Homework 5, DATA 556: Due Tuesday, 10/31/2018

Alexander Van Roijen

October 31, 2018

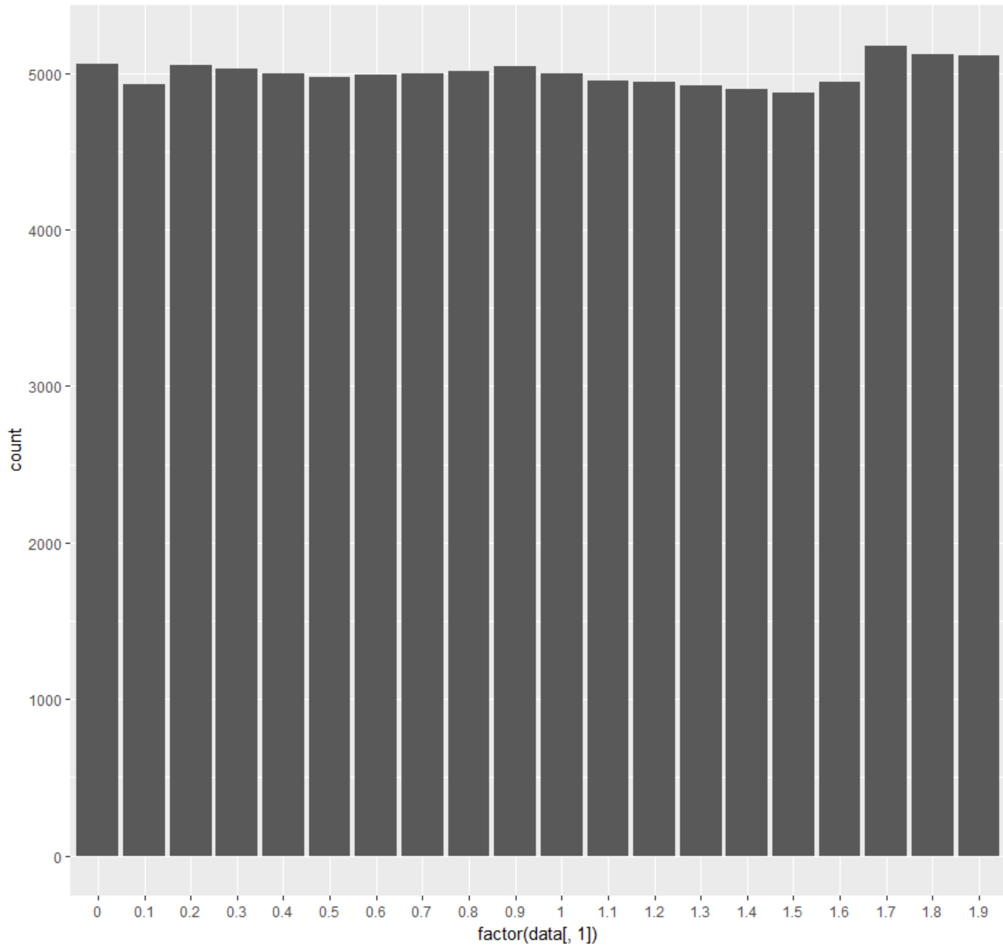
Please complete the following:

1. Problem 1 Let U ~Unif(a, b).

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

```
> binUnif= function(n,a,b)
+ {
+   resultsog = runif(n,a,b)
+   results = floor(resultsog*10)
+   data = data.frame(results)
+   vals = seq(a,b,.1)
+   p<-ggplot(data=data, aes(x=factor(data[,1]))) +geom_bar(stat="count") + scale_x_
+   p
+   #barplot(results,main="whatev",width=0.5)
+   print(median(resultsog))
+   getmode = table(resultsog)
+   print(which.max(getmode))
+   #print(forMode[c(1:10),])
+ }
> #this is 1a
> set.seed(123)
> n=1000000
> binUnif(n,0,2)
[1] 0.9983065
0.0349205480888486
17519
```

