Analytics 511 HW 5

Please Note that all plots are located after the written up solutions before the R code. Due to the new Mac Software, I can't directly create PDF for tex.

Exercise 33

Thh ECDF plot looks kinda of like a log plot going from 0 to 60, with 0 to 60 representing the minute mark on the hour. The average value of W_a is around 10.

Exercise 34

The peak of the histogram is around the value of 1. The histogram also has a tail towards the right hand side

Exercise 35

I was interested in learning about the age rendering website that guessed the age of a person based on an image of a person's face shown in office hours. I was very surprised how the demo became so viral in Turkey. He really drives home the idea of presenting machine learning concepts to the layman. I'm surprised that it took 7 days to reach 50 million users via this method. However, it does make sense, since it was a simple idea. I was surprised that people would share the website when results were wrong. The tips for how to look younfer was particularily amusing, I plan to wear contacts and shave more often.

The tips for dealing with big data were particularity informative, such as how more effective building in the cloud really in.

Exercise 36

Based on the qqnorm plot, I'm inclined to say that Z has a normal distribution. The Mean calculated was approximately 0.739 and I calculated the standard deviation two different ways, the built in method gave an SD of 0.897021 whereas my calculated, modeling the actual formula, was 0.7818569.

Exercise 37

$E(Y|X) = E(aX + b + Z|X) = E(aX|X) + E(b) + E(Z) = aE(X|X) + b + 0 = aX + b$

Exercise 38

Part A

Since is an event the a point $x \in R$ is accepted, then $U < \frac{l(x)}{M}$, so X = x, thus $P(Y = x|A) = P(X = x)$

Part B

$P(A|Y = x) = \frac{l(x)}{M}$

Part C

$P(A) = \sum_{x \in R} P(A|Y = x) * P(Y = x) = \frac{\sum_{x \in R} l(x)}{M} * \frac{1}{N} = \frac{c}{MN}$

Part D

$P(X = x) = P(Y = x|A) = \frac{P(Y=x\&A)}{P(A)} = \frac{P(A|Y=x)*P(Y=x)}{P(A)} = \frac{\frac{l(x)}{M}*\frac{1}{N}}{\frac{C}{MN}} = \frac{l(x)}{c} = p(x)$

