Homework 2

Erik Lee

Mon Jul 9 14:39:14 2018

Homework policy: This homework is due by 5:00pm (EST) on the due date. Homework is to be handed in via the course website in pdf format. Although we prefer you use Rmarkdown or Word, you do not need to type the homework; there are many ways (scanner in the library or phone apps) to convert written homework into a pdf file. Ask the teaching staff if you need assistance.

Late homework will not be accepted. You are encouraged to discuss homework problems with other students (and with the instructor and TFs, of course), but you must write your final answer in your own words. Solutions prepared "in committee" or by copying someone else's paper are not acceptable.

1) A fair six-sided die is rolled.

a) What are the possible outcomes of this event?

{1,2,3,4,5,6}

b) Calculate the probability of rolling a prime number.

P(prime) = 3/6 = 1/2 {2,3,5} (1 is not a prime number)

c) Calculate the probability of rolling an even number.

P(even) = 3/6 = 1/2 {2,4,6}

d) What is the probability of rolling a number greater than seven?

P(X>7) = 0

---

2) The probability that a driver is speeding on a stretch of road is 0.27. What is the probability that a driver is not speeding?

P(not speeding) = P(complement of speeding) = 1- 0.27 = 0.73

---

3) A department store manager has monitored the numbers of complaints received per week about poor service. The probabilities for numbers of complaints in a week, established by this review, are shown in the table below.

NUMBER OF COMPLAINTS	0	1-3	4-6	7-9	10-12	More than 12
PROBABILITY	.15	.29	.16	?	.14	.06

Let A be the event "There will be at least one complaint in a week," and B the event "There will be less than 10 complaints in a week."

a) Find the value of ?

? = 1 - (.15+.29+.16+.14+.06) = 0.2

b) Find the probability of A.

P(A) = .29+.16+.2+.14+.06 = 0.85

c) Find the probability of B.

P(B) = 1 - (.14+.06) = 0.8

d) Find the probability of the complement of A.

P(A') = 1-0.85 = 0.15

e) Find the probability of A or B.

P(A or B) = 0.85+0.8-(.29+.16+.2) = 1

f) Find the probability of A and B.

P(A and B) = 0.85*0.8 = 0.68

---

4) Answer the following questions using the following joint probability table

	No wind	Some wind	Strong wind	Storm
No rain	0.1	0.2	0.05	0.01
Light rain	0.05	0.1	0.15	0.04
Heavy rain	0.05	0.1	0.1	0.05

a) Find the marginal probability P(light rain).

P(light rain) = 0.05+0.1+0.15+0.04 = 0.34

b) Find the marginal probability P(strong wind).

P(strong wind) = 0.05+0.15+0.1 = 0.30

c) Find the conditional probability P(heavy rain | strong wind) .

P(heavy rain | strong wind) = 0.1/0.3 = 1/3

d) Find the conditional probability P(some wind | light rain) .

P(some wind | light rain) = 0.1/0.34 = 10/34 = 5/17

---

5) A journal article reported in September 2003 that a Swiss dermatologist established a link between smoking and gray hair in women under 40, with results as shown in this table. Let G denote the event a women in the study has gray hair, and S denote the event a women in the study is a smoker.

	Gray	Not Gray	Total
Smoker	13	10	23
Nonsmoker	0	32	32
Total	13	42	55

a) Was this an observational study or an experiment?

This is an observational study. The journal article uses data based on findings from a Swiss dermatologist. The article is not conducting an experiment to gather data linking smoking and gray hair in women under 40, nor is it reporting to repeat the results found.

b) Find P(G)

P(G) = 13/55

c) Find P(S or G)

P(S or G) = 23/55 + 13/55 - 13/55 = 23/55

d) Find P(S|G)

P(S|G) = (13/55)/(13/55) = 13/13 = 1

e) Are having Gray hair and being a smoker independent or dependent events? Explain

Having gray hair and being a smoker are dependent events. This can be proven by checking if P(S|G) = P(S) or P(G|S) = P(G). If these equalities are true, then there is an independence between S (smoking) and G (gray hair). Checking the first equality, P(S|G) = P(S), P(S|G) = 1 (from part d) and P(S) = 23/55. From this equality we find that P(S|G) != P(S) in this instance and it is false. This leads to the conclusion that Gray hair and smoking are dependent events.

