MIT Introduction to Statistics 18.05 Problem Set 3

John Hancock

March 16, 2017

Contents

1 References and License

We are answering questions in the material from MIT OpenCourseWare course 18.05, Introduction to Probability and Statistics.
In this document we are answering questions Orloff and Bloom ask in [2].
Please see the references section for detailed citation information.
The material for the course is licensed under the terms at `http://ocw.mit.edu/terms`.
We use documentation in [9], [6], [7] to write LaTeXsource code for this document.

2 Independence

In this section we answer a problem in [2] that involves rolling two six sided dice.

2.1 Pairwise and mutual independence

We define two events, A, and B to be pairwise independent if $P(A \cap B) = P(A)\,P(B)$.

For this problem Orloff and Bloom give us the definition of mutual independence for three events, A, B, and C. A, B, and C are mutally independent if

$$P(A \cap B \cap C) = P(A)\, P(B)\, P(C) \tag{1}$$

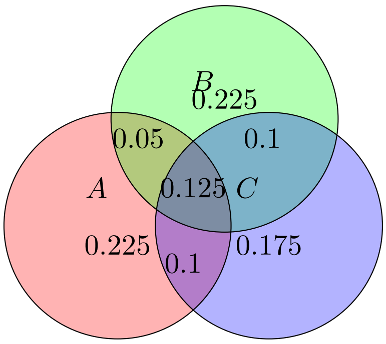
In this section, Orloff and Bloom give the following definitions for events A, B, and C:

- A is the event that we roll an odd number with the first die.
- B is the event that we roll an odd number with the second die.
- C is the event that the sum of the numbers we roll is odd.

A, B, and C are not mutually independent. Whatever the A, B, and C probabilities of A, B, and C are individually, the probability of $P(A \cap B \cap C)$ is 0 since the sum of two odd numbers is always an even number.

2.2 Venn diagram

Orloff and Bloom give the following Venn diagram:



And ask us whether or not the events in the Venn diagram above are mutually independent.
These events are not mutually independent because

$$P(A)\, P(B)\, P(C) = 0.225 \times 0.225 \times 0.175 = 0.008859375. \tag{2}$$

However, in the Venn diagram above, Orloff and Bloom give us that $P(A \cap B \cap C) = 0.125$
Therefore the events are not mutually independent.

2.3 How many kids

For this question we use the same assumptions about the probability of the gender that a child is born with that Orloff and Bloom use in example 9 of [4].
We define the following events:

- A is the event that the children in a family are both boys and girls.
- B is the event that at most one of the children is a girl.
- $C_{i,b}$ is the event that child number i is a boy.
- $C_{i,g}$ is the event that child number i is a girl.

Our goal is to construct a sample space such that A and B are independent. The definition of independent events is in [3].
We rely on the same assumption that Orloff and Bloom make in [4] regarding the probability of the genders of sequences of children.
Therefore we assume $P\left(C_{i,b}\right) = 0.5$, and $P\left(C_{i,g}\right) = 0.5$, independent of the event that any other child is a boy or a girl.
We write the following table to discover the number of children where A, and B will meet the definition of independent events.
We fill in one cell in the table below for each possible sequnce of three children in the family being boys or girls.

$C_{1,b}C_{2,b}C_{3,b}$	$C_{1,b}C_{2,b}C_{3,g}$	$C_{1,b}C_{2,g}C_{3,b}$	$C_{1,b}C_{2,g}C_{3,g}$
$C_{1,g}C_{2,b}C_{3,b}$	$C_{1,g}C_{2,b}C_{3,g}$	$C_{1,g}C_{2,g}C_{3,b}$	$C_{1,g}C_{2,g}C_{3,g}$

In the table above there are 6 sequences that are in A, so $P\left(A\right) = \frac{6}{8}$.
Also, there are 4 sequences in B, so $P\left(B\right) = \frac{4}{8}$.
Moreover, there are 3 sequences where there is at most one girl, and the children are both boys and girls. Therefore $P\left(A \cap B\right) = \frac{3}{8}$.
A and B are indpendent since

$$P\left(A\right) P\left(B\right) = \left(\frac{6}{8}\right)\left(\frac{4}{8}\right) = \frac{24}{64} = \frac{3}{8}. \tag{3}$$

Therefore $P\left(A\right) P\left(B\right) = P\left(A \cap B\right)$, so A and B must be independent events.
We made these calculations assuming that there are 3 children, therefore the number of children we require in order for A, and B to be independent events is 3.

3 Dice

In this section we will deal with problems that Orloff and Bloom ask about the random variable X, that is equal to the value we roll with a fair 4-sided die, the random variable Y, that is equal to the value we roll with a fair 6 sided die, and the random variable Z, that is equal to the average of X and Y.

3.1 Standard deviation of X, Y, and Z

We use the definition of variance and standard deviation in [5] to calculate the standard deviations $\sigma(X)$, $\sigma(Y)$.
We use the exact same method to calculate the variance of a discrete random variable many times. For details on how to do the calculation see [1]. We calculate the variance of X, and Y, then take the square root of the variance to obtain the standard deviation.
Here are the results:

$$\sigma_X \approx 1.118. \tag{4}$$

$$sigma_Y \approx 1.708. \tag{5}$$

In order to calculate the variance of Z, we can use properties of variance that Orloff and Bloom show in [5].
X and Y are independent random variables; the value we roll with a fair four sided die has no effect on the value we roll with a fair six sided die. Therefore the equation

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y). \tag{6}$$

We will also use the property of variance Orloff and Bloom show in [5] that

$$\mathrm{Var}(X) = a^2 \mathrm{Var}(X). \tag{7}$$

Therefore, in the scope of this problem,

$$\mathrm{Var}(fracX+Y2) = \mathrm{Var}\left(\frac{X}{2}+\frac{Y}{2}\right) = \frac{1}{4}\left(\mathrm{Var}(X)+\mathrm{Var}(Y)\right). \tag{8}$$

