Math3C: Homework 5

by Wesley Kerr (TA: Section 3a and 3b)

You should NOT look at these answers before you attempt the problems. If you do, then you're not going to learn as well. Note that these answers are not reviewed by Dr. Weisbert and/or may include some errors (as we figured out first week). If you find one or more, let me know and I'll edit the document. My email is wesleytk@ucla.edu.

Exercise 1. You toss a fair coin five times. The results of different tosses are independent events, that is, the event that you get a heads on the third toss and the event that you get a heads on the fifth toss are independent. What is the probability that you get all tails?

These are independent events, so we just need to think of one event. What is $P(T_i)$ on flip i? $\frac{1}{2}$.

Okay, so now you have 5 flips:

$$P(TTTTT) = P(T_1)P(T_2)P(T_3)P(T_4)P(T_5) = \frac{1}{2^5}$$

Exercise 2. Suppose that A and B are two events in Ω. Suppose that $P(A^c) = 0.6$, $P(A \cup B) = 0.6$, and $P(A \cap B^c) = 0.3$. Is the set $\{A, B\}$ a set of independent events?

If they are independent, then $P(A \cap B) = P(A)P(B)$. Let's figure out what each of those things are.

First the easy one: $P(A) = 1 - P(A^c) = 1 - 0.6 = 0.4$.

Then draw the Venn diagram. We then can see $P(A \cap B^c) = 0.3 = P(A) - P(A \cap B) \Rightarrow P(A \cap B) = 0.4 - 0.3 = 0.1$.

Then let's use this:

$$\begin{aligned} P(A \cup B) =& P(A \cap B^c) + P(A \cap B) + P(B \cap A^c) \\ 0.6 =& 0.3 + 0.1 + P(B \cap A^c) \\ 0.2 =& P(B \cap A^c) \end{aligned}$$

Fill that into the Venn Diagram. We can then see that $P(B) = P(B \cap A^c) + P(A \cap B) = 0.3$.

Let's assess:

$$0.1 = P(A \cap B) \stackrel{?}{=} P(A)P(B) = 0.4 \cdot 0.3$$

Note that $0.1 \neq 0.12$ so the events are NOT independent.

Exercise 3. Let Ω be a set in correspondence with the outcomes of three coin tosses with a single fair coin. Give an example of three events in Ω that are pairwise independent but not independent.

For this, the internet (wikipedia "pairwise independence") is a great place. Suppose event A and B are coin tosses of fair coins. Event C is a success if and only if exactly one of those events is a "heads". These are pairwise independent, but they are not triple-independent.

Exercise 4. A family has three children. Let A be the event that the first child is a girl. Let B be the event that at least one child is a boy. Is the set $\{A, B\}$ a set of independent events?

Let's think about B, the event that at least one child is a boy. There are 4 total options for combinations of children: $\{GG, GB, BG, BB\}$. How many of those possibilities have at least 1 boy? $\frac{3}{4}$.

Now let's think about $P(B|A)$. If the events are independent, this is $P(B) = \frac{3}{4}$. The total options if we know A is true is: $\{GG, GB\}$. Therefore, $P(B|A) = \frac{1}{2}$. That doesn't match, so they are NOT independent.

[Notice that I did this for 2 children, instead of 3, as the question states. The form should be the same (and the answer should be the same as well)]. Here's the question with 3 kids. There are now 8 equally likely options of kids:

$$\{GGG, GGB, GBG, BGG, BBG, BGB, GBB, BBB\}$$

There are $\frac{7}{8}$ that have at least one boy. Now let's look at the set of kids where the first is a girl:

$$\{GGG, GGB, GBG, GBB\}$$

How many of those have at least one boy, now just $\frac{3}{4}$. Therefore, they are NOT independent.

Exercise 5. You draw 5 cards from a standard shuffled deck. Let A be the event that you draw a flush. Let B be the event that the first card you draw is a diamond. Are A and B independent?

