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

Alexander Van Roijen

October 29, 2018

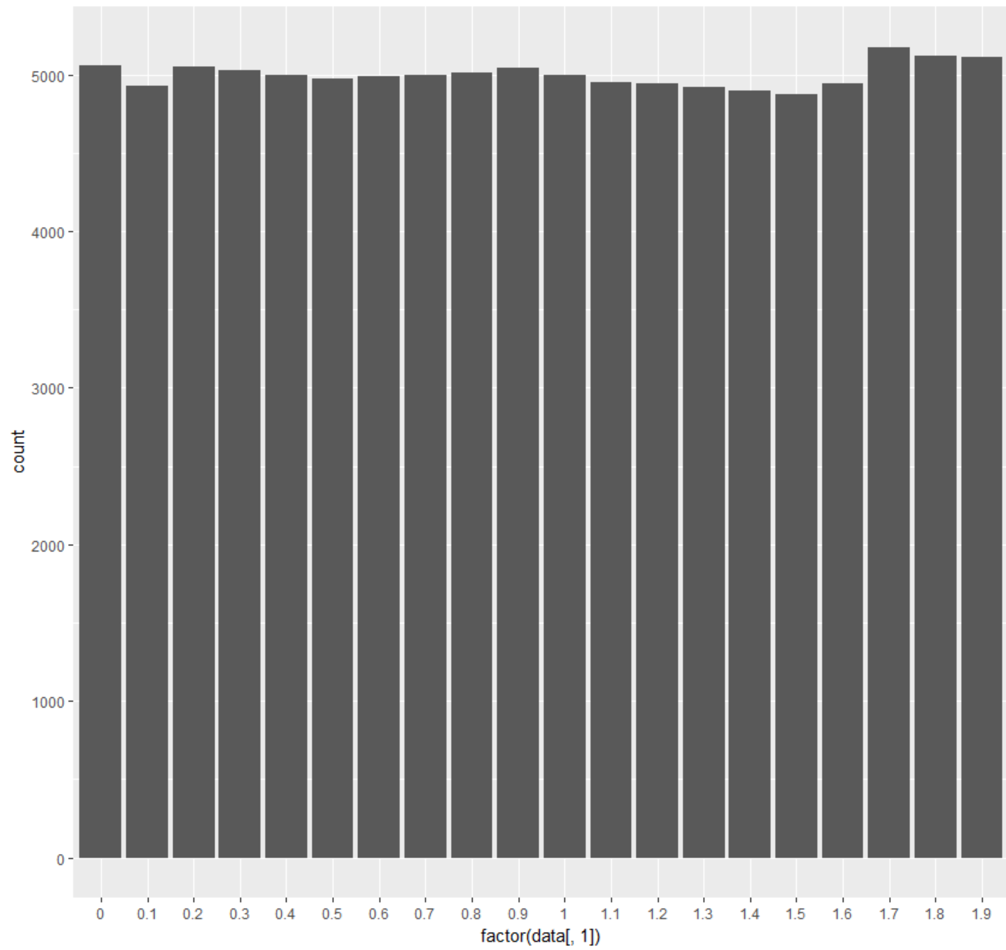
Please complete the following:

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)

$$\text{Take the result form above and take the derivative} \tag{1}$$

$$f(X|X>a) = F'(X|X>a) = \frac{dF(X|X>a)}{dx} = \frac{F'(x)-F'(a)}{1-F(a)} \text{ with} \frac{dF(a)}{dx} = 0 \tag{2}$$

