Data 556 - Homework 5

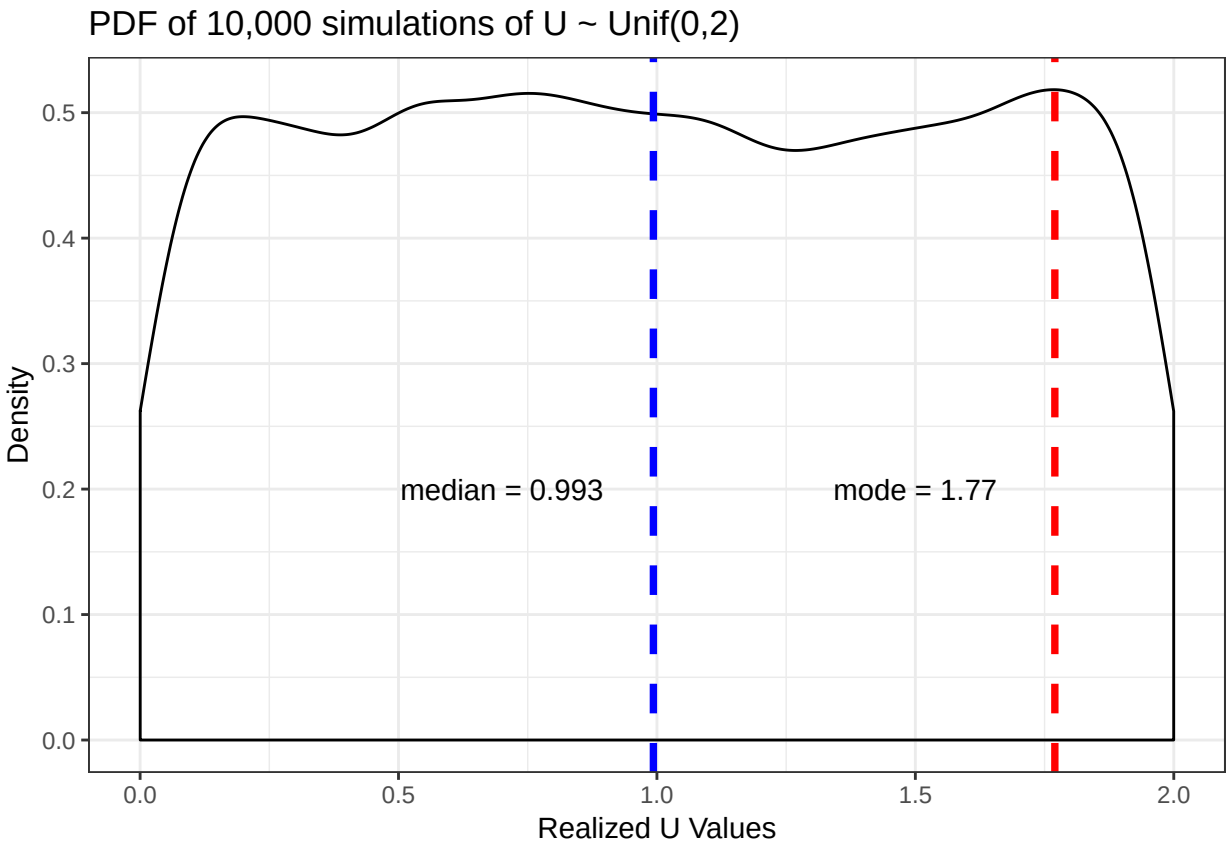
Will Wright

October 27, 2018

Problem 1a

Use simulations in R to numerically estimate the median and the mode of U for a = 0 and b = 2.