Let's do the $P(A|B) \stackrel{?}{=} P(A)$ trick. We know that the probability of a flush is:

$$P(\text{flush}) = \frac{4 \cdot \binom{13}{5}}{\binom{52}{5}}$$

Now $P(\text{flush}|\text{first is a diamond})$. That means that we need to choose the other 4 cards. The probability that the rest will also be diamonds:

$$P(\text{flush}|\text{first is a diamond}) = \frac{1 \cdot \binom{13}{5}}{13 \cdot \binom{51}{4} \cdot \frac{1}{5}}$$

The 5 is because, in your hand, you can order the diamond anywhere. By multiplying, we said order matter, so we need to divide by 5 to make order not matter. We now need to check if those match:

$$\begin{aligned}
\frac{4 \cdot \binom{13}{5}}{\binom{52}{5}} &\stackrel{?}{=} \frac{1 \cdot \binom{13}{5}}{13 \cdot \binom{51}{4} \cdot \frac{1}{5}} \\
\frac{4 \cdot 5! \cdot 47!}{52!} &\stackrel{?}{=} \frac{4!47!}{13 \cdot 51! \cdot \frac{1}{5}} \\
\frac{4 \cdot 5}{52} &\neq \frac{1}{13 \cdot \frac{1}{5}} \\
\frac{5}{52} &\neq \frac{5}{13 \cdot 4} \\
5 &\stackrel{\checkmark}{=} 5
\end{aligned}$$

They do match, therefore they are independent.

Exercise 6. You draw 5 cards from a standard shuffled deck. Let A be the event that you draw a straight. Let B be the event that the first card you draw is a diamond. Are A and B independent?

We know from previous homeworks that the number of possible straights are 10. There are 4^5 possible combinations of each such straight, because you have 5 suits to choose. That means that there are:

$$P(\text{straight}) = \frac{10 \cdot 4^5}{\binom{52}{5}}$$

Now $P(\text{straight}|\text{first is a diamond})$. We still have 10 possible straights because we don't know the value of the first card.There are 4 cards that we need to choose the suit for. For the denominator, we have 13 choices for the diamond. We then have $\binom{51}{4}$ choices for the remaining cards. This is:

$$P(\text{straight}|\text{first is a diamond}) = \frac{10 \cdot 4^4}{13 \cdot \binom{51}{4} \cdot \frac{1}{5}}$$

Let's check if they match.

$$\frac{10 \cdot 4^5}{\binom{52}{5}} \stackrel{?}{=} \frac{10 \cdot 4^4}{13 \cdot \binom{51}{4} \cdot \frac{1}{5}}$$

$$\frac{4 \cdot 5! \cdot 47!}{52!} \stackrel{?}{=} \frac{4!47!}{13 \cdot 51! \cdot \frac{1}{5}}$$

$$\frac{5}{52} \stackrel{\checkmark}{=} \frac{5}{13 \cdot 4}$$

Now they're independent.

Exercise 7. You draw 5 cards from a standard shuffled deck. Let A be the event that you draw a straight. Let B be the event that the first card you draw is an Ace. Are A and B independent?

This is very similar to the previous one, only instead of chosing the suit, we're choosing the value. We can then get that:

$$P(\text{straight}|\text{first is an ace}) = \frac{2 \cdot 4^5}{\frac{4}{5} \cdot \binom{51}{4}}$$

There are two such straights with aces (where it's low and high). We then need to choose the suit for 5 cards. In the denominator, we choose the suit of the ace, then choose the other 4 cards.

There's no way that we can make that match, so they are NOT independent.

Exercise 8. You roll a fair six sided die. If you roll a 1, you pick a ball from bag 1. If you roll a 2, 3, or 4, you pick a ball from bag 2. If you roll a 5, you pick a ball from bag 3. If you roll a 6, you pick a ball from bag 4. Bag 1 has 2 red balls and 3 green balls. Bag 2 has three red balls and four green balls. Bag 3 has one red ball and five green balls. Bag 4 has eight red balls and two green balls. What is the probability that you rolled a 1 if you pick a green ball?

Remember the rule of total probability. Let's use that to figure out $P(1)$. This is easy because it's the first step. It's $\frac{1}{6}$. We'd like to figure out $P(1|G)$, but that's sorta hard. It's much easier to figure out $P(G|1) = \frac{3}{5}$. From Bayes theorem and the law of total probability, we know:

$$P(1|G) = \frac{P(1 \cap G)}{P(G)} = \frac{P(G|1)P(1)}{P(G)} = \frac{\frac{3}{5}\frac{1}{6}}{P(G|1)P(1) + P(G|2,3,4)P(2,3,4) + P(G|5)P(5) + P(G|6)P(6)}$$
$$= \frac{\frac{3}{5}\frac{1}{6}}{\frac{3}{5}\frac{1}{6} + \frac{4}{7}\frac{3}{6} + \frac{5}{6}\frac{1}{6} + \frac{2}{10}\frac{1}{6}}$$

Exercise 9. A certain disease has a prevalence of one in every 100 individuals in the population. A test for this disease has a false positive rate of 10% and a false negative rate of 4%. What is the probability that an individual who tests positive for the disease has the disease? What is the probability that an individual who tests negative for the disease has the disease?