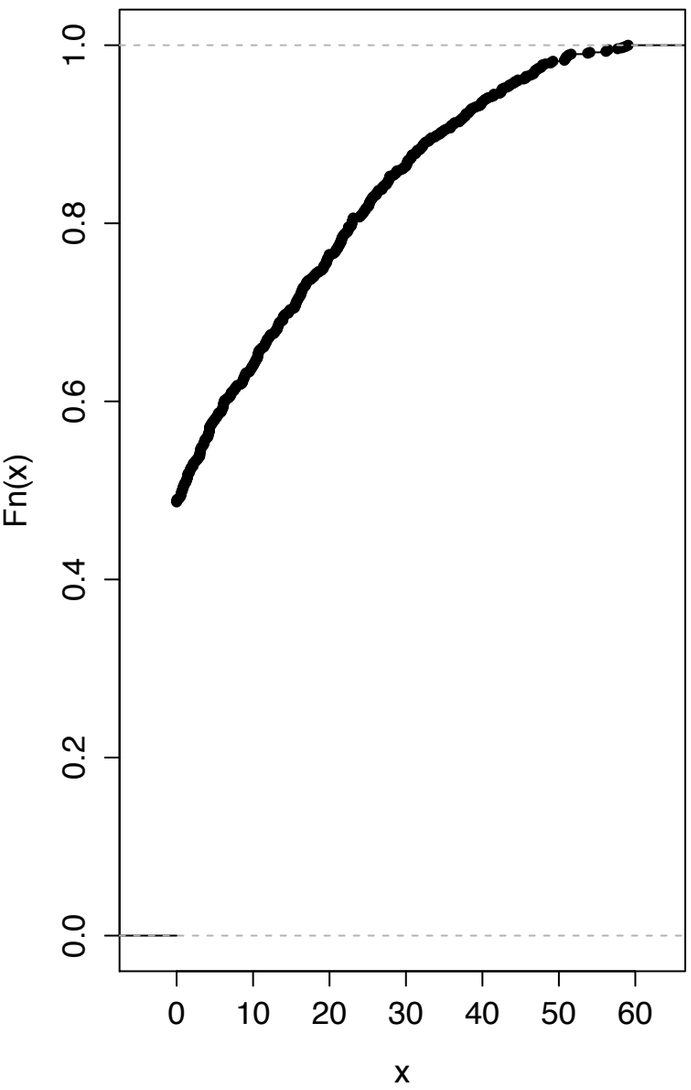
Exercise 40

Part 2

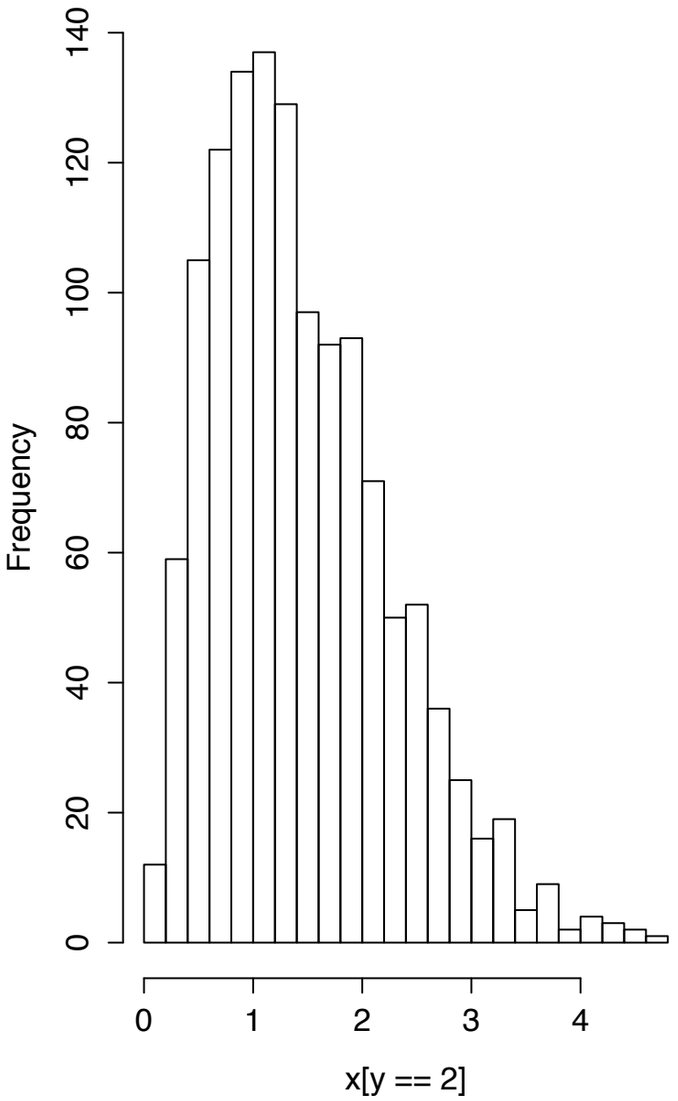
$P(X = 1|Y = y) = \frac{P(X=1\&Y=y)}{P(Y=y)} = \frac{P(Y=y|X=1)*P(X=1)}{P(Y=y|X=1)*P(X=1)+P(Y=y|X=2)*P(X=2)} = \frac{w_1 p_1(y)}{w_1 p_1(y)+w_2 p_2(y)}$

$$P(X=2|Y=y) = \frac{P(X=2\&Y=y)}{P(Y=y)} = \frac{P(Y=y|X=2)*P(X=2)}{P(Y=y|X=1)*P(X=1)+P(Y=y|X=2)*P(X=2)} = \frac{w_2 p_2(y)}{w_1 p_1(y)+w_2 p_2(y)}$$

