

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

Alexander Van Roijen

October 31, 2018

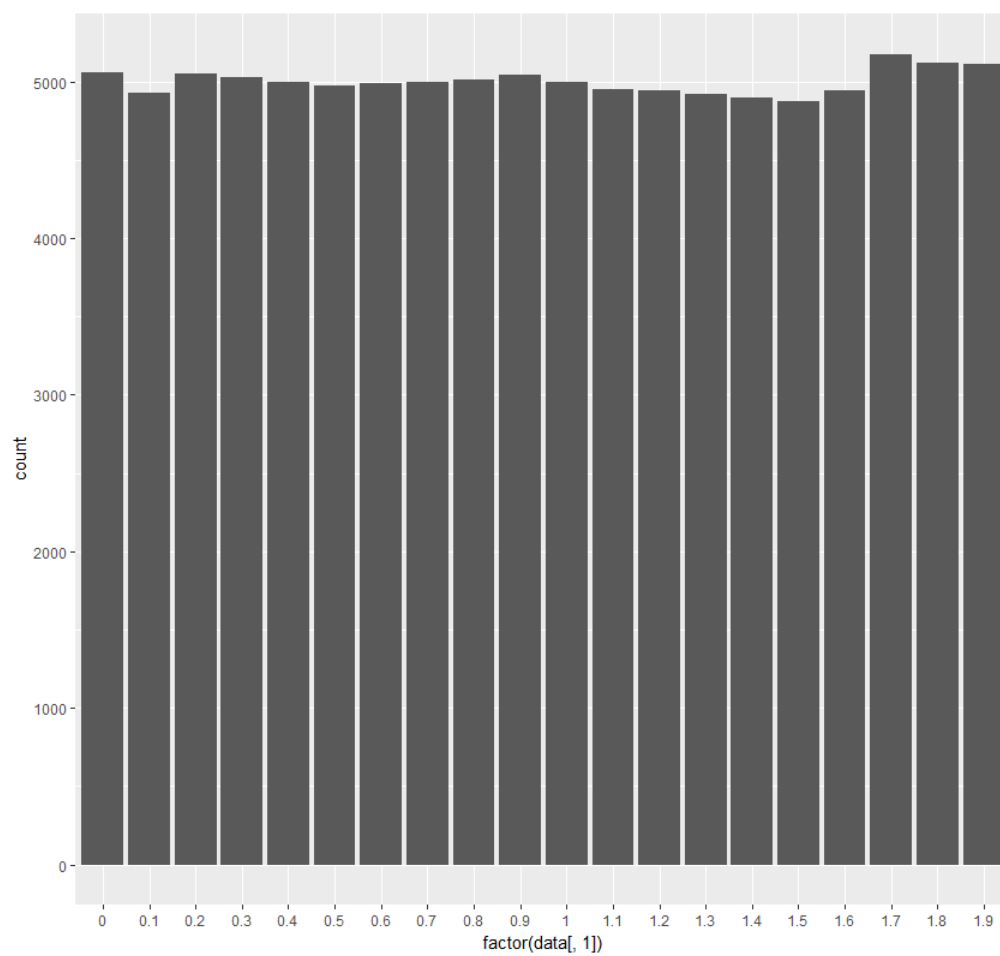
Please complete the following:

1. Problem 1 Let $U \sim \text{Unif}(a, b)$.

- (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 $U \sim \text{Unif}(a,b)$

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

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

$$\Rightarrow \frac{c-a}{b-a} \geq \frac{1}{2} \Rightarrow c \geq \frac{b-a}{2} + \frac{2a}{2} \Rightarrow c \geq \frac{b+a}{2} \quad (3)$$

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

Note that this unique mode will not apply in the discrete case of the uniform distribution

(5)

Mode: (6)

Want: $f(c) \geq f(x) \forall x$ (7)

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

\Rightarrow Mode of X is entire support of X = $[a..b]$ (9)