First the first question:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)}$$

We are given that $P(D) = \frac{1}{100}$ and can figure out that $P(D^c) = \frac{99}{100}$. We are given that $P(+|D^c) = 0.1$ therefore $P(-|D^c) = 0.9$. Similarly, if $P(-|D) = 0.04$ then $P(+|D) = 0.96$.

We then have all of the necessary parts:

$$P(D|+) = \frac{0.96 \cdot \frac{1}{100}}{0.96 \cdot \frac{1}{100} + 0.1 \cdot \frac{99}{100}}$$

We actually have all the necessary parts for the second question:

$$P(D^c|-) = \frac{P(-|D^c)P(D^c)}{P(-|D^c)P(D^c) + P(-|D)P(D)} = \frac{0.9 \cdot \frac{99}{100}}{0.9 \cdot \frac{99}{100} + 0.04 \cdot \frac{1}{100}}$$

Notice I answered the wrong question intially. The form, however, is correct. What you're looking for is the other term in the numerator:

$$P(D|-) = \frac{P(-|D)P(D)}{P(-|D)P(D) + P(-|D^c)P(D^c)} = \frac{0.04 \cdot \frac{1}{100}}{0.04 \cdot \frac{1}{100} + 0.9 \cdot \frac{99}{100}}$$

Exercise 10. A certain disease has a prevalence of one in every 20 individuals in the population. A test for this disease has a false positive rate of 15% and a false negative rate of 2%. What is the probability that an individual who tests positive for the disease has the disease? What is the probability that an individual who tests negative for the disease has the disease?

Yay doing the same problem with different numbers:

$$P(D) = \frac{1}{20}$$
$$P(D^c) = \frac{19}{20}$$
$$P(+|D^c) = 0.15$$
$$P(-|D^c) = 1 - 0.15 = 0.85$$
$$P(-|D) = 0.02$$
$$P(+|D) = 1 - 0.02 = 0.98$$

Plugging these numbers in:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} = \frac{0.98 \cdot \frac{1}{20}}{0.98 \cdot \frac{1}{20} + 0.85 \cdot \frac{19}{20}}$$

$$P(D^c|-) = \frac{P(-|D^c)P(D^c)}{P(-|D^c)P(D^c) + P(-|D)P(D)} = \frac{0.85 \cdot \frac{19}{20}}{0.85 \cdot \frac{19}{20} + 0.02 \cdot \frac{1}{20}}$$

Exercise 11. A certain disease has a prevalence of one in every 100 individuals in the population. A test for this disease has a false positive rate of 10% and a false negative rate of 3%. The rate of false positives and false negatives depends only on whether or not the tested individual has the disease. What is the probability that an individual who tests positive twice for the disease has the disease? What is the probability that an individual who tests negative twice for the disease has the disease?

This is a combination of the exact problem we did twice, with slight modification. While the question is a little vague, you get the impression that the additional sentence is saying that the two tests are independent (otherwise, we wouldn't know how to solve the problem). So let's run with that assumption.

Let's first redo the problem:

$$\begin{aligned} P(D) &= \frac{1}{100} \\ P(D^c) &= \frac{99}{100} \\ P(+|D^c) &= 0.1 \\ P(-|D^c) &= 1 - 0.1 = 0.9 \\ P(-|D) &= 0.03 \\ P(+|D) &= 1 - 0.03 = 0.97 \end{aligned}$$

Plugging these numbers in:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} = \frac{0.97 \cdot \frac{1}{100}}{0.97 \cdot \frac{1}{100} + 0.1 \cdot \frac{99}{100}}$$

$$P(D^c|-) = \frac{P(-|D^c)P(D^c)}{P(-|D^c)P(D^c) + P(-|D)P(D)} = \frac{0.9 \cdot \frac{99}{100}}{0.9 \cdot \frac{99}{100} + 0.03 \cdot \frac{1}{100}}$$