$$=> f(X|X>a) = \frac{f(x)}{1-F(a)} \tag{3}$$

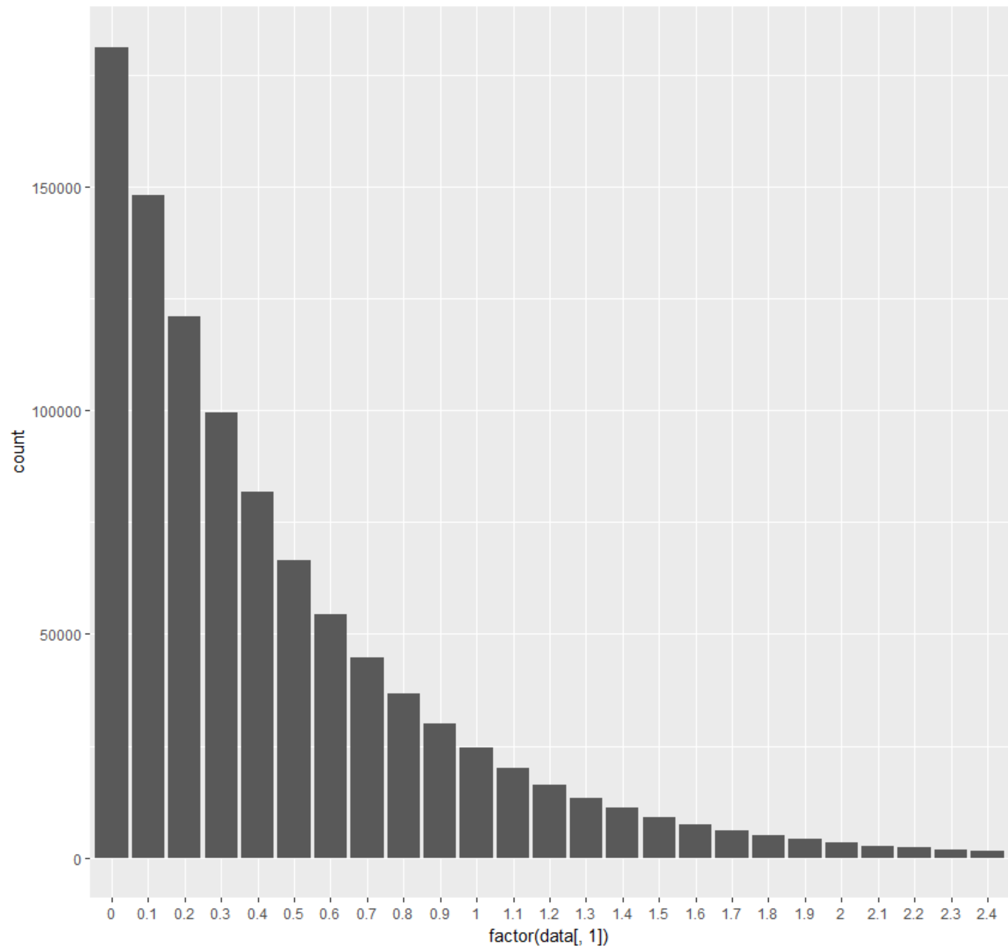
2. Let $X \sim$ Expo(λ)

(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 $X \sim \text{Expo}(\lambda)$

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 demonstrated from Unif(1,1)

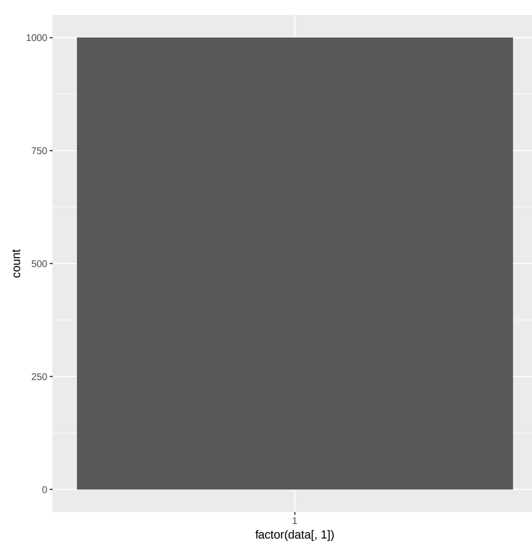


Figure 4: Mode across demonstrated from Unif(1,1)

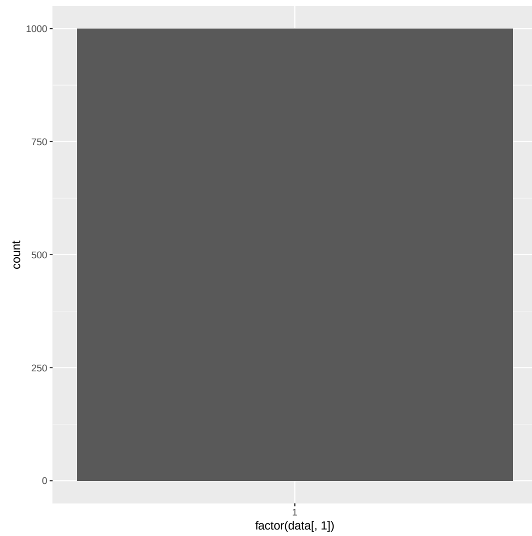


Figure 5: Mode across demonstrated from Unif(1,1)