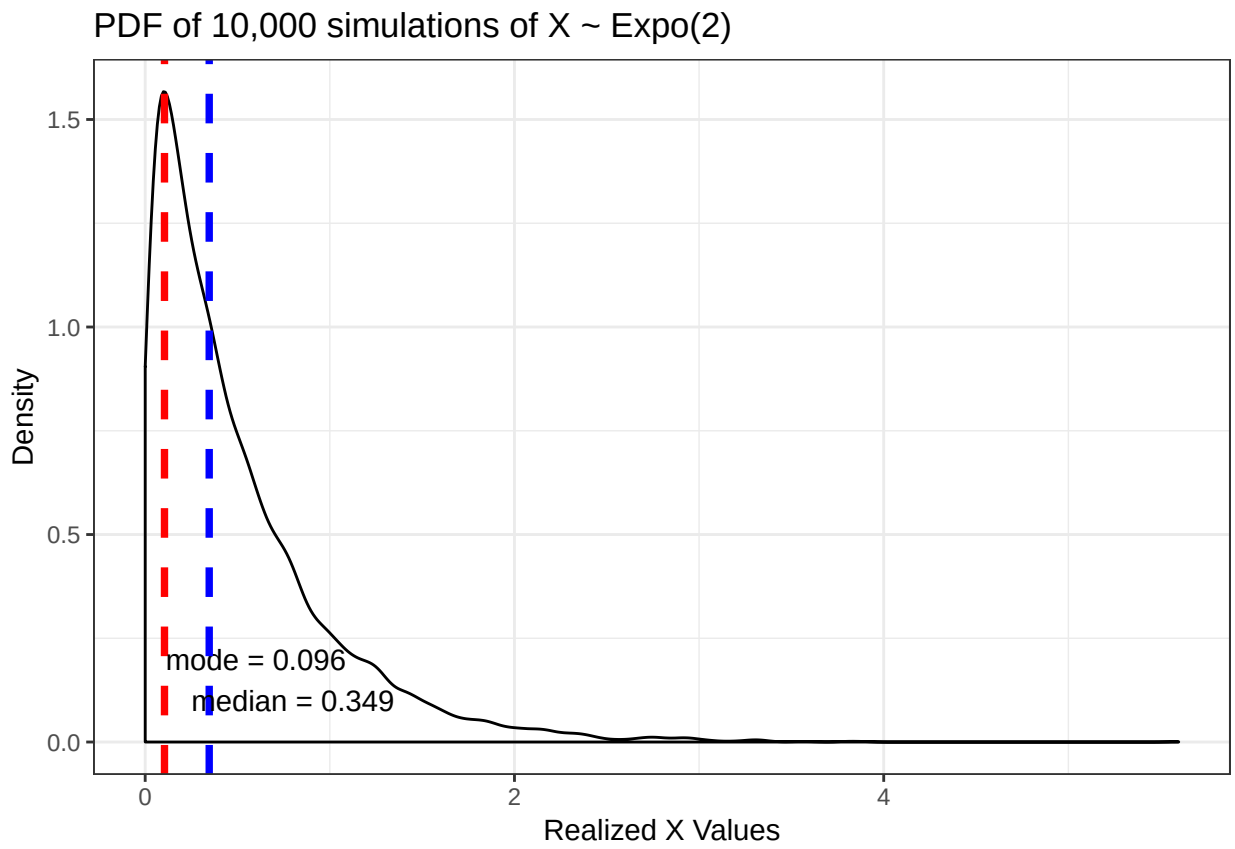
```
a <- 0
b <- 2
n <- 10000
U <- runif(n,a,b)
Umedian <- median(U)

# The mode is a value that has the greatest mass or density out of all values in the support of X.
Udensity <- density(U)
Umode <- Udensity$x[which(Udensity$y==max(Udensity$y))]
g <- ggplot(data.frame(U), aes(x = U))
g + geom_density() +
  geom_vline(xintercept = Umode, col = "red", linetype = "dashed", size = 1.3) +
  annotate("text", x = 1.5, y = 0.2, label = "mode = 1.77") +
  geom_vline(xintercept = Umedian, col = "blue", linetype = "dashed", size = 1.3) +
  annotate("text", x = 0.7, y = 0.2, label = "median = 0.993") +
  labs(title = "PDF of 10,000 simulations of U ~ Unif(0,2)",
       y = "Density",
       x = "Realized U Values") +
  theme_bw()
```



Problem 2a

Use simulations in R to numerically estimate the median and the mode of X ~ Expo (2)

```
X <- rexp(10000, rate = 2)
Xmedian <- median(X)

# The mode is a value that has the greatest mass or density out of all values in the support of X.
Xdensity <- density(X)
Xmode <- Xdensity$x[which(Xdensity$y==max(Xdensity$y))]
g <- ggplot(data.frame(X), aes(x = X))
g + geom_density() +
  geom_vline(xintercept = Xmode, col = "red", linetype = "dashed", size = 1.3) +
  annotate("text", x = 0.6, y = 0.2, label = "mode = 0.096") +
  geom_vline(xintercept = Xmedian, col = "blue", linetype = "dashed", size = 1.3) +
  annotate("text", x = 0.8, y = 0.1, label = "median = 0.349") +
  labs(title = "PDF of 10,000 simulations of X ~ Expo(2)",
       y = "Density",
       x = "Realized X Values") +
  theme_bw()
```