6) Suppose P(A)=0.76, P(B|A)=0.30 and P(B|notA) = 0.02. It might be helpful to construct the 2x2 probability table at this point. Find the following probabilities:

a) P(notB|A)

P(notB|A) = 0.7

b) P(notB|notA)

P(notB|notA) = 0.98

c) P(B)

P(B) = 0.2328

d) P(notB)

P(notB) = 0.7672

e) P(A|B)

P(A|B) = 0.228/0.2328 = 0.9794 (rounded to nearest 4th decimal)

f) P(notA|B)

P(notA|B) = 0.0048/0.2328 = 0.0206 (rounded to nearest 4th decimal)

g) P(A|notB)

P(A|notB) = 0.532/0.7672 = 0.6934 (rounded to nearest 4th decimal)

h) P(notA|notB)

P(notA|notB) = 0.2352/0.7672 = 0.3066 (rounded to nearest 4th decimal)

---

7) Suppose the probability of having schizophrenia P(s) = 0.01 in the population, and the conditional probability of "hearing voices" given schizophrenia P(hv | s) = 0.66, and the probability of "hearing voices" P(hv) = 0.75. Find the probability of having schizophrenia given "not hearing voices".

P(s|not hv) = 0.0034/0.25 = 0.0136

---

8) There are two boxes, Box B1 and Box B2. Box B1 contains 2 red balls and 8 blue balls. Box B2 contains 7 red balls and 3 blue balls. Suppose Jane first randomly chooses one of two boxes B1 and B2, with equal probability, 1/2, of choosing each. Suppose Jane then randomly picks one ball out of the box she has chosen (without telling you which box she had chosen), and shows you the ball she picked. Suppose you only see that the ball Jane picked is red. Given this information, what is the probability that Jane chose box B1?

P(B1|R) = 0.1/0.45 = 1/45 = 0.0222..

---

9) Let X be a discrete random variable with PMF (probability mass function) given by

$$p_X(x) = \begin{cases} x^2/a, & if\ x = -3, -2, -1, 0, 1, 2, 3 \\ 0, & otherwise \end{cases}$$

a) Find the value of a.

a = 28

b) Calculate $E(X)$

E(X) = 0

---

10) The probability that a cellular phone company kiosk sells X number of new phone contracts per day is shown below.

x	4	5	6	8	10
P(x)	0.2	0.25	0.35	0.10	0.10

a) Find the mean, variance, and standard deviation for this probability distribution.

Mean X = 5.95$Var(X)$ = (ˆ2)3.0475 SD(X) = $1.745709

b) Suppose the kiosk salesperson makes $80/day (8 hours at $10/hour), plus a $25 bonus for each new phone contract sold. What is the mean and variance of the salesperson's daily salary?

Mean X = 80 + 25(5.95) = $228.75

Var(X) = 80 + (25ˆ2)(3.0475) = 1984.6875 ($ˆ2 or square dollars) Variance, when multiplied by a constant value c ($25 per phone in this instance), increases the vaiance by the square of the constant (25ˆ2 = 625).

---

11) In a population of students, the number of calculators owned is a random variable X with P(X = 0) = 0.4, P(X = 1) = 0.5, P(X = 2) = 0.1.

a) Find E(x)

E(X) = 0.7

b) Find Var(X)

Var(X) = 0.41

---

12) You roll two dice.

a) What is the probability of two sixes? Of exactly one 6? Of no sixes?

P(two sixes) = (1/6)*(1/6) = 1/36 P(exacly one six) = (1/6)(5/6) + (5/6)(1/6) = 10/36 = 5/18 P(no sixes) = (5/6)*(5/6) = 25/36

b) What is the expected number of sixes that will show?