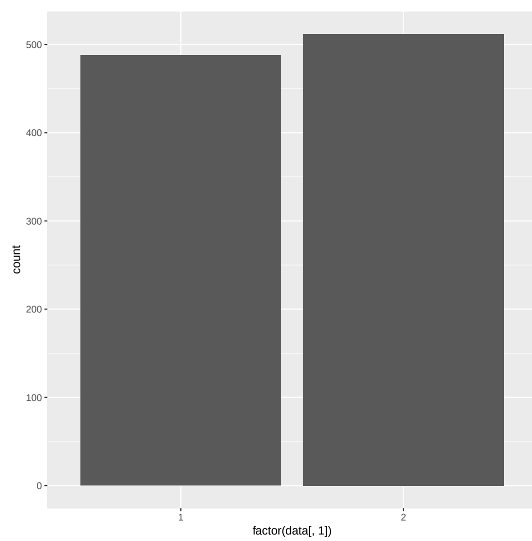


Figure 6: Mode across demonstrated from Unif(1,1)

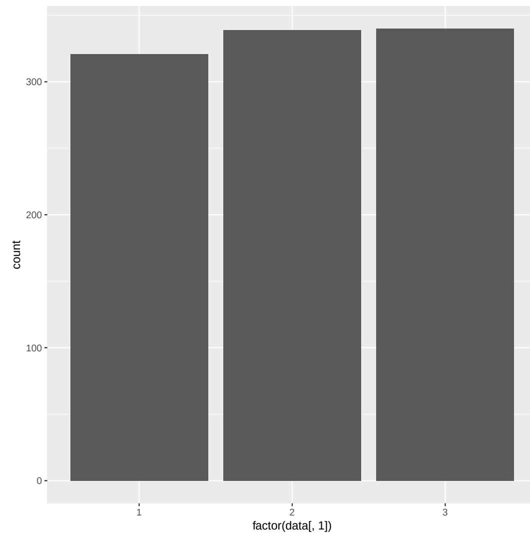


Figure 7: Mode across demonstrated from Unif(1,1)

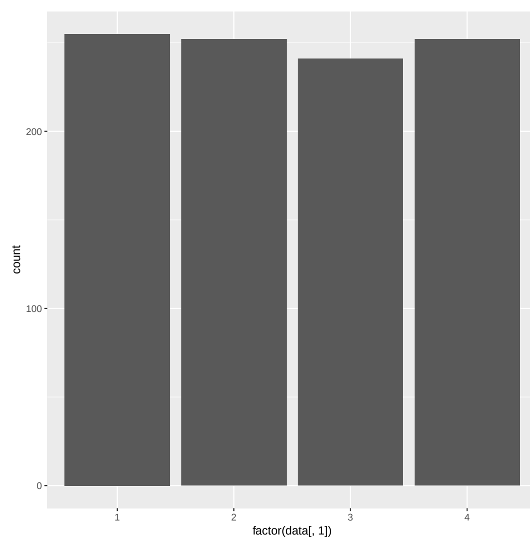


Figure 8: Mode across demonstrated from Unif(1,1)

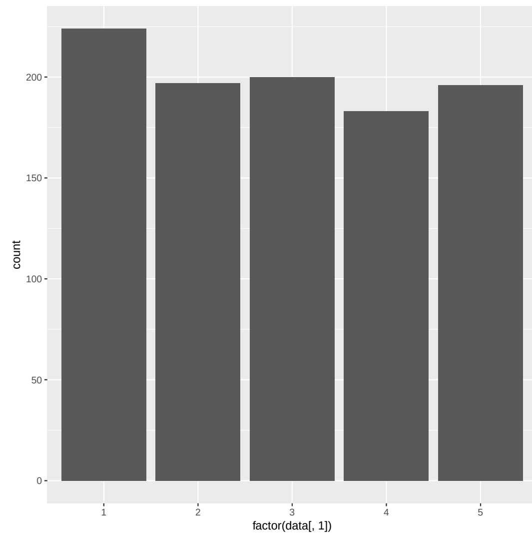


Figure 9: Mode across demonstrated from Unif(1,1)