Figure 1: Mode roughly equivalent across all values of x



(b) Find the Median and Mode of $U \sim$ Unif(a,b)

$$\text{Median: } P(X \leq c) \geq \frac{1}{2} \& P(X \geq c) \geq \frac{1}{2} \tag{1}$$

$$=> P(X \leq x) = \int_a^c f(x)dx = \int_a^c \frac{1}{b-a}dx = \frac{c}{b-a} - \frac{a}{b-a} = \frac{c-a}{b-a} \tag{2}$$

$$=> \frac{c-a}{b-a} \geq \frac{1}{2} => c \geq \frac{b-a}{2} + \frac{2a}{2} => c \geq \frac{b+a}{2} \tag{3}$$

$$\text{coming from the other inequality, we similarly get } c \leq \frac{b+a}{2} \tag{4}$$

$$\text{Note that this unique mode will not apply in the discrete case of the uniform distribution} \tag{5}$$

$$\text{Mode:} \tag{6}$$

$$\text{Want: } f(c) \geq f(x) \forall x \tag{7}$$

$$f(c) = f(x) = \frac{1}{b-a} \forall x \text{ by def of DUnif} \tag{8}$$

$$=> \text{Mode of X is entire support of X} = [a...b] \tag{9}$$

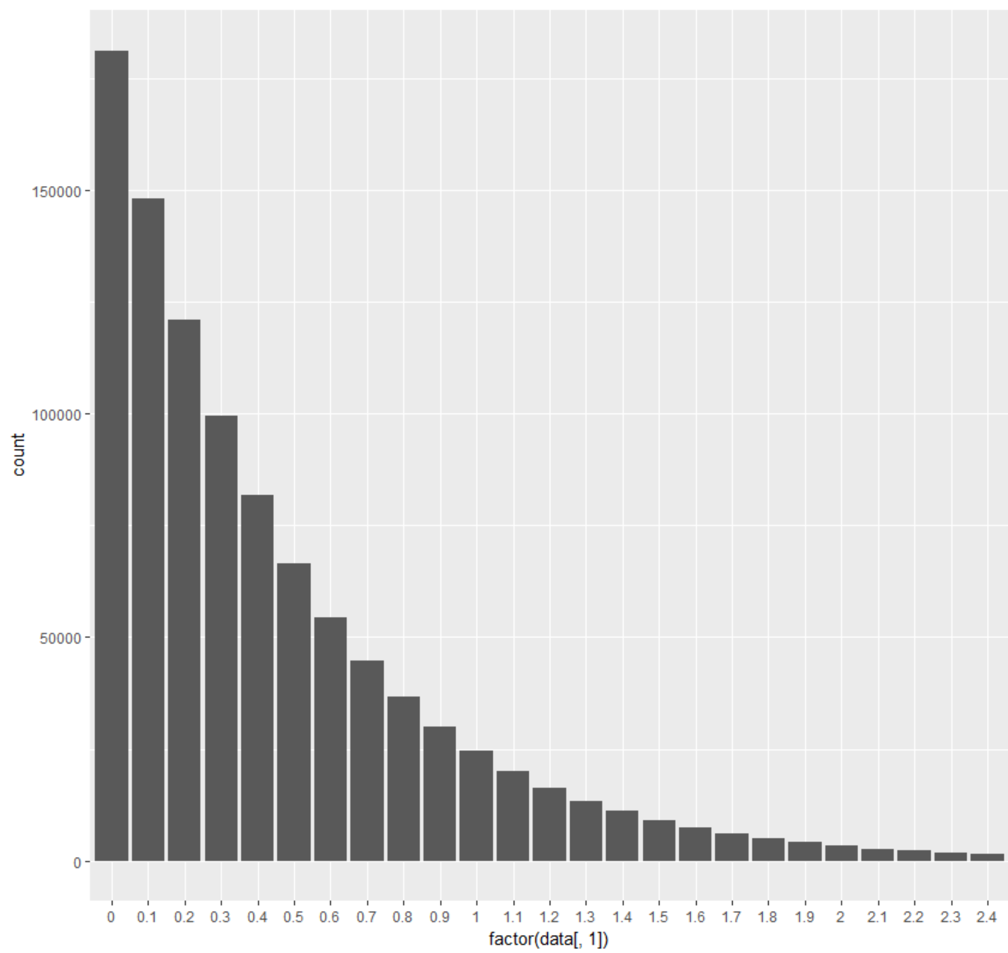
2. Let $X \sim$ Expo(λ)

(a) Use R to simulate median and mode of Expo(2)

```
expoMedMode = function(n,rate)
+ {
+   resultsog = rexp(n,rate=rate)
+   results = floor(resultsog*10)
+   data = data.frame(results)
+   data = data.frame(data[data[,1]<25,])
+   vals = seq(0,2.5,.1)
+   p<-ggplot(data=data, aes(x=factor(data[,1]))) +geom_bar(stat="count") + scale_x_
+   print(p)
+   #barplot(results,main="whatev",width=0.5)
+   print(median(resultsog))
+   getmode = table(resultsog)
```

```
+   print(which.max(getmode))
+ }
> #this is 2a
> set.seed(123)
> n=1000000
> expoMedMode(n,2)
[1] 0.3467215
0.00334986066445708
6662
```

Figure 2: Mode of an exponential with rate of 2 around 0 and median around .3467



This overall makes sense, as we will show below, 0 is the mode of the exponential distribution. This is confirmed as well by the barplot shown above.