E(X = number of sixes that will show) = 2(1/36) + 1(5/18) + 0(25/36) = 12/36 = 1/3

---

13) We can simulate the expected value result in part (b) above. Follow the following steps in R:

i. Simulate two dice rolls using (use similar code for die2) `die1=sample(1:6,10000,replace=TRUE)`

ii. Combine the two dice rolls into a matrix using `dicerolls=cbind(die1,die2)`

iii. Each row of dicerolls represents the outcome of rolling two dice. We want to count how many 6's appear each time we roll two dice. We do that as follows: `num6=rowSums(dicerolls==6)`

iv. Take the mean of the num6 variable and compare it to part (b) above. How does this mean change if we instead use 1000000 rolls?

If 1,000,000 rolls were taken, the calculated mean would be closer to 0.3333.... or 1/3 than the mean of num6 for 10,000 rolls. This follows the Law of Large numbers, which states that as the number of trials (sample size) conducted increases, the observed (sample) mean appproaches/gets closer in vlaue to the expectation. In this instance, with 1,000,000 rolls, the observed mean will be clsoer to $1/3 = 0.3333..$ than to the mean of numb6.

Here is code to test the difference in mean. The first value is the mean for 10,000 rolls and the second is the mean for 1,000,000 rolls:

```
# Running 10,000 rolls
die1=sample(1:6,10000,replace=TRUE)
die2=sample(1:6,10000,replace=TRUE)
dicerolls=cbind(die1,die2)
num6=rowSums(dicerolls==6)
mean(num6)
```

```
## [1] 0.3307
```

```
# Running 1,000,000 rolls
die1=sample(1:6,1000000,replace=TRUE)
die2=sample(1:6,1000000,replace=TRUE)
dicerolls=cbind(die1,die2)
num6=rowSums(dicerolls==6)
mean(num6)
```

```
## [1] 0.3329
```

Based on the R code above, when 1,000,000 rolls were conducted, the mean of num6 was closer to the expected 1/3 than 10,000 rolls. We see this, when we run the code, as the second mean is closer to $1/3 = 0.3333$ thant the first mean.

---

14) If random variable X has mean μ and variance σ^2, show (using the a+bX rule) what the mean and variance of $Z = (X - \mu)/\sigma$ are.

Mean = X - sigma*Z

Variance = ((X - mu)/Z)^2

---

15) Find the variance of each of the following bets from the class notes. Which bet is riskiest and which best is safest?

a) You get $5 with probability 1.0.

Var(X) = 0

b) You get $10 with probability 0.5, or $0 with probability 0.5.

Var(X) = 25

c) You get $5 with probability 0.5, $10 with probability 0.25 and $0 with probability 0.25.

Var(X) = 12.5

d) You get $5 with probability 0.5, $105 with probability 0.25 or lose $95 with probability 0.25.

Var(X) = 5000

The riskiest is bet d. This has the highest variance of 5000 indicating that the amount recieved can diverge from the mean significantly, with regards to spread. There is a 0.25 chance to lose $95 and a 0.25 chance to win $105. It gives the best chance to win big or lose big.

The safest is bet a. This has a variance of 0 which indicates that the winnings do not diverge from the mean at all. There is a 100% chance and expectation to win $5.

---

16) Let X be a random variable with E(X) = 20 and Var(X) = 10. Find the following.