2. Let $X \sim \text{Expo}(\lambda)$

(a) Use R to simulate median and mode of $\text{Expo}(2)$

```
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_discrete()
+   print(p)
+   #barplot(results,main="whatev",width=0.5)
+   print(median(resultsog))
+   getmode = table(resultsog)
```

```

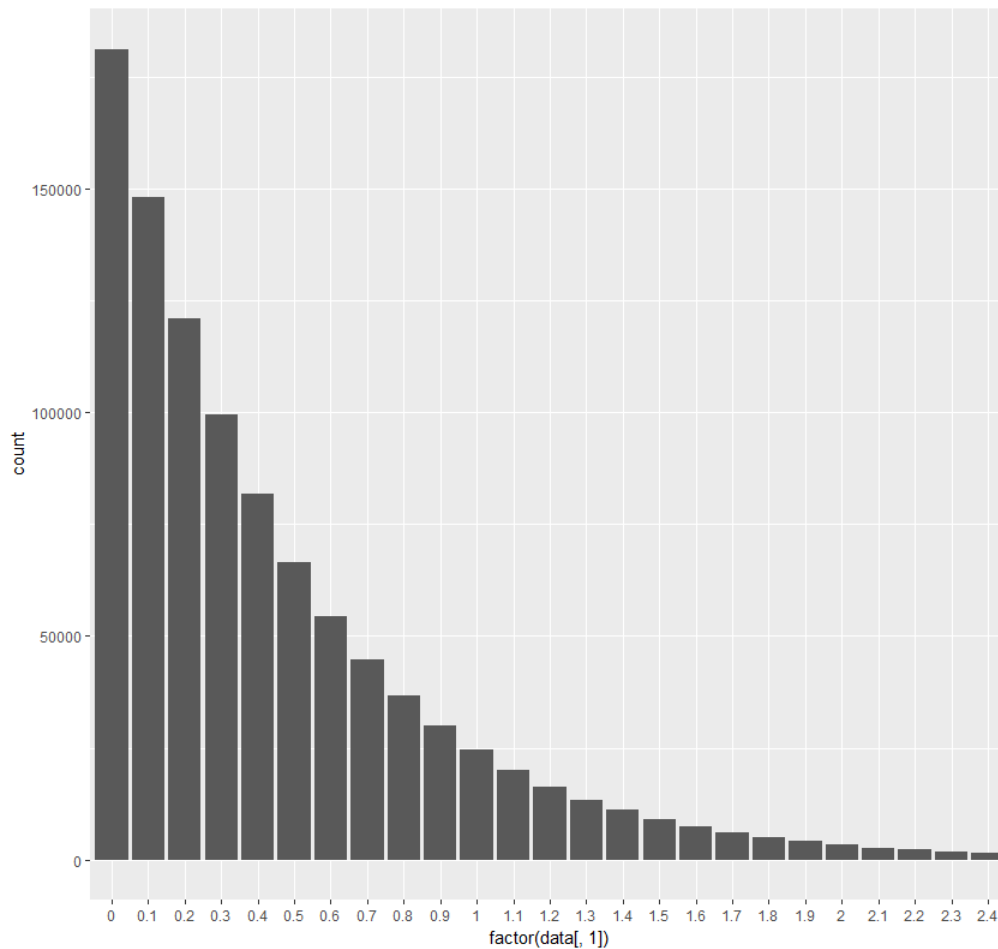
+   print(which.max(getmode))
+ }

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

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

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

$$\text{Coming from the other side we get} \quad (12)$$

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

So again, we have our unique median. This time as a function of lambda which makes sense

$$(14)$$

$$\text{Mode:} \quad (15)$$

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

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

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

a decreasing value of $f(x)$ that starts at $x = 0$ indicates our mode lies at $x = 0$ / $wf(x) = \lambda$

$$(19)$$

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)

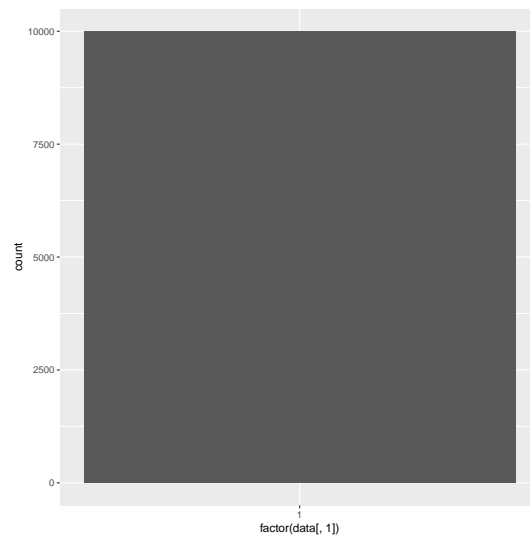


Figure 4: Mode across X demonstrated from Unif(1,2)