(b) Find the Median and Mode of $X \sim \text{Expo}(\lambda)$

$$\text{Median: } P(X \leq c) \geq \frac{1}{2} \& P(X \geq c) \geq \frac{1}{2} \tag{10}$$

$$P(X \leq c) = \int_0^c f(x)dx = \int_0^c \lambda e^{-\lambda x} dx = 1 - e^{-\lambda c} \geq \frac{1}{2} => c \geq \frac{\ln(2)}{\lambda} \tag{11}$$

$$\text{Coming from the other side we get} \tag{12}$$

$$P(X \geq c) = \int_c^\infty f(x)dx = \int_c^\infty \lambda e^{-\lambda x} dx = e^{-\lambda c} \geq \frac{1}{2} => c \leq \frac{\ln(2)}{\lambda} \tag{13}$$

$$\text{So again, we have our unique median. This time as a function of lambda which makes sense} \tag{14}$$

$$\text{Mode:} \tag{15}$$

$$\text{Want: } f(c) \geq f(x) \forall x \tag{16}$$

$$\text{Note that: } f(x) = \lambda e^{-\lambda x} => f'(x) = -(\lambda^2) e^{-\lambda x} \tag{17}$$

$$\text{This along with } f(0) = \lambda \text{ explains the whole story} \tag{18}$$

$$\text{a dereasing value of f(x) that starts at } x = 0 \text{ indicates our mode lies at } x = 0 / w f(x) = \lambda \tag{19}$$

3. Let X be Discrete Uniform on 1,2,3,4,5...n .

(a) Use simulations in R to numerically estimate all medians and all modes of X for n = 1,2,3...10.

```
> #this is 3a
> set.seed(433)
> counter = 1
> n=10
> size = 1000
> while(counter <= n)
+ {
+   binUnif(size,1,counter,1,1)
+   counter = counter + 1
+ }
[1] "From 1 to 1"
```

```
[1] 1

1

1

[1] "From 1 to 2"

[1] 1.505739

1.0011387350969

1

[1] "From 1 to 3"

[1] 2.025223

1.00209179287776

1

[1] "From 1 to 4"

[1] 2.502714

1.00117637915537

1

[1] "From 1 to 5"

[1] 2.957827

1.00255306344479

1

[1] "From 1 to 6"

[1] 3.368198

1.00341863720678

1

[1] "From 1 to 7"

[1] 4.004081

1.0143808578141

1

[1] "From 1 to 8"

[1] 4.470458

1.00075664301403

1
```

```
[1] "From 1 to 9"

[1] 5.058819

1.0038467105478

1

[1] "From 1 to 10"

[1] 5.422321

1.01815693522803

1
```

Figure 3: Mode across X demonstrated from Unif(1,1)

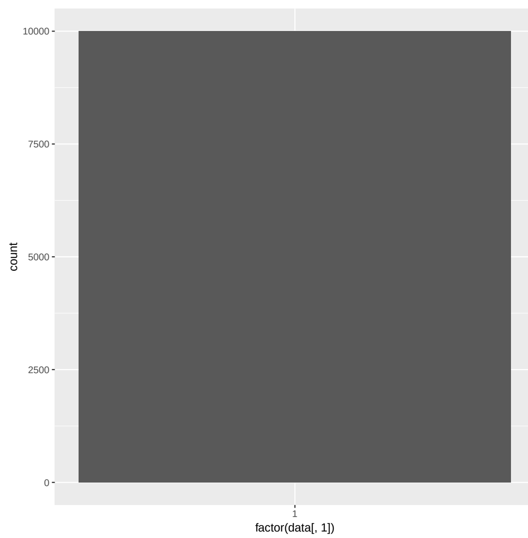


Figure 4: Mode across X demonstrated from Unif(1,2)