a) $E(X^2)$

E(Xˆ2) = (20)20 = 400 Var(Xˆ2) = (10ˆ2)10 = 1000

b) $E(3X + 10)$

E(3X+10) = 3(20)+10 = 70 Var(3X+10) = (3ˆ2)(10) + 10 = 100

c) $E(-X)$

E(-X) = (-1)20 = -20 Var(-X) = ((-1)ˆ2)(10) = 10

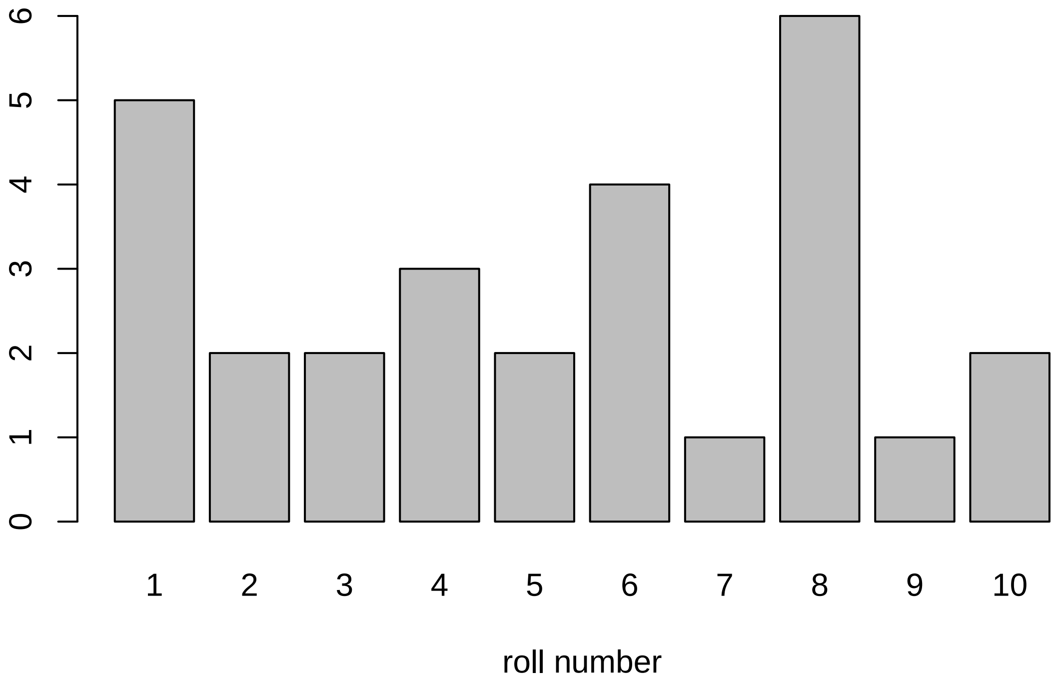
d) Standard deviation of –2X ?

Var(-2X) = ((-2)ˆ2)10 = 40 SD(-2X) = (sqrt(Var(-2X))) = sqrt(40) = 6.325

---

17) Probability via simulation. We can simulate rolling a 6 sided dice many times in R using the following code. In this example we roll a dice 10 times.

```
n=10
dicerolls = sample(1:6,n,replace=TRUE)
barplot(dicerolls,xlab="roll number",names=1:10)
```



Suppose we want to find P(get a 6 on a dice roll). Pretty boring [the answer is 1/6] but this is how you would find it via simulation. Of course this value will change every time we run the simulation [the random dice rolls change every time]

```
n=1000
dicerolls = sample(1:6,n,replace=TRUE)
prob6 = sum(dicerolls==6)/n
cat("The probability of rolling a 6 is",prob6,"\n")
```

```
## The probability of rolling a 6 is 0.144
```

We could easily extend this code to two dice (or more). Extend this code to find the probability the sum of three dice is 13. You can check your simulated answer with the answer given here https://www.thoughtco.com/probabilities-for-rolling-three-dice-3126558

```
n=1000
diceroll1 = sample(1:6,n,replace=TRUE)
diceroll2 = sample(1:6,n,replace=TRUE)
diceroll3 = sample(1:6,n,replace=TRUE)

prob13 = sum(diceroll1+diceroll2+diceroll3==13)/n
cat("The probability of rolling a 13 is",prob13,"\n")
```

```
## The probability of rolling a 13 is 0.112
```

The link states that the probability of rolling a 13 with 3 dies is 9.7%. After running the code several times, the probability floats around 0.1 or 10%, which matches the link's probability.