Let's think about the second test. It's tempting to say it's the same thing:

$$P(D| +_1 +_2) = \frac{P(+_2|D)P(D)}{P(+_2|D)P(D) + P(+_2|D^c)P(D^c)}$$

However, think about that. Is $P(D)$ the same if the first test was positive? No. Let's figure out what it should be:

$$P(D| +_1 +_2) = \frac{P(D \cap +_1 \cap +_2)}{P(+_1 +_2)} = \frac{P(+_2|D \cap +_1)P(D \cap +_1)}{P(+_1 +_2)} = \frac{P(+_2|D)P(+_1|D)P(D)}{P(+_1 +_2)}$$

Expanding the bottom by the law of total probability

$$= \frac{P(+_2|D)P(+_1|D)P(D)}{P(+_2|D)P(+_1|D)P(D) + P(+_2|D^c)P(+_1|D^c)P(D^c)}$$

Note that $P(+_2|D \cap +_1) = P(+_2|D)$ because "the false positive and negative rate dependly on whether the patient has the disease." (aka not on if the test was positive the first time.) Now all we have to do is plug into that equation:

$$P(D| +_1 +_2) = \frac{0.97 \cdot 0.97 \frac{1}{100}}{0.97 \cdot 0.97 \frac{1}{100} + 0.1 \cdot 0.1 \frac{99}{100}}$$

Now let's do the second problem, which is done in the same way.

$$\begin{aligned} P(D^c|-_2) =& \frac{P(-_2|D^c)P(-_1|D^c)P(D^c)}{P(-_2|D^c)P(-_1|D^c)P(D^c) + P(-_2|D)P(-_1|D)P(D)} \\ =& \frac{0.9 \cdot 0.9 \cdot \frac{99}{100}}{0.9 \cdot 0.9 \cdot \frac{99}{100} + 0.03 \cdot 0.03 \cdot \frac{99}{100}} \end{aligned}$$

Once again, I answered $P(D^c| -_1 -_2)$ instead of $P(D| -_1 -_2)$. They add to 1, so it's really that the other term should be in the denominator.

$$\begin{aligned} P(D|-_2) =& \frac{P(-_2|D)P(-_1|D)P(D)}{P(-_2|D^c)P(-_1|D^c)P(D^c) + P(-_2|D)P(-_1|D)P(D)} \\ =& \frac{0.03 \cdot 0.03 \cdot \frac{99}{100}}{0.9 \cdot 0.9 \cdot \frac{99}{100} + 0.03 \cdot 0.03 \cdot \frac{99}{100}} \end{aligned}$$

Exercise 12. A certain disease has a prevalence of one in every 200 individuals in the population. A test for this disease has a false positive rate of 12% and a false negative rate of 10%. The rate of false positives and false negatives depends only on whether or not the tested individual has the disease. What is the probability that an individual who tests positive for the disease and then tests negative has the disease? What is the probability that an individual who tests negative and then tests positive for the disease has the disease?

Again, redoing the problem. At this point, you should be really comfrotable with that.

$$\begin{aligned} P(D) =& \frac{1}{200} \\ P(D^c) =& \frac{199}{200} \\ P(+|D^c) =& 0.12 \\ P(-|D^c) =& 1 - 0.12 = 0.88 \\ P(-|D) =& 0.1 \\ P(+|D) =& 1 - 0.1 = 0.9 \end{aligned}$$

Plugging these numbers in:

$$\begin{aligned} P(D|+) =& \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} = \frac{0.9 \cdot \frac{1}{200}}{0.9 \cdot \frac{1}{200} + 0.88 \cdot \frac{199}{200}} \\ P(D^c|-) =& \frac{P(-|D^c)P(D^c)}{P(-|D^c)P(D^c) + P(-|D)P(D)} = \frac{0.88 \cdot \frac{199}{200}}{0.88 \cdot \frac{199}{200} + 0.1 \cdot \frac{1}{200}} \end{aligned}$$

The logic works the same way as the last problem.

$$
\begin{aligned}
P(D|-_1+_2) =& \frac{P(+_2|D)P(-_1|D)P(D)}{P(+_2|D)P(-_1|D)P(D)+P(+_2|D^c)P(-_1|D^c)P(D^c)} \\
=& \frac{0.9\cdot 0.1\cdot \frac{1}{200}}{0.9\cdot 0.1\cdot \frac{1}{200}+0.12\cdot 0.88\cdot \frac{199}{200}}
\end{aligned}
$$