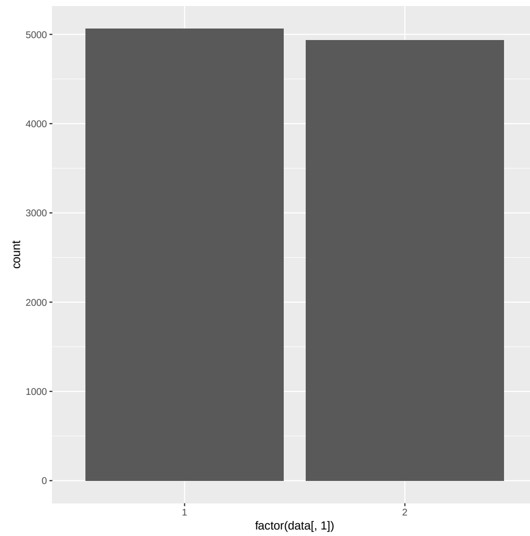


Figure 5: Mode across X demonstrated from Unif(1,3)

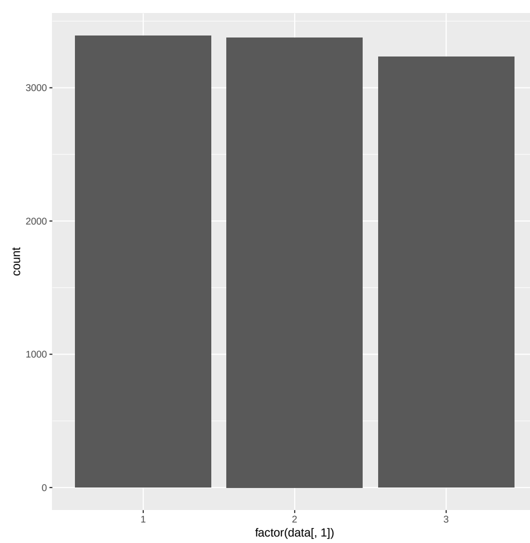


Figure 6: Mode across X demonstrated from Unif(1,4)

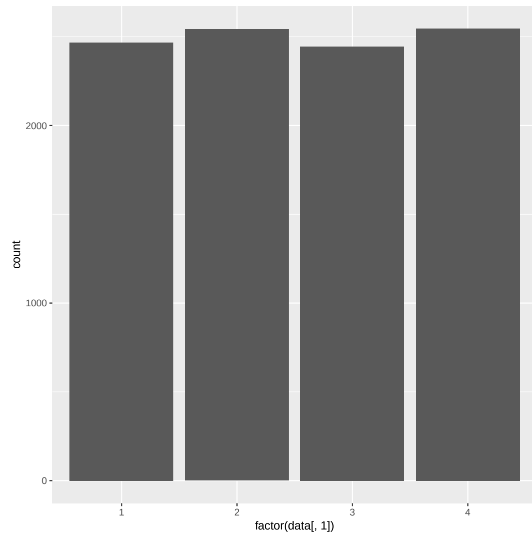


Figure 7: Mode across X demonstrated from Unif(1,5)

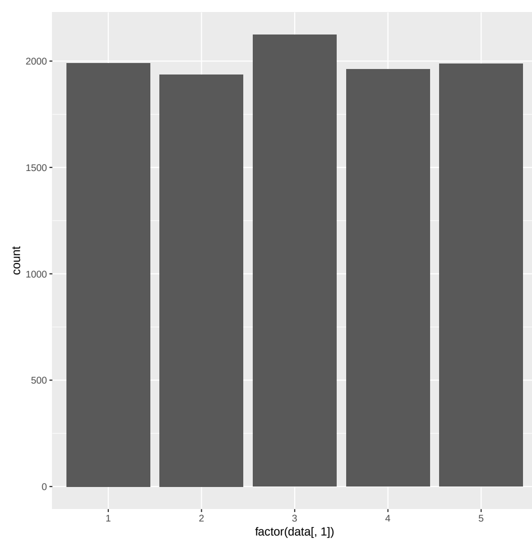


Figure 8: Mode across X demonstrated from Unif(1,6)