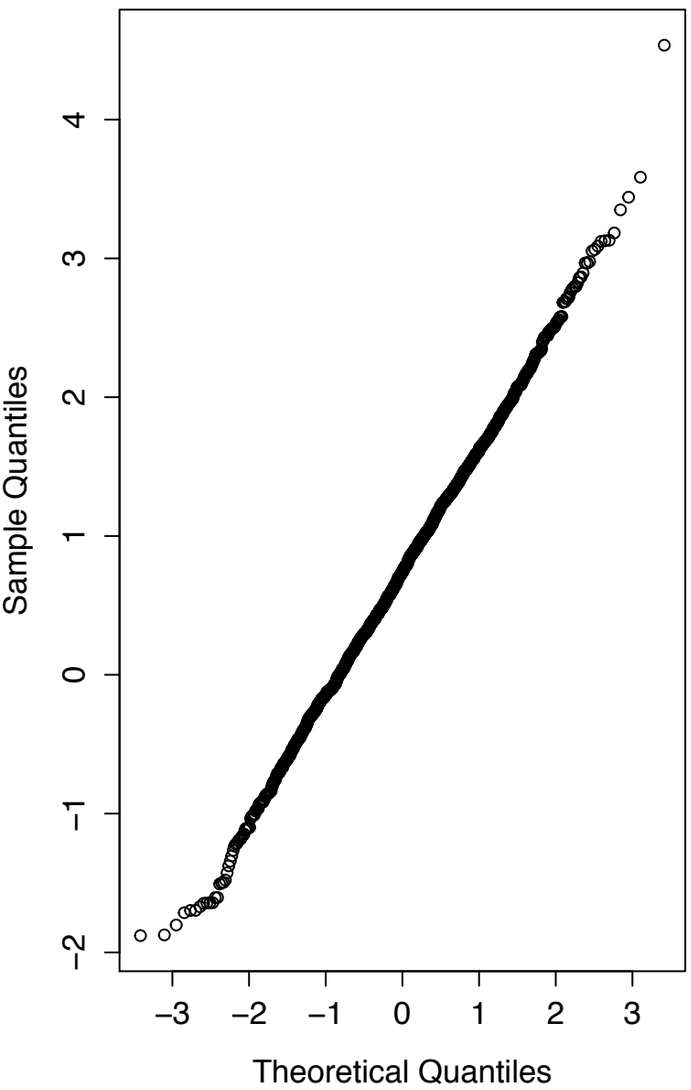
Exercise 33



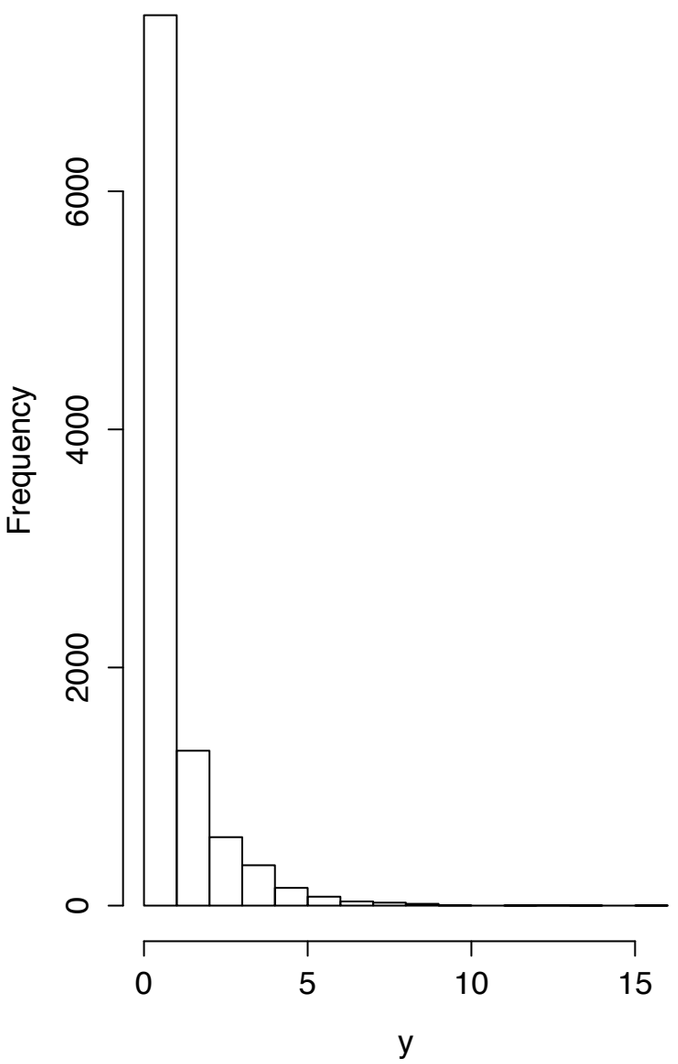
Exercise 34

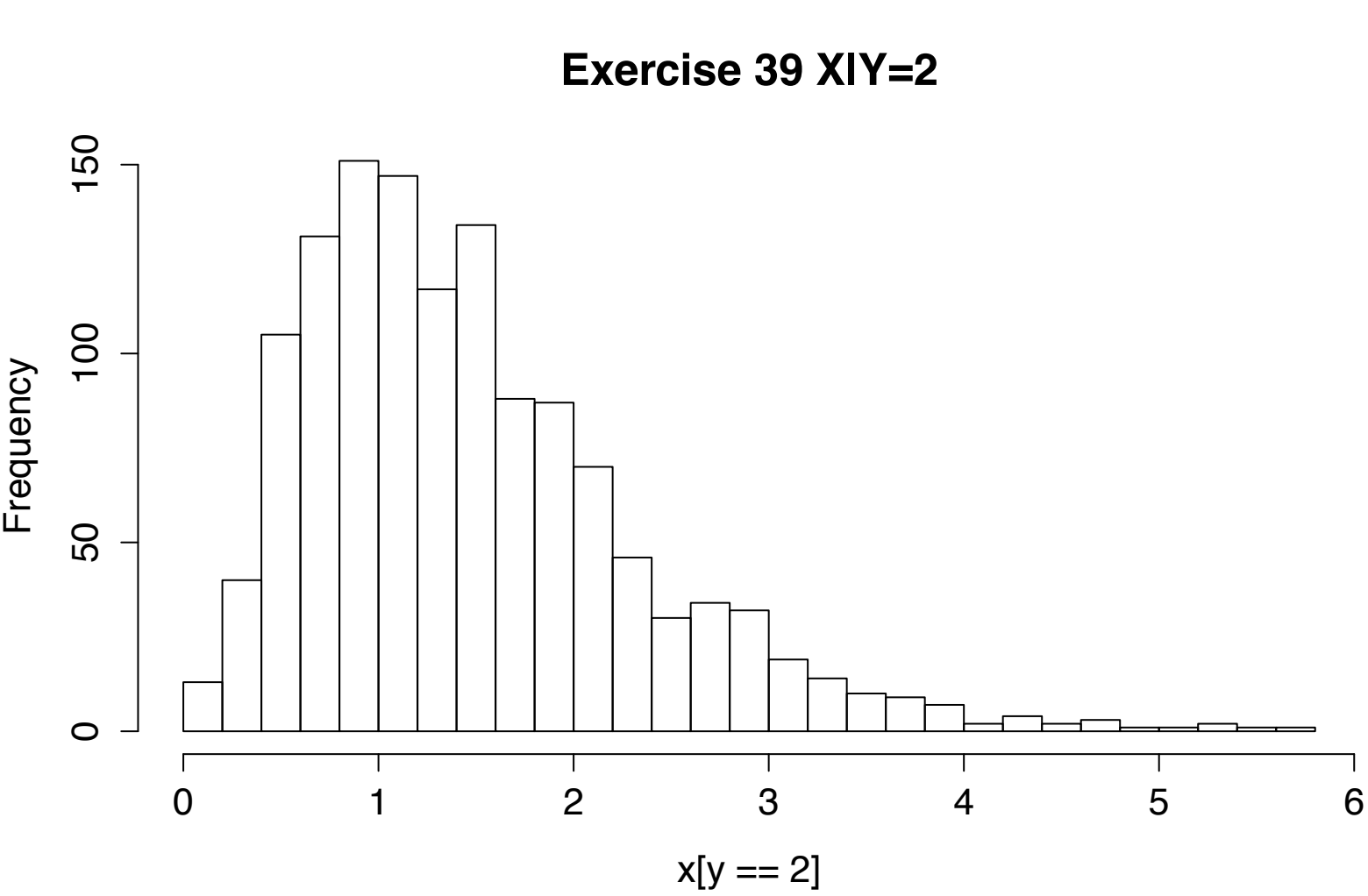


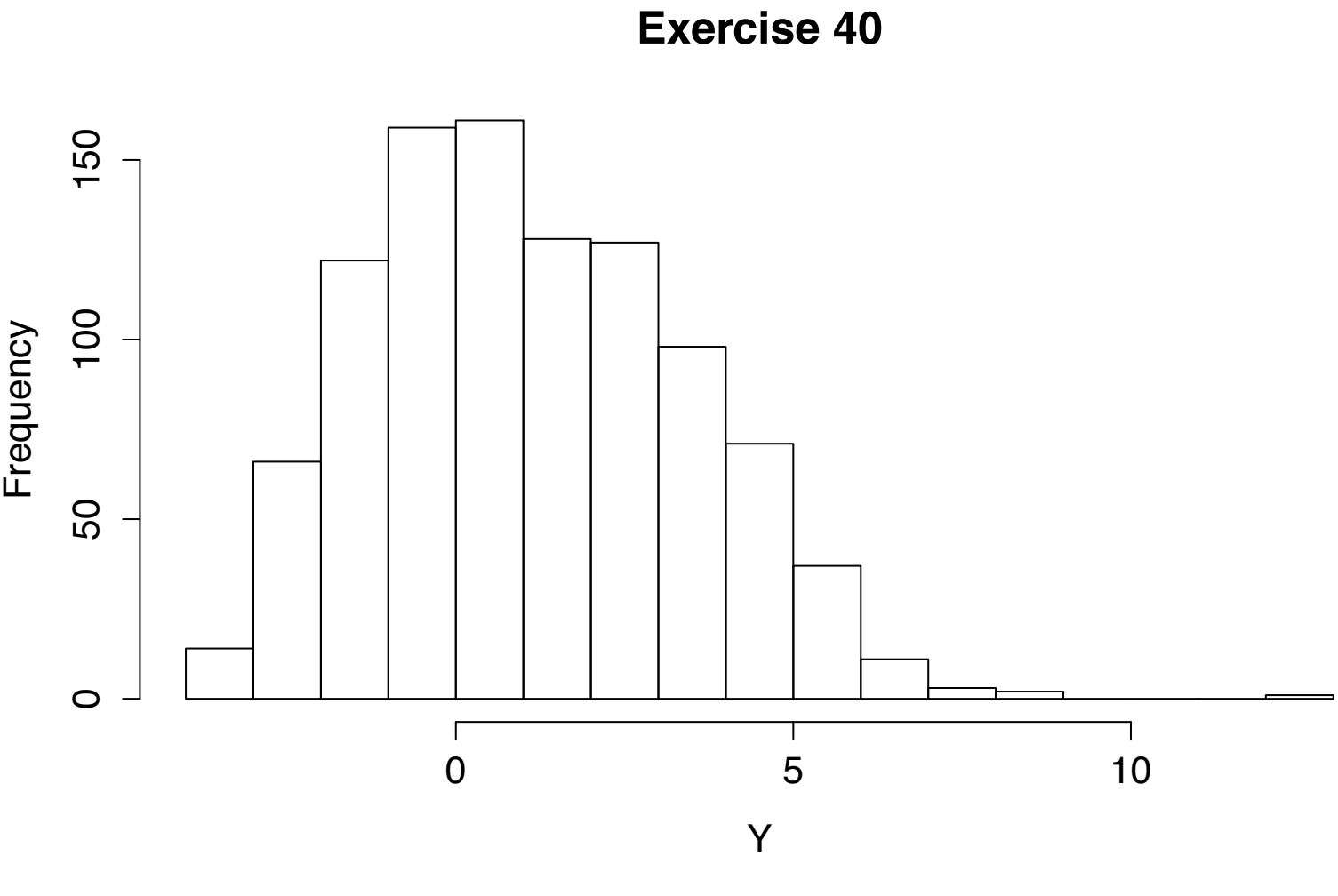
Exercise 36



Exercise 39 Y




Exercise 39 X|Y=2
Frequency
0
50
100
150
0
1
2
3
4
5
6
x[y == 2]


Exercise 40
Frequency
0
50
100
150
0
5
10
Y

```
#####Exercise 33
par(mfrow = c(2,2))
Xa = runif(1000, 0, 60)
Xb = runif(1000, 0, 60)
Wa = (Xb-Xa)
Wa[Wa<0] = 0
mean(Wa)
plot.ecdf(Wa, main = "Exercise 33")

#####Exercise 34
x = c()
y = c()
for(i in 1:10000){