To get the second part, we notice that it doesn't actually matter what's first and what is second (multiplication commutes). The false positive and negative rate are the same. Therefore:

$$
P(D|-_1+_2) = P(D|+_1-_2)
$$

Exercise 13. A certain disease has a prevalence of one in every 100 individuals in the population. A test for this disease has a false positive rate of 10% and a false negative rate of 3%. An individual who tests positive for the disease will be twice as likely to test positive in future tests. An individual who tests negative for the disease will be three times more likely to test negative again. What is the probability that an individual who tests positive twice for the disease has the disease? What is the probability that an individual who tests negative twice for the disease has the disease?

Let's redo the problem for the first test:

$$
\begin{aligned}
P(D) =& \frac{1}{100} \\
P(D^c) =& \frac{99}{100} \\
P(+|D^c) =& 0.1 \\
P(-|D^c) =& 1-0.1 = 0.9 \\
P(-|D) =& 0.03 \\
P(+|D) =& 1-0.03 = 0.97
\end{aligned}
$$

Let's plug this in:

$$
\begin{aligned}
P(D|-) =& \frac{P(-|D)P(D)}{P(-|D)P(D)+P(-|D^c)P(D^c)} = \frac{0.03\cdot\frac{1}{100}}{0.03\cdot\frac{1}{100}+0.9\cdot\frac{99}{100}} \\
P(D|+) =& \frac{P(+|D)P(D)}{P(+|D)P(D)+P(+|D^c)P(D^c)} = \frac{0.97\cdot\frac{1}{100}}{0.97\cdot\frac{1}{100}+0.1\cdot\frac{99}{100}}
\end{aligned}
$$

If the first test was positive, then $P(+|+) = 2P(+)$. That can be done in a lot of ways. Let's assume that it has no relation to the actual diagnosis of the patient, so $P(+_2|D\cap +_1) = 2P(+_1|D)$ and $P(+_2|D^c\cap +_1) = 2P(+_1|D^c)$. Let's think if that matters. For the second test, we would then have:

$$
\begin{aligned}
P(D|+_1+_2) =& \frac{P(D\cap +_1 \cap +_2)}{P(+_1+_2)} = \frac{P(+_2|D\cap +_1)P(D\cap +_1)}{P(+_1+_2)} = \frac{2P(+_1|D)P(+_1|D)P(D)}{P(+_1+_2)} \\
& \text{Expanding the bottom by the law of total probability} \\
=& \frac{2P(+_1|D)P(+_1|D)P(D)}{2P(+_1|D)P(+_1|D)P(D) + 2P(+_1|D^c)P(+_1|D^c)P(D^c)} \\
& \text{Factoring out that 2} \\
=& \frac{2}{2}\frac{P(+_1|D)P(+_1|D)P(D)}{P(+_1|D)P(+_1|D)P(D) + P(+_1|D^c)P(+_1|D^c)P(D^c)} \\
=& \frac{P(+_1|D)P(+_1|D)P(D)}{P(+_1|D)P(+_1|D)P(D) + P(+_1|D^c)P(+_1|D^c)P(D^c)}
\end{aligned}
$$

That's the same as before. Note that it doesn't matter if that factor was a 2 or if it was a 3, so we learned something for the second question. So the final answer is:

$$
P(D|+_1+_2) = \frac{0.97\cdot 0.97\cdot \frac{1}{100}}{0.97\cdot 0.97\cdot \frac{1}{100} + 0.1\cdot 0.1\cdot \frac{99}{100}}
$$

By the same logic, we get:

$$
\begin{aligned}
P(D|-_1-_2) =& \frac{P(-_1|D)P(-_2|D)P(D)}{P(-_1|D)P(-_2|D)P(D) + P(-_1|D^c)P(-_2|D^c)P(D^c)} \\
=& \frac{0.03\cdot 0.03\cdot \frac{1}{100}}{0.03\cdot 0.03\cdot \frac{1}{100} + 0.9\cdot 0.9\cdot \frac{99}{100}}
\end{aligned}
$$

Exercise 14. Disease C has a prevalence of one in every 1000 individuals in the population. Test A for this disease has a false positive rate of 10% and a false negative rate of 1%. Test B for this disease has a false positive rate of 2% and a false negative rate of 12%. The false positive and false negative rates depend only on whether or not an individual has the disease. Test A is said to have high sensitivity while Test B is a test with high specificity. You screen a population for disease C by administering Test A and then test all who tests positive using Test B. If an individual tests positive on both, what is the probability that the individual has the disease. What is the probability that an individual who tests positive twice for Disease C has the disease? What is the probability that an individual who tests tests positive first and then negative second has Disease C? What is the probability that an individual who tests positive first will test positive second?

