
}  
Fr

## [1] 0.09090909 0.16666667 0.23076923 0.33333333 0.50000000 0.66666667  
## [7] 0.80000000 0.83333333 0.88888889 0.90909091

*#From above, we have F(0.1) = 0.0909, F(0.2) = 0.167,*  
*#F(0.3) = 0.231, F(0.5) = 0.333, F(1) = 0.5, F(2) = 0.667m*  
*#F(4) = 0.8, F(5) = 0.833, F(8) = 0.889, F(10) = 0.909*  
  
proportions = rep(0,10)  
**for** (i **in** 1:10)  
{  
 proportions[i] = length(Rsamples[Rsamples < r[i]]) / 200  
}  
plot(proportions, Fr, xlab = 'Proportion of R values', ylab = 'CDF', main = 'Proportions of R Values vs CDF Values')



*#The plot shows a linear correlation between the Values*  
*#of F(r) and the proportion of values in Rsamples less than*  
*#The corresponding values of r used in calculations of F(r)*  
*#The plot also shows that the values of the two variables are*   
*#Relatively close*  
  
*#d*  
n=10000  
Xsamples2 = rep(0,n)  
Ysamples2 = rep(0,n)  
Rsamples2 = rep(0,n)  
proportions2 = rep(0,10)  
**for**(i **in** 1:n)  
{  
 Xsamples2[i] = rexp(1, rate = 1/400)  
 Ysamples2[i] = rexp(1, rate = 1/400)  
 **if** (Ysamples2[i]==0){  
 Rsamples2[i] = 0  
 } **else**{  
 Rsamples2[i] = Xsamples2[i]/Ysamples2[i]  
 }  
**for** (i **in** 1:10)  
 {  
 proportions2[i] = length(Rsamples2[Rsamples2 < r[i]]) / n  
 }  
}  
plot(proportions2, Fr, xlab = 'Proportion of R values', ylab = 'CDF', main = 'Proportions of R Values vs CDF Values for larger sample')



*#Based off the larger sample, the values for F(r) and the proportions*  
*#are extremely close, leading me to believe that the professor's*  
*#claim is correct, as a larger sample led to closer values*