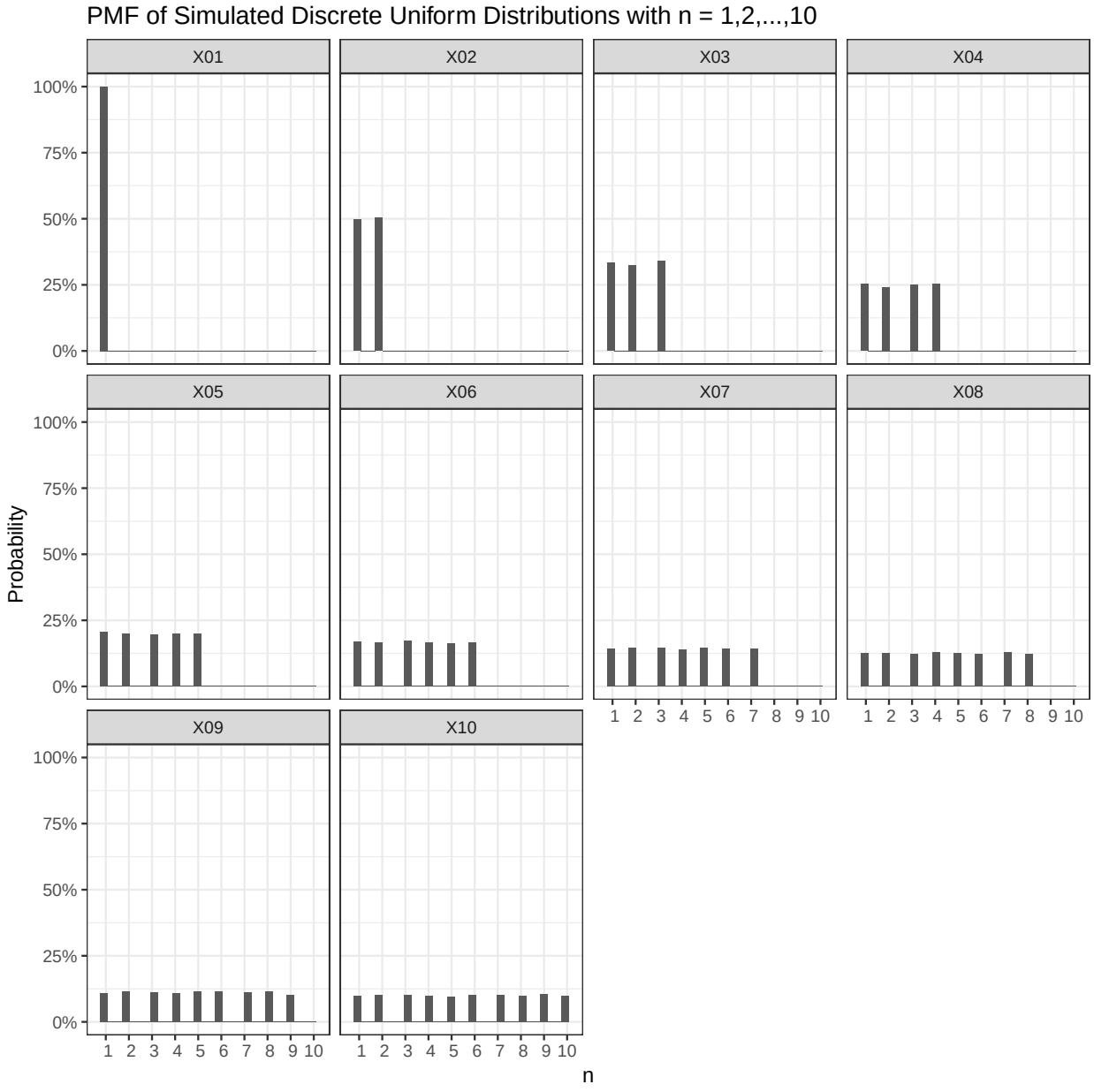


Problem 3a

Use simulations in R to numerically estimate all medians and all modes of X for n = 1, 2,..., 10.

```
n <- 1:10
# simulate draws from n = 1, 2,..., 10
simulator <- function(i) ceiling(runif(10000)*i) # function that converts continuous uniforms to discre
draws <- sapply(n, simulator)

Xmedians <- apply(draws, 2, median) # apply median() to the 10 columns for n = 1,2,...,10

# function to calculate modes
moder <- function(nSimulations) {
  freqTable <- data.frame(table(nSimulations))
  freqTable$nSimulations[which(freqTable$Freq==max(freqTable$Freq))]
}
Xmodes <- apply(draws, 2, moder)

# convert to long format for facet_wrap()
longDraws <- gather(data.frame(draws), n, draw)

# clean labels to display in the correct order by adding a 0 in front of X1-9
longDraws[,1][which(longDraws$n!="X10")] <- gsub("X(.*)", paste0("X0","\\1"),longDraws[,1])

g <- ggplot(longDraws, aes(x = draw))
g + geom_histogram(aes(y = ..count../10000)) +
```

```
  facet_wrap(n ~ .) +
  theme_bw() +
  labs(title = "PMF of Simulated Discrete Uniform Distributions with n = 1,2,...,10",
       x = "n",
       y = "Probability") +
  scale_x_discrete(limits = 1:10) +
  scale_y_continuous(labels = percent_format())
```




```
# build table of n, median, mode

summary <- data.frame(n = n, median = Xmedians, mode = as.numeric(Xmodes))
kable(summary) %>% kable_styling()
```

n	median	mode
1	1	1
2	2	2
3	2	3
4	3	4
5	3	1
6	3	3
7	4	3
8	4	4
9	5	5
10	6	9