x[i] = rexp(1,1)
y[i] = rpois(1,x[i])}
length(x[y == 2])
hist(x[y == 2], breaks = 20, main = "Exercise 34")
#####Exercise 36
#from ANALY511functions
mybivariate = function(n, mx = 0, my = 0,  sigx = 1, sigy = 1, rho =
1/2){
  Z1 <- rnorm(n)
  Z2 <- rnorm(n)
  X <- sigx*Z1 + mx
  Y <- rho*sigy*Z1 + sqrt(1-rho^2)*sigy*Z2
  z <- matrix(c(X,Y),ncol = 2)
  #plot(z[,1],z[,2],asp = 1)
  #grid(col = 1)
  return(z)
}
Exercise36 = mybivariate(10000)
cor(Exercise36[,1], Exercise36[,2])
Z= Exercise36[,2][Exercise36[,1]>=1]
mean(Z)
VarZ = mean(Z^2)-(mean(Z))
sqrt(VarZ)
sd(Z)
qqnorm(Z,main = "Exercise 36")


par(mfrow = c(2,1))
#####Exercise 39
x = c()
y = c()
for(i in 1:10000){
  x[i] = rexp(1,1)
  y[i] = rpois(1,x[i])}
hist(y , breaks = 20,main = "Exercise 39 Y")
hist(x[y == 2], breaks = 20,main = "Exercise 39 X|Y=2")
```

```
#####Exercise 40
###Part 1
w1 = 1/3
w2 = 2/3

Y = rep(0, times = 1000)

X = sample(c(1,2),1000,  prob = c(w1, w2), replace = T)
Y[X == 1] = rnorm(length(Y[X == 1]), -1, 1)
Y[X == 2] = rnorm(length(Y[X == 2]), 2, 2)


hist(Y, breaks = 20,main = "Exercise 40")
```