We use the values we calculated for $\mathrm{Var}(X)$, and $\mathrm{Var}(Y)$ above,

$$\mathrm{Var}(Z) \approx \frac{1}{4}(1.118+1.708) \approx 0.707 \tag{9}$$

Standard deviation is the square root of variance, so

$$\sigma_z \approx \sqrt{\frac{1}{4}(1.118^2+1.708^2)} \approx 1.021 \tag{10}$$

3.2 Graph pmf and cdf of Z

We use R to simulate rolling the dice. We find documentation in [8] helpful in writing this source code. Here is a listing of the R source code:

```
y=sample(c(1:6), replace = TRUE, 1000000)
x=sample(c(1:4), replace = TRUE, 1000000)
z=(x+y)/2
zTable = table(z)
zTable/1000000
```

The output of this program is:

```
          z
         1        1.5         2        2.5         3        3.5
4        4.5         5
0.041375 0.083538 0.125244 0.167014 0.166669 0.167029 0.124364 0.083153 0
```

We look at the numerical inverses of the values in the second row of the table above to get a clue about the probability mass of each possible value of Z. Here again we utilize R to compute the inverses:

```
freqs=c(0.041375, 0.083538, 0.125244, 0.167014, 0.166669, 0.167029, 0.124
1/freqs
```

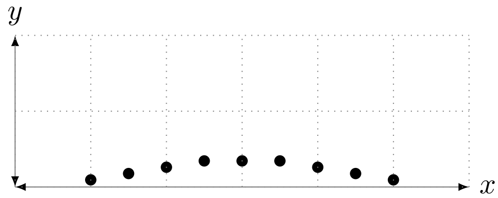
We execute the code above to get the result

```
24.169184 11.970600  7.984414  5.987522  5.999916  5.986984
8.040912 12.026024 24.030374
```

Now it becomes clear that we can approximate the frequencies of the values of Z as fractions with 24 in the denominator. We summarize this approximation in the table below, and make a guess as to the value of the pmf of Z:

Z	1	1.5	2	2.5	3	3.5	4	4.5	5
pmf $p(z)$?	$\frac{1}{24}$	$\frac{2}{24}$	$\frac{3}{24}$	$\frac{4}{24}$	$\frac{4}{24}$	$\frac{3}{24}$	$\frac{2}{24}$	$\frac{1}{24}$	

The data from the sumulation inspires us to make the following reasoning about the pmf of Z: There are 6×4 possible pairs of values we can roll with the two dice. To compute the value of Z we add the pair of values, and divide by 2. Some of the values of Z will occur more frequently because there are more values of X and Y divided by 2 that equal a particular value of Z. $1.5 = \frac{1+2}{2}$, and $1.5 = \frac{2+1}{2}$. The order of terms in the sum is important because the first number is the value we roll with the 4-sided die, and the second number is the value we roll with the 6-sided die. $2 = \frac{1+3}{2} = \frac{2+2}{2} = \frac{3+1}{2}$. $2.5 = \frac{1+4}{2} = \frac{2+3}{2} = \frac{3+2}{2} = \frac{4+1}{2}$. $3 = \frac{1+5}{2} = \frac{2+4}{2} = \frac{3+3}{2} = \frac{4+2}{2}$. $3.5 = \frac{1+6}{2} = \frac{2+5}{2} = \frac{3+4}{2} = \frac{4+3}{2}$. $4 = \frac{2+6}{2} = \frac{3+5}{2} = \frac{4+4}{2}$. $4.5 = \frac{3+6}{2} = \frac{4+5}{2}$ $5 = \frac{4+6}{2}$. Hence the pmf for a particular value of Z is the number of ways of summing a value between 1 and 4, and a value between 1 and 6, and dividing by 2 to equal Z. Therefore the tentative pmf we write in the table above is the pmf for Z. Here is a plot of the pmf:



References

[1] John Hancock. *MIT Introduction to Statistics 18.05 Reading 5a Questions*. Available at https://github.com/jhancock1975/mit-intro-to-stats-18.05/blob/prob-set-3/reading-5a-questions/reading5aQuestions.pdf (2016/3/16).

[2] Jeremy Orloff and Jonathan Bloom. *18.05 Problem Set 3, Spring 2014*. Available at https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/assignments/MIT18_05S14_ps3.pdf (Spring 2014).

[3] Jeremy Orloff and Jonathan Bloom. *Conditional Probability, Independence and Bayes Theorem Class 3, 18.05, Spring 2014*. Available at https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading3.pdf (Spring 2014).

[4] Jeremy Orloff and Jonathan Bloom. *Discrete Random Variables Class 4, 18.05, Spring 2014 Jeremy Orloff and Jonathan Bloom*. Available at https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading4a.pdf (Spring 2014).

[5] Jeremy Orloff and Jonathan Bloom. *Variance of Discrete Random Variables Class 5, 18.05, Spring 2014 Jeremy Orloff and Jonathan Bloom*. Available at https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading5a.pdf (Spring 2014).

[6] StackExchange.com user Paul Gessler. *TikZ: How to set a node on an exact position on a line?* Available at http://tex.stackexchange.com/questions/147052/tikz-how-to-set-a-node-on-an-exact-position-on-a-line (2015/2/08).

[7] StackExchange user Peter Grill. *Draw a Plot with Point - Answer*.

[8] StackOverflow.com user Shane. *Counting the number of elements with the values of x in a vector*.

[9] Texample.net user Till Tantau. *Example: Venn diagram*. Available at http://www.texample.net/tikz/examples/venn-diagram/ (2006/11/08).