Note how I fixed the typo about disease D that doesn't exist. We're only talking about disease C.

This question is fun (and actually relevant: we do this for HIV). In this case, the numbers are going to be unhappily asymmetric. Let's get all the parts we need

$$
\begin{aligned}
P(D) =& \frac{1}{100} \\
P(D^c) =& \frac{99}{100} \\
P(+_1|D^c) =& 0.1 \\
P(-_1|D^c) =& 1 - 0.1 = 0.9 \\
P(-_1|D) =& 0.01 \\
P(+_1|D) =& 1 - 0.01 = 0.99 \\
P(+_2|D^c) =& 0.02 \\
P(-_2|D^c) =& 1 - 0.02 = 0.98 \\
P(-_2|D) =& 0.12 \\
P(+_2|D) =& 1 - 0.12 = 0.88
\end{aligned}
$$

Let's use the equations that we know from the previous problems:

$$
\begin{aligned}
P(D| +_1 +_2) =& \frac{P(+_2|D)P(+_1|D)P(D)}{P(+_2|D)P(+_1|D)P(D) + P(+_2|D^c)P(+_1|D^c)P(D^c)} \\
=& \frac{0.88 \cdot 0.99 \cdot \frac{1}{100}}{0.88 \cdot 0.99 \cdot \frac{1}{100} + 0.02 \cdot 0.1 \cdot \frac{99}{100}} \\
P(D| +_1 -_2) =& \frac{P(-_2|D)P(+_1|D)P(D)}{P(-_2|D)P(+_1|D)P(D) + P(-_2|D^c)P(+_1|D^c)P(D^c)} \\
=& \frac{0.12 \cdot 0.99 \cdot \frac{1}{100}}{0.12 \cdot 0.99 \cdot \frac{1}{100} + 0.98 \cdot 0.1 \cdot \frac{99}{100}}
\end{aligned}
$$

The third question just uses some of the denominators that we used before:

$$
\begin{aligned}
P(+_2|+_1) =& \frac{P(+_1 \cap +_2)}{P(+_1)} = \frac{P(+_2|D)P(+_1|D)P(D) + P(+_2|D^c)P(+_1|D^c)P(D^c)}{P(+_1|D)P(D) + P(+_1|D^c)P(D^c)} \\
=& \frac{0.88 \cdot 0.99 \cdot \frac{1}{100} + 0.02 \cdot 0.1 \cdot \frac{99}{100}}{0.99 \cdot \frac{1}{100} + 0.1 \cdot \frac{99}{100}}
\end{aligned}
$$

Exercise 15. Twelve students are in a class: two are 63" tall, three are 68" tall, one is 69" tall, three are 72" tall, and one is 75" tall. Let H be the random variable that gives the height of an individual. Find a histogram for the distribution of H. What is the probability mass function of H? What is the cumulative mass function for H? What is the mean of H?

We then realize that we've given the heights of 10 students, not 12. So the total is really 10.

The histogram just plots the probabilty mass function (PMF), which let's call P, has peaks at:

$$P(63) = \frac{2}{10}$$
$$P(68) = \frac{3}{10}$$
$$P(69) = \frac{1}{10}$$
$$P(72) = \frac{3}{10}$$
$$P(75) = \frac{1}{10}$$

The cumulative mass function (CMF), g, is:

$$g(63) = \frac{2}{10}$$
$$g(68) = \frac{5}{10}$$
$$g(69) = \frac{6}{10}$$
$$g(72) = \frac{9}{10}$$
$$g(75) = \frac{10}{10}$$

The mean of H is:

$$E(H) = \sum_{h_i} p(h_i)h_i = 63 \cdot \frac{2}{10} + 68\frac{3}{10} + 69\frac{1}{10} + 72\frac{3}{10} + 75\frac{1}{10}$$

Exercise 16. You have three dice and each die has three sides. You roll all three dice. Let S be the sum of the faces on the dice. Find a histogram for the distribution of S. What is the probability mass function for S? What is the cumulative mass function for S? What is the mean of S?