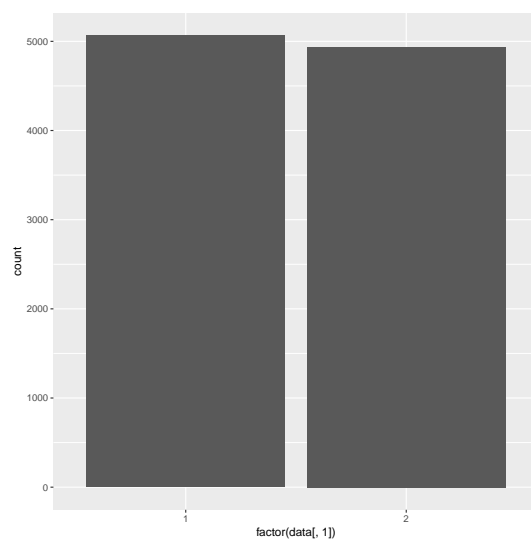


Figure 5: Mode across X demonstrated from Unif(1,3)

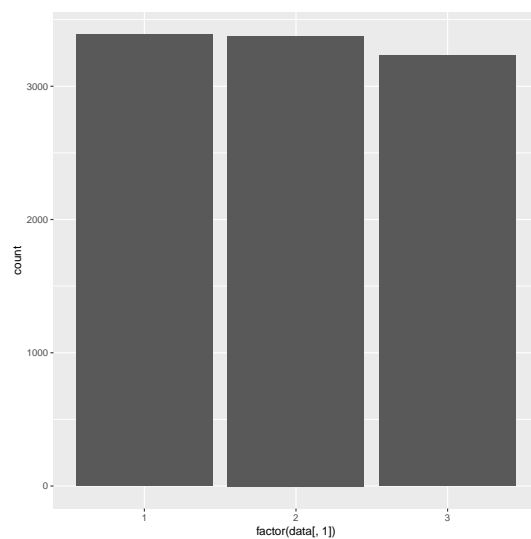


Figure 6: Mode across X demonstrated from Unif(1,4)

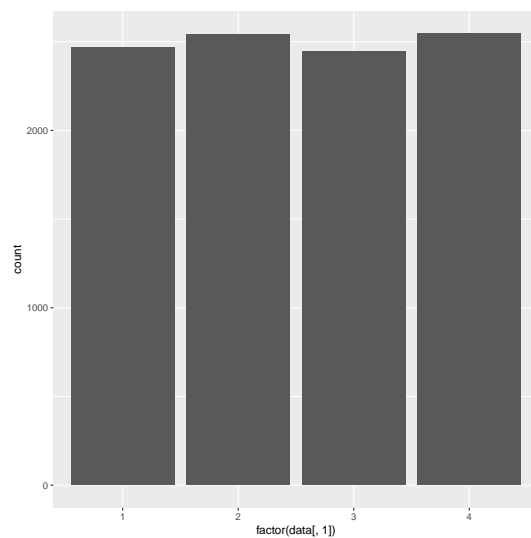


Figure 7: Mode across X demonstrated from Unif(1,5)

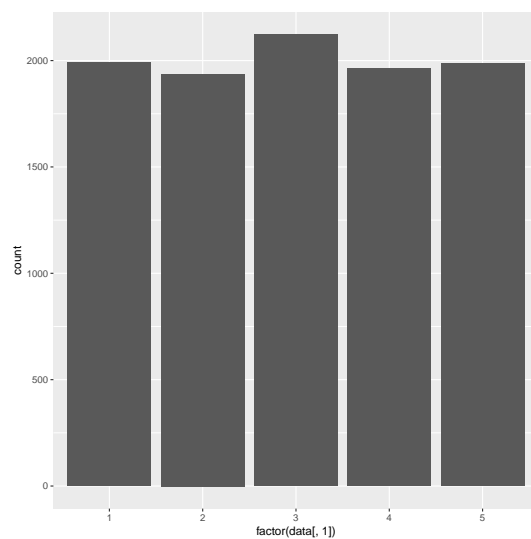


Figure 8: Mode across X demonstrated from Unif(1,6)

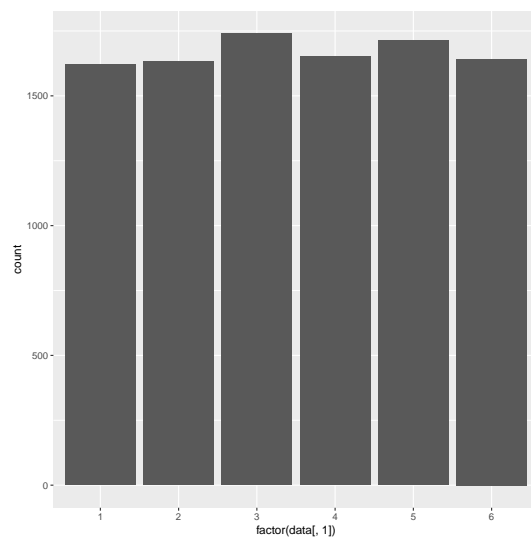


Figure 9: Mode across X demonstrated from Unif(1,7)

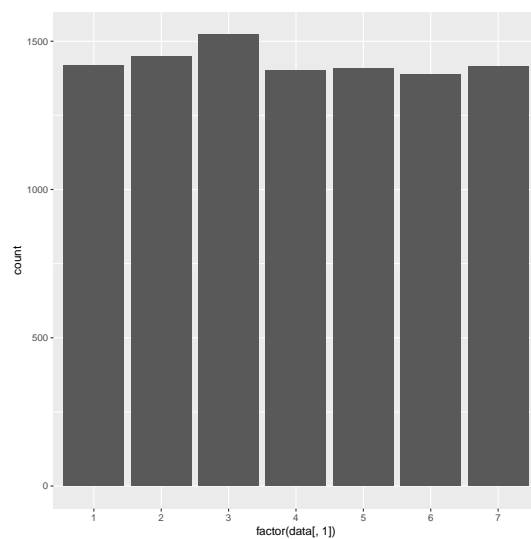


Figure 10: Mode across X demonstrated from Unif(1,8)

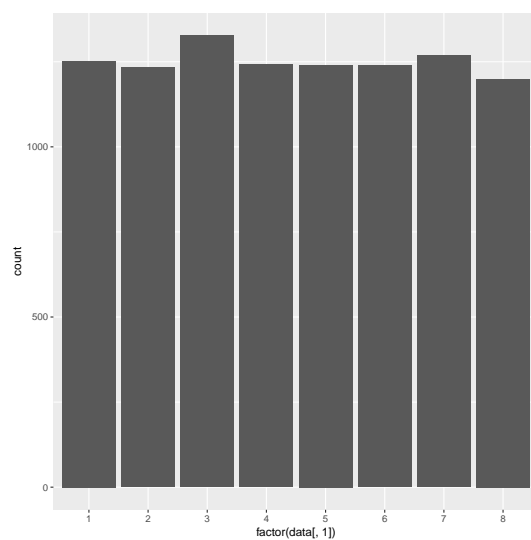


Figure 11: Mode across X demonstrated from Unif(1,9)

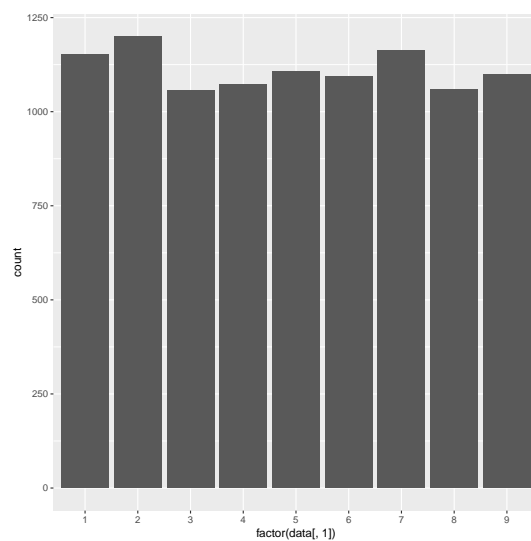
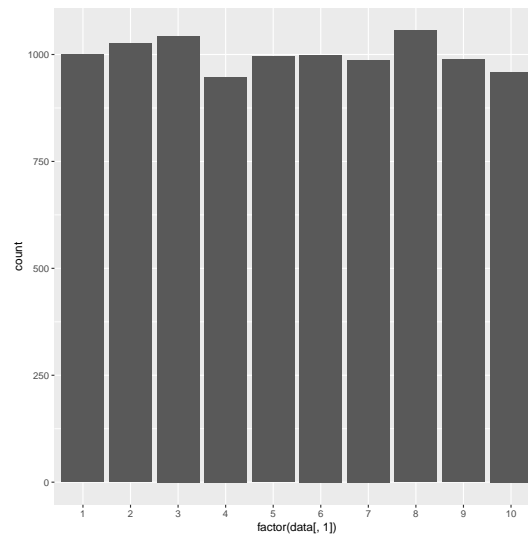


Figure 12: Mode across X demonstrated from Unif(1,10)



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.

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

$$\Rightarrow \frac{1}{n-1+1} \geq \frac{1}{n-1+1} \text{ by def of discrete Uniform} \quad (21)$$

$$\Rightarrow c = x \forall x \in 1, 2, 3, \dots, n \text{ which is almost the same as } 1 \quad (22)$$

Notable difference is that in the discrete case, c can only take discrete values (23)

Now the median (24)