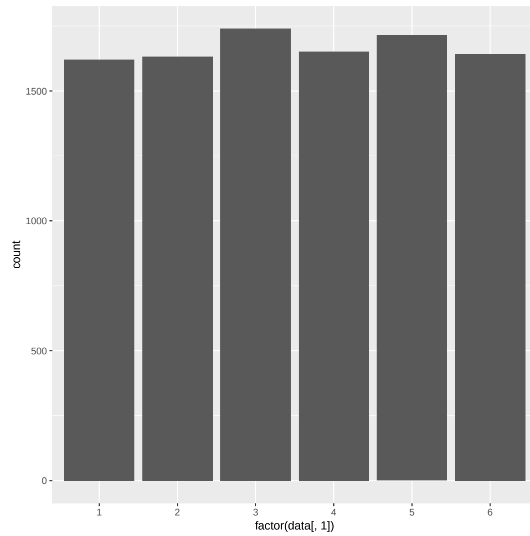


Figure 9: Mode across X demonstrated from Unif(1,7)

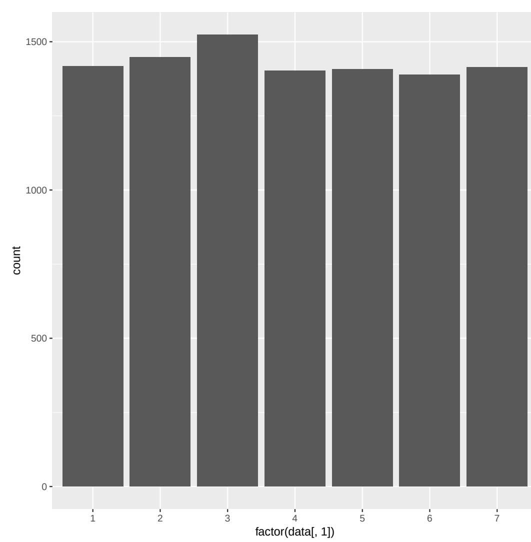


Figure 10: Mode across X demonstrated from Unif(1,8)

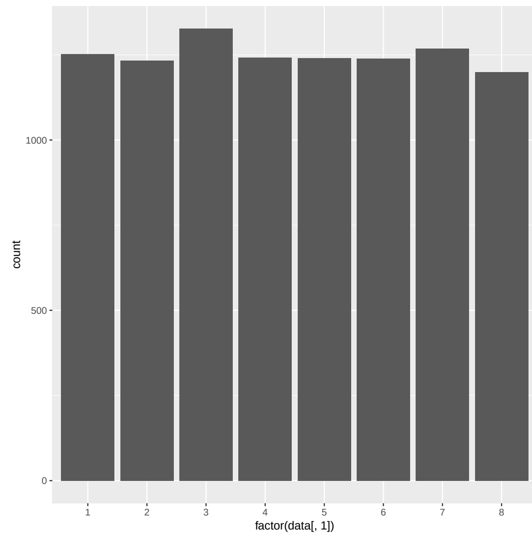


Figure 11: Mode across X demonstrated from Unif(1,9)

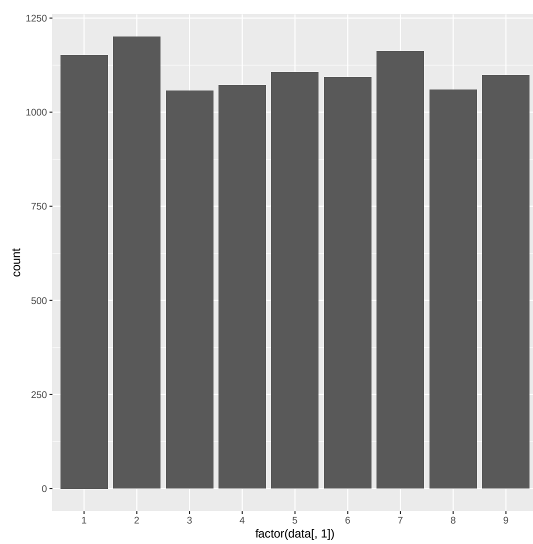
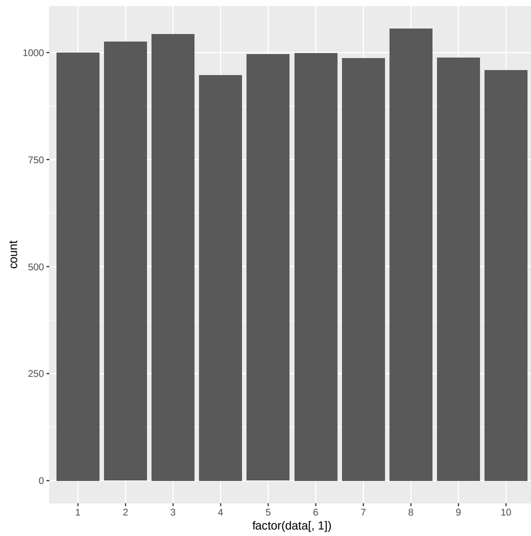