This will be fun. We have to figure out the probability of each sum. The probability of each combination of 3 die rolls are: 3^3. How many ways can we get:

(1) 0

(2) 0

(3) $\#\{111\} = 1$

(4) $\#\{211\} = 3$

(5) $\#\{221, 311\} = 3 + 3 = 6$

(6) $\#\{222, 321\} = 1 + 3! = 7$

(7) $\#\{223, 331\} = 3 + 3 = 6$

(8) $\#\{233\} = 3$

(9) $\#\{333\} = 1$

To get the PMF, we divide all these numbers by 3^3. The CMF is then:

$$\begin{aligned}
g(3) =& \frac{1}{27} \\
g(4) =& \frac{4}{27} \\
g(5) =& \frac{10}{27} \\
g(6) =& \frac{17}{27} \\
g(7) =& \frac{23}{27} \\
g(8) =& \frac{26}{27} \\
g(9) =& \frac{27}{27}
\end{aligned}$$

We could do the calculation for the mean, or we could notice that the PMF is symmetric around the middle and the numbers grow linearly. That means that the mean is the middle: $E(S) = 6$.

Exercise 17. You play the following game with a friend. You each have a fair six sided die. Your friends rolls the die. You then roll the die at most two times. When you stop rolling, if you rolled a number higher than your friend's roll on your final roll, then you earn in dollars the difference between the die rolls. When you stop rolling, if you rolled a number lower than your friend's roll on your final roll, then you pay your friend in dollars the difference between the die rolls. When you stop rolling, if you rolled the same number that your friend rolled, then neither of you pay the other. What is the expected amount you earn per game if you play a large number of games? Here is an example: Your friend rolled a 3. You then roll a 4. If you stop, you get \$1. Suppose that you continue, if you roll a 1 on your second roll, then you'd have to pay your friend \$2. In that case you'd regret rolling again. However, if you had rolled a 6, you would have earned \$3.

Let's assume that you play intelligently, aka, you know the probabilities and when to roll again. If you're going to, in the long run, lose money, then you won't make that decision. That means that we need to figure out what that decision will be. This is going to be a LOT of cases.

Suppose your friend rolled a 6 (with probability $\frac{1}{6}$). You've got a $\frac{1}{6}$ probability of getting -5, -4, -3, -2, -1, and 0 (all equally likely). The average of those numbers is -\$2.5. If we get less than that, let's roll again. (see next case for more explanation). That gives us:

$$E(\$|f_6) = \frac{3}{6} \cdot (-2.5) + \frac{1}{6} \cdot (-2 - 1 + 0) = -\$1.75$$

Suppose your friend rolled a 5 (with probability $\frac{1}{6}$). If you roll, you've got an equal ($\frac{1}{6}$) probability of getting -4, -3, -2, -1, 0, 1. The average of that is -\$1.5. Intuitively, in the long run we're going to get that. That means if we get more than than that (-1, 0, 1), we'll be happy and stop rolling. If we get less than that, we'll try again and get, overall, -\$1.5. That gives us:

$$E(\$|f_5) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 0 + \frac{1}{6} \cdot (-1) + \frac{3}{6} \cdot (-1.5) = -\$0.75$$

Suppose your friend rolled a 4 (with probability $\frac{1}{6}$). If you roll, you've got an equal probability of getting -3, -2, -1, 0, 1, 2. The average of that is -\$0.5. So again, we'll roll again if we get less than that. That gives us:

$$E(\$|f_4) = \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 0 + \frac{3}{6}\cdot(-0.5) = \$0.25$$