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

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

However, we can observe the patterns noted in the graphs below (27)

If n is odd, the previous conclusion is the only solution as (28)

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

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

In the case n is even, you have some wiggle room (31)

There is no jump at $\frac{n}{2}$ (32)

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

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

One last important factor to note, is that this result relies heavily on (35)

how we define the median and our environment (36)

In general we have (37)

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

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

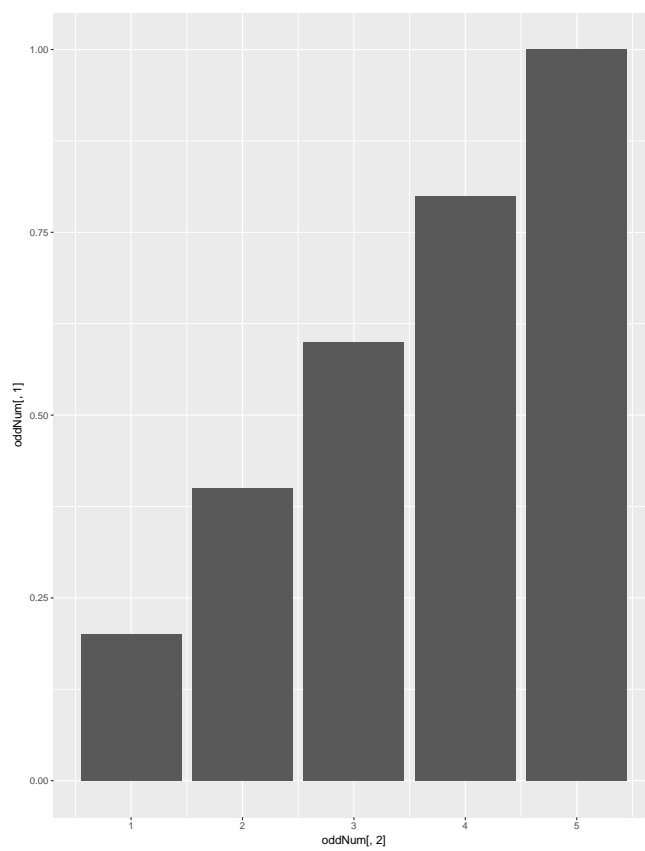
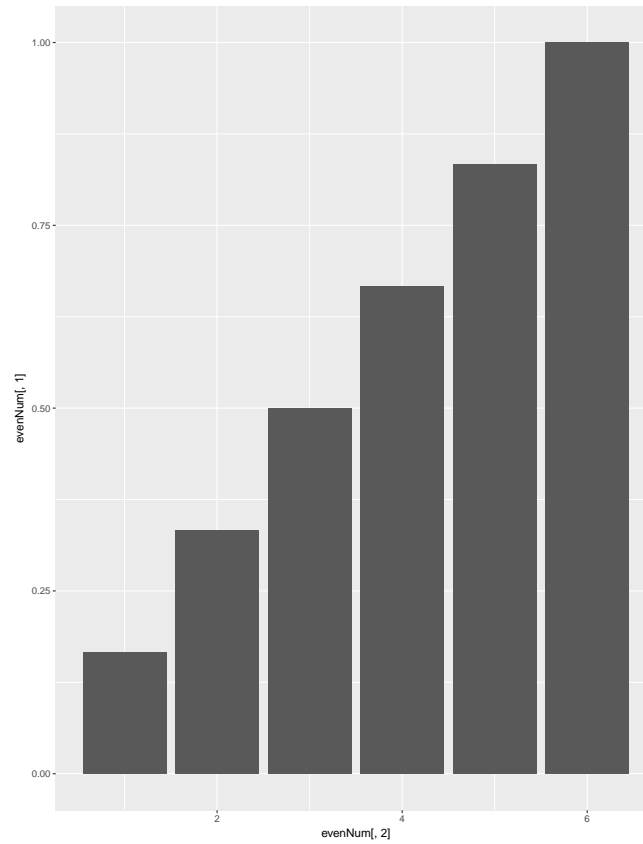


Figure 14: (n=6) These jump patterns hold for all discrete n



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.

First note that from our notes, we know the following (39)

1: The mean and median are on in the same by definition (40)

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

Where μ is the mean of our distribution, or central point (42)