Figure 12: Mode across X demonstrated from Unif(1,10)



Again, we expect some variability here, but with large enough N we can understand that the mode of a discrete uniform distribution is over all discrete support of the RV

(b) Find All medians and modes of X

Note the mode is trivial, as it is a similar case to 1.

$$\text{Want } P(X = c) \geq P(X = x) \forall x \in 1, 2, ...n \tag{20}$$

$$=> \frac{1}{n - 1 + 1} \geq \frac{1}{n - 1 + 1} \text{ by def of discrete Uniform} \tag{21}$$

$$=> c = x \forall x \in 1, 2, 3, ...n \text{ which is almost the same as 1} \tag{22}$$

$$\text{Notable difference is that in the discrete case, c can only take discrete values} \tag{23}$$

$$\text{Now the median} \tag{24}$$

$$\text{Want } P(X \leq x) \geq \frac{1}{2} \& P(X \geq x) \geq \frac{1}{2} \tag{25}$$

$$P(X \leq x) = \sum_{i=1}^{x} \frac{1}{n} = \frac{x}{n} \geq \frac{1}{2} => x \geq \frac{n}{2} \tag{26}$$

$$\text{However, we can observe the patterns noted in the graphs below} \tag{27}$$

$$\text{If n is odd, the previous conclusion is the only solution as} \tag{28}$$

$$\text{if you go above or below } \frac{n}{2} \text{ you lose the probability @ } x = \frac{n}{2} \tag{29}$$

$$\text{This is true as we have a jump exactly at } \frac{n}{2} \tag{30}$$

$$\text{In the case n is even, you have some wiggle room} \tag{31}$$

$$\text{There is no jump at } \frac{n}{2} \tag{32}$$

$$\text{In fact, the next jump is at } \frac{n}{2} + 1 \tag{33}$$

$$\text{This means we have medians from } [\frac{n}{2}, \frac{n}{2} + 1] \tag{34}$$

$$\text{One last important factor to note, is that this result relies heavily on} \tag{35}$$

$$\text{how we define the median and our environment} \tag{36}$$

$$\text{In general we have} \tag{37}$$

$$Median(X) = \begin{cases} \frac{n}{2} & n = \text{odd} \\ [\frac{n}{2}, \frac{n}{2} + 1] & n = \text{even} \end{cases} \tag{38}$$