Suppose your friend rolled a 3 (see the pattern here yet?). If you roll, you're going to get -2, -1, 0, 1, 2, 3. The average of that is \$0.5. So again, we roll again if we get less than that. That gives us:

$$E(\$|f_3) = \frac{1}{6}\cdot(3+2+1) + \frac{3}{6}\cdot 0.5 = \$1.25$$

Suppose your friend rolled a 2. If you roll, you're going to get -1, 0, 1, 2, 3, 4. The average of that is \$1.5. So again, we roll again if we get less than that. This gives us:

$$E(\$|f_2) = \frac{1}{6}\cdot(4+3+2) + \frac{3}{6}\cdot 1.5 = \$2.25$$

Suppose your friend rolled a 1. If you roll, you're going to get 0, 1, 2, 3, 4, 5. The average of that is \$2.5. This gives us:

$$E(\$|f_1) = \frac{1}{6}\cdot(5+4+3) + \frac{3}{6}\cdot 2.5 = \$3.25$$

These are all equally likely, so the final amount is the average of them, or:

$$E(\$) = \sum_{i=1}^{6} E(\$|f_i)P(f_i) = \$0.75$$

Emailed question. Disease C has prevalence 1/10 in a population and Disease H has prevalence 1/4. A test for Disease C has 5% false positives and 10% false negatives. A test for Disease H has 8

Let's assume that the two diseases and tests are independent (not given, but we need that to solve them). First think about disease C. We know:

$$\begin{aligned}
P(C) =& \frac{1}{10} \\
P(C^c) =& \frac{9}{10} \\
P(C^c|+_C) =& 0.05 \\
P(C|+_C) =& 1-0.05 = 0.95 \\
P(C|-_C) =& 0.10 \\
P(C^c|-_C) =& 1-0.10 = 0.90
\end{aligned}$$

Now let's think about disease H. We know:

$$\begin{aligned}
P(H) =& \frac{1}{4} \\
P(H^c) =& \frac{3}{4} \\
P(H^c|+_H) =& 0.08 \\
P(H|+_H) =& 1-0.08 = 0.92 \\
P(H|-_H) =& 0.02 \\
P(H^c|-_H) =& 1-0.02 = 0.98
\end{aligned}$$

Now we're asked for:

$$\begin{aligned}P(\text{at least } C \text{ or } H| -_C -_H) =& P(C \cap H| -_C -_H) + P(C^c \cap H| -_C -_H) + P(C \cap H^c| -_C -_H)\\ =& P(C|-_C)P(H|-_H) + P(C^c|-_C)P(H|-_H) + P(C|-C)P(H^c|-_H)\end{aligned}$$

We know the form for all of these necessary parts:

$$\begin{aligned}P(C|-_C) =& \frac{P(-_C|C)P(C)}{P(-_C|C)P(C) + P(-_C|C^c)P(C^c)}\\ =& \frac{0.10 \cdot \frac{1}{10}}{0.10 \cdot \frac{1}{10} + 0.90 \cdot \frac{9}{10}}\\ P(C^c|-_C) =& \frac{P(-_C|C^c)P(C^c)}{P(-_C|C)P(C) + P(-_C|C^c)P(C^c)}\\ =& \frac{0.90 \cdot \frac{9}{10}}{0.10 \cdot \frac{1}{10} + 0.90 \cdot \frac{9}{10}}\\ P(H|-_H) =& \frac{P(-_H|H)P(H)}{P(-_H|H)P(H) + P(-_H|H^c)P(H^c)}\\ =& \frac{0.02\frac{1}{4}}{0.02\frac{1}{4} + 0.98\frac{3}{4}}\\ P(H^c|-_H) =& \frac{P(-_H|H^c)P(H^c)}{P(-_H|H)P(H) + P(-_H|H^c)P(H^c)}\\ =& \frac{0.98\frac{3}{4}}{0.02\frac{1}{4} + 0.98\frac{3}{4}}\end{aligned}$$

Plugging each of those in, we get:

$$\begin{aligned}P(\text{at least } C \text{ or } H| -_C -_H) =& \frac{0.10 \cdot \frac{1}{10}}{0.10 \cdot \frac{1}{10} + 0.90 \cdot \frac{9}{10}} \frac{0.02\frac{1}{4}}{0.02\frac{1}{4} + 0.98\frac{3}{4}}\\ &+ \frac{0.10 \cdot \frac{1}{10}}{0.10 \cdot \frac{1}{10} + 0.90 \cdot \frac{9}{10}} \frac{0.98\frac{3}{4}}{0.02\frac{1}{4} + 0.98\frac{3}{4}}\\ &+ \frac{0.90 \cdot \frac{9}{10}}{0.10 \cdot \frac{1}{10} + 0.90 \cdot \frac{9}{10}} \frac{0.02\frac{1}{4}}{0.02\frac{1}{4} + 0.98\frac{3}{4}}\end{aligned}$$