All that is left to show is that the mode is equivalent to either our median or mean

(43)

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

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

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

Lets proove via contradiction. If $c \neq \mu \Rightarrow c > \mu$ or $c < \mu$ (47)

$\Rightarrow f(c) = f(2\mu - c)$ with $c \neq 2\mu - c$ by our previous statement (48)

$\Rightarrow \exists c_1$ s.t. $f(c_1) \geq f(x) \forall x$ & c_2 s.t. $f(c_2) \geq f(x) \forall x$ with $c_1 \neq c_2$ (49)

This is a contradiction! as now we have two unique modes (50)

which violates the property of a unimodal distribution \square (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.

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

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

since $X' & Y'$ independet $\Rightarrow M_{X'+Y'} = M_{X'} * M_{Y'} = M_{X^2} * M_{Y^2} = ((1 - 2t)^{-\frac{1}{2}})^2 = \frac{1}{1 - 2t}$ (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 .

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

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

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

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

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

7. Let X_1, \dots, X_n be i.i.d. with mean μ , variance σ^2 , and MGF M . Let

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \text{ and } Z_n = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$$

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

$$E[\bar{X}_n] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} * n\mu = \mu \quad (60)$$

$$\Rightarrow E[Z_n] = E\left[\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}\right] = \sqrt{n} \frac{E[\bar{X}_n] - \mu}{\sigma} = 0 \quad (61)$$

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \quad (62)$$

$$\Rightarrow \text{Var}[Z_n] = \text{Var}\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}\right) = \frac{n}{\sigma^2} \text{Var}(\bar{X}_n - \mu) = \frac{n}{\sigma^2} \text{Var}(\bar{X}_n) = \frac{n}{\sigma^2} \frac{\sigma^2}{n} = 1 \quad (63)$$

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

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

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

$$\text{Now let } \frac{t\sqrt{n}}{n\sigma} = t' \quad (66)$$

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