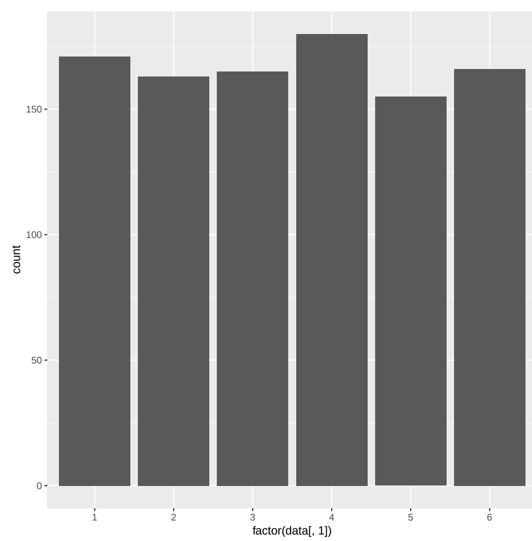


Figure 10: Mode across demonstrated from Unif(1,1)

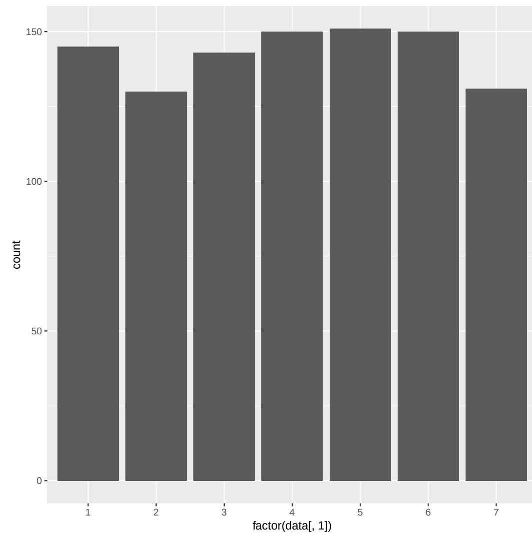


Figure 11: Mode across demonstrated from Unif(1,1)

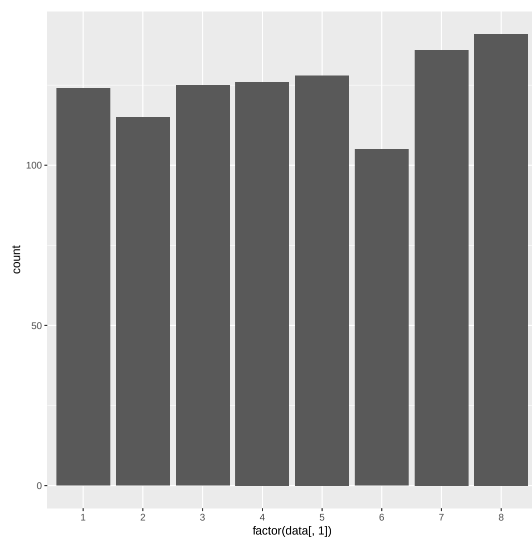
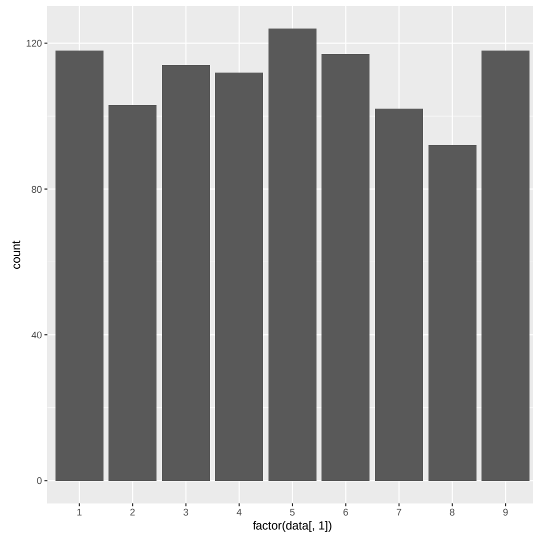


Figure 12: Mode across demonstrated from Unif(1,1)