Figure 13: (n=5)Note the jumps taking place on the discrete values

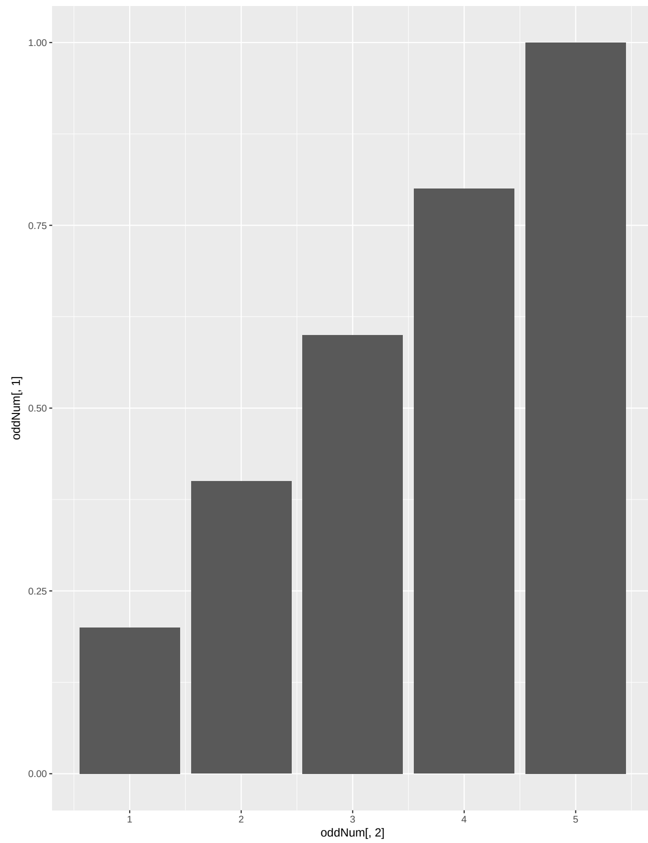
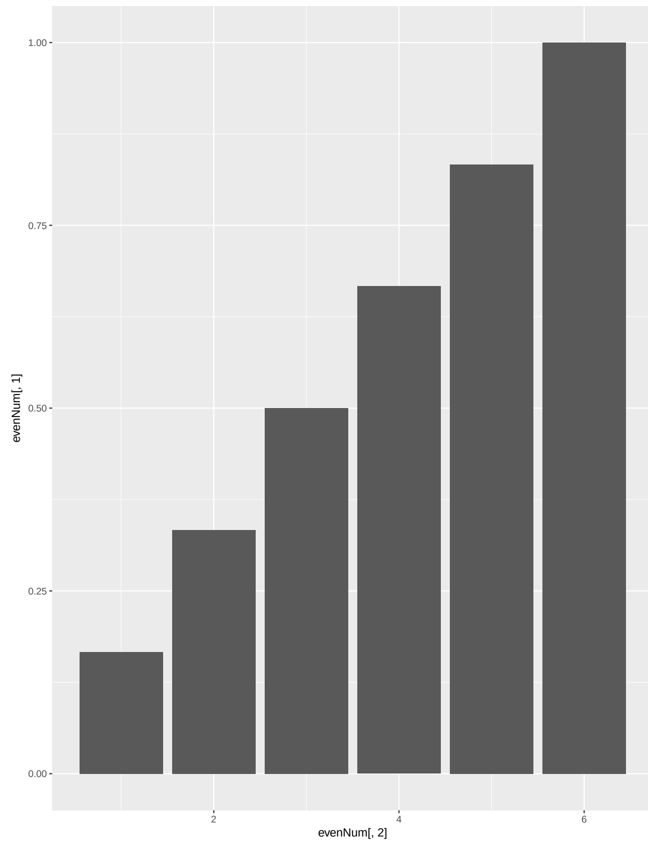


Figure 14: (n=6)These jump patterns hold for all discrete n



4. A distribution is called symmetric unimodal if it is symmetric (about some point) and has a unique mode. For example, any Normal distribution is symmetric unimodal. Let X have a continuous symmetric unimodal distribution for which the mean exists. Show that the

mean, median, and mode of X are all equal.

$$\text{First note that from our notes, we know the following} \tag{39}$$

$$1\text{: The mean and median are on in the same by definition} \tag{40}$$

$$2\text{: } f(x) = f(2\mu - x)\forall x \& P(X \geq X + \mu) = P(X \leq X - \mu) \tag{41}$$

$$\text{Where } \mu \text{ is the mean of our distribution, or central point} \tag{42}$$

$$\text{All that is left to show is that the mode is equivalent to either our median or mean} \tag{43}$$

$$\int_{-\infty}^{\mu-x} f(x)dx = \int_{\mu+x}^{\infty} f(x)dx = \int_{-\infty}^{\mu-x} f(2\mu-x)dx = \int_{\mu+x}^{\infty} f(2\mu-x)dx \tag{44}$$

$$\int_{-\infty}^{\mu} f(x)dx = \frac{1}{2} = \int_{\mu}^{\infty} f(x)dx \tag{45}$$

$$\text{By the defintion of the mode, we have some c } s.t. f(c) \geq f(x)\forall x \tag{46}$$

$$\text{Lets proove via contradiction. If } c \neq \mu => c > \mu \text{ or } c < \mu \tag{47}$$

$$=> f(c) = f(2\mu - c) \text{ with } c \neq 2\mu - c \text{ by our previous statement} \tag{48}$$

$$=> \exists c_1\, s.t. f(c_1) \geq f(x)\forall x \& c_2\, s.t. f(c_2) \geq f(x)\forall x \text{ with } c_1 \neq c_2 \tag{49}$$

$$\text{This is a contradiction! as now we have two unique modes} \tag{50}$$

$$\text{which violates the property of a unimodal distribution}\Box \tag{51}$$

5. Let $W = X^2 + Y^2$, with X, Y i.i.d. N(0, 1). You can assume you know that the MGF of X^2 is $(1-2t)^{\frac{-1}{2}}$ for $t < \frac{1}{2}$. Find the MGF of W.