(b) Find All medians and modes of X

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

$$\text{Want } P(X=c) \geq P(X=x) \forall x \in 1,2,...n \tag{4}$$

$$=> \frac{1}{n-1+1} \geq \frac{1}{n-1+1} \text{ by def of discrete Uniform} \tag{5}$$

$$=> c = x \forall x \in 1,2,3,...n \text{ which is almost the same as 1} \tag{6}$$

$$\text{Notable difference is that in the discrete case, c can only take discrete values} \tag{7}$$

$$\text{Now the median} \tag{8}$$

$$\text{Want } P(X \leq x) \geq \frac{1}{2} \& P(X \geq x) \geq \frac{1}{2} \tag{9}$$

$$P(X \leq x) = \sum_{i=1}^{x} \frac{1}{n} = \frac{x}{n} \geq \frac{1}{2} => x \geq \frac{n}{2} \tag{10}$$

$$\text{However, we can observe the patterns noted in the graphs below} \tag{11}$$

$$\text{If n is odd, the previous conclusion is the only solution as} \tag{12}$$

$$\text{if you go above or below } \frac{n}{2} \text{ you lose the probability @ } x = \frac{n}{2} \tag{13}$$

$$\text{This is true as we have a jump exactly at } \frac{n}{2} \tag{14}$$

$$\text{In the case n is even, you have some wiggle room} \tag{15}$$

$$\text{There is no jump at } \frac{n}{2} \tag{16}$$

$$\text{In fact, the next jump is at } \frac{n}{2}+1 \tag{17}$$

$$\text{This means we have medians from } [\frac{n}{2}, \frac{n}{2}+1] \tag{18}$$

$$\text{One last important factor to note, is that this result relies heavily on} \tag{19}$$

$$\text{how we define the median and our environment} \tag{20}$$

$$\text{In general we have} \tag{21}$$

$$Median(X) = \begin{cases} \frac{n}{2} & n = \text{odd} \\ [\frac{n}{2}, \frac{n}{2}+1] & n = \text{even} \end{cases} \tag{22}$$

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

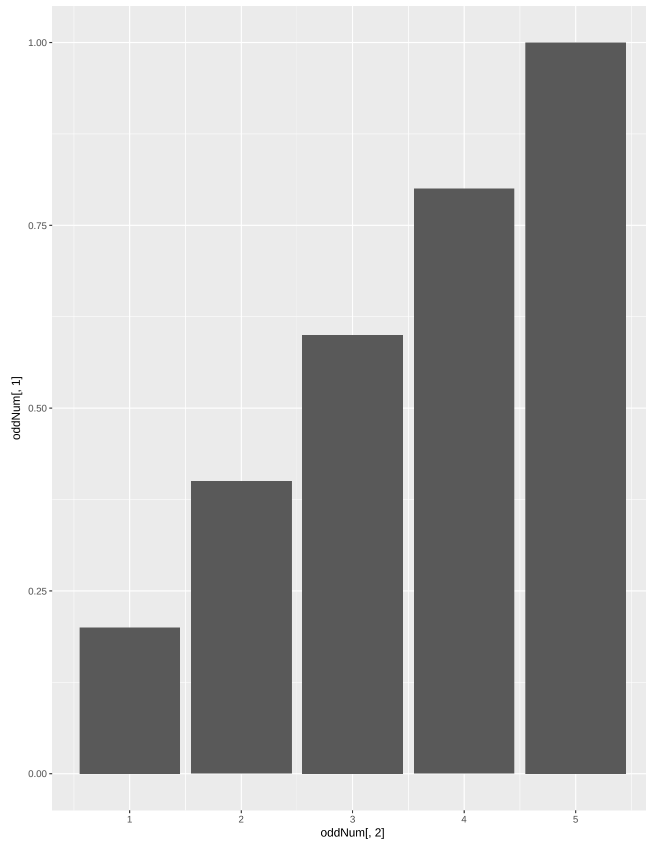
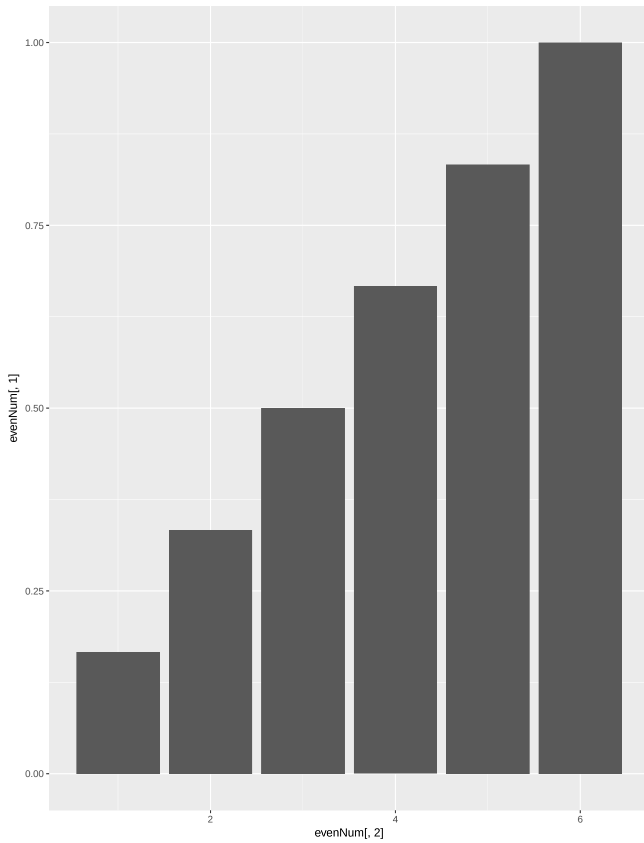