$$\text{We know for independent X, Y random variables that } M_{X+Y}(t) = M_X(t) * M_Y(t) \tag{52}$$

$$\text{We know that } M_{X^2}(t) = M_{Y^2}(t) \text{ and lets rewrite } X' = X^2 andY' = Y^2 \tag{53}$$

$$\text{since } X' \& Y' \text{ independet } => M_{X'+Y'} = M_{X'} * M_{Y'} = M_{X^2} * M_{Y^2} = ((1-2t)^{\frac{-1}{2}})^2 = \frac{1}{1-2t} \tag{54}$$

6. Let $X \sim Expo(\lambda)$. You can assume you know that $\lambda X \sim$ Expo(1), and that the nth

moment of an Expo(1) random variable is n!. Find the skewness of X.

$$\text{Def of Skewness(X): } E[(\frac{X-\mu}{\sigma})^3] \tag{55}$$

$$= E[(\frac{X-E[X]}{\sqrt{\text{Var}(X)}})^3] = E[(\frac{X-\frac{1}{\lambda}}{\sqrt{\frac{1}{\lambda^2}}})^3] = E[(\lambda(X-\frac{1}{\lambda}))^3] = E[(\lambda X-1)^3] \tag{56}$$

$$= E[(\lambda X)^3 - 3(\lambda X)^2 + 3(\lambda X) - 1] = E[(\lambda X)^3] - 3E[(\lambda X)^2] + 3E[(\lambda X)] - 1 \tag{57}$$

$$\text{With } E[(\lambda X)^3] = 3!, E[(\lambda X)^2] = 2!, E[(\lambda X)] = 1! \text{ from problem statement} \tag{58}$$

$$=> E[(\frac{X-E[X]}{\sqrt{\text{Var}(X)}})^3] = 3! - 3*2 + 3*1 - 1 = 2 \tag{59}$$

7. Let $X_1, ...X_n$ be i.i.d. with mean μ, variance σ^2, and MGF M. Let $\bar{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$ and $Z_n = \sqrt{n}\frac{\bar{X_n}-\mu}{\sigma}$

(a) Show that Z_n has mean 0 and variance 1

$$E[\bar{X_n}] = E[\frac{1}{n}\sum_{i=1}^n X_i] = \frac{1}{n}\sum_{i=1}^n E[X_i] = \frac{1}{n} * n\mu = \mu \tag{60}$$

$$=> E[Z_n] = E[\sqrt{n}\frac{\bar{X_n}-\mu}{\sigma}] = \sqrt{n}\frac{E[\bar{X_n}]-\mu}{\sigma} = 0 \tag{61}$$

$$\text{Var}(\bar{X_n}) = \text{Var}(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n^2}\sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n} \tag{62}$$

$$=> \text{Var}[Z_n] = \text{Var}(\sqrt{n}\frac{\bar{X_n}-\mu}{\sigma}) = \frac{n}{\sigma^2}\text{Var}(\bar{X_n}-\mu) = \frac{n}{\sigma^2}\text{Var}(\bar{X_n}) = \frac{n}{\sigma^2}\frac{\sigma^2}{n} = 1 \tag{63}$$

(b) Find the MGF of Z_n in terms of M, the MGF of each X_i

$$MGF_{Z_n}(t) = e^{Z_n t} = e^{\sqrt{n}\frac{\bar{X_n}-\mu}{\sigma}t} = e^{t\sqrt{n}\frac{\bar{X_n}}{\sigma} - t\sqrt{n}\frac{\mu}{\sigma}} = e^{\frac{t\sqrt{n}}{n\sigma}\sum_{i=1}^n X_i} e^{-t\sqrt{n}\frac{\mu}{\sigma}} \tag{64}$$

$$\text{Note MGF}(X_i) = M = e^t X \text{ where t is any real} \tag{65}$$

$$\text{Now let } \frac{t\sqrt{n}}{n\sigma} = t' \tag{66}$$

$$=> MGF_{Z_n} = e^{t'\sum_{i=1}^n X_i} e^{-t\sqrt{n}\frac{\mu}{\sigma}} = e^{t'X_1 + t'X_2 ... t'X_n} e^{-t\sqrt{n}\frac{\mu}{\sigma}} = \prod_{i=1}^n (M) e^{-t\sqrt{n}\frac{\mu}{\sigma}} = M^n e^{-t\sqrt{n}\frac{\mu}{\sigma}} \tag{67}$$

Happy halloween!