Figure 14: (n=6)These jump patterns hold for all discrete n



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.

$$\text{First note that from our notes, we know the following} \tag{23}$$

$$1: \text{ The mean and median are on in the same by definition} \tag{24}$$

$$2: \ f(x) = f(2\mu - x) \forall x \& P(X \geq X + \mu) = P(X \leq X - \mu) \tag{25}$$

$$\text{Where } \mu \text{ is the mean of our distribution, or central point} \tag{26}$$

$$\text{All that is left to show is that the mode is equivalent to either our median or mean} \tag{27}$$

$$\int_{-\infty}^{\mu - x} f(x)dx = \int_{\mu + x}^{\infty} f(x)dx = \int_{-\infty}^{\mu - x} f(2\mu - x)dx = \int_{\mu + x}^{\infty} f(2\mu - x)dx \tag{28}$$

$$\int_{-\infty}^{\mu} f(x)dx = \frac{1}{2} = \int_{\mu}^{\infty} f(x)dx \tag{29}$$

$$\text{By the defintion of the mode, we have some c } s.t. \, f(c) \geq f(x) \forall x \tag{30}$$

$$\text{Lets proove via contradiction. If } c \neq \mu => c > \mu \text{ or } c < \mu \tag{31}$$

$$=> f(c) = f(2\mu - c) \text{ with } c \neq 2\mu - c \text{ by our previous statement} \tag{32}$$

$$=> \exists c_1 \, s.t. f(c_1) \geq f(x) \forall x \& c_2 \, s.t. f(c_2) \geq f(x) \forall x \text{ with } c_1 \neq c_2 \tag{33}$$

$$\text{This is a contradiction! as now we have two unique modes} \tag{34}$$

$$\text{which violates the property of a unimodal distribution}\square \tag{35}$$

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.

$$\text{Var}(X - Y) = \text{Var}(X) + Var(Y) = b^2 + d^2 \tag{36}$$

$$E[X - Y] = E[X] - E[Y] = a - c \tag{37}$$

$$=> P(X < Y) = P(X - Y < 0) \sim N(a - b, c^2 + d^2) \text{ determined via hint} \tag{38}$$

$$= \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}} \text{ Where } \sigma = \sqrt{c^2 + d^2} \text{ and } \mu = a - b \tag{39}$$

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.

Similar to 5, except now we need to calculate the expected value and variance over the 100 samples

$$E[x-y] = E[\frac{1}{n}\sum_{n=1}^{100} x_n - y_n] = \frac{1}{n}\sum_{n=1}^{100} E[X_n] - E[y_n] = \frac{1}{n}\sum_{n=1}^{100}(69.1-63.7) = \frac{1}{n}n*(69.1-63.7) \tag{40}$$

$$\tag{41}$$

$$\text{we got this from the distributions of the random variables given} \tag{42}$$

$$=> E[x-y] = 5.4 \tag{43}$$

$$\text{Var}[x-y] = \text{Var}[\frac{1}{n}\sum_{n=1}^{100} x_n - y_n] = \frac{1}{n^2}\text{Var}[\sum_{n=1}^{100} x_n - y_n] \text{ from independence we get} \tag{44}$$

$$= \frac{1}{n^2}\sum_{n=1}^{100}(\text{Var}[x_n]+\text{Var}[y_n]) = \frac{1}{n^2}\sum_{n=1}^{100}(2.9^2+2.7^2) = \frac{1}{n^2}n(15.7) = \frac{1}{100}(15.7) \tag{45}$$

$$=> \text{Var}[x-y] = \sqrt{\frac{15.7}{100}}^2 \tag{46}$$

$$=> x-y \sim N(5.4, \sqrt{\frac{15.7}{100}}) \tag{47}$$

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

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