Problem Sheet 3 (Solutions)

MATH1710 Probability and Statistics I

University of Leeds, 2022-23

A: Short questions

A1. Consider dealing two cards (without replacement) from a pack of cards. Which of the following pairs of events are independent?

(a) "The first card is a Heart" and "The first card is Red".

> *Solution.* We have
>
> $$\begin{aligned} \mathbb{P}(\text{first Heart}) &= \frac{13}{52} = \frac{1}{4} \\ \mathbb{P}(\text{first Red}) &= \frac{26}{52} = \frac{1}{2} \\ \mathbb{P}(\text{first Heart and first Red}) &= \mathbb{P}(\text{first Heart}) = \frac{1}{4}. \end{aligned}$$
>
> So $\mathbb{P}(\text{first Heart and first Red}) \neq \mathbb{P}(\text{first Heart})\,\mathbb{P}(\text{first Red})$, and the events are not independent.

(b) "The first card is a Heart" and "The first card is a Spade".

> *Solution.* We have
>
> $$\begin{aligned} \mathbb{P}(\text{first Heart}) &= \frac{13}{52} = \frac{1}{4} \\ \mathbb{P}(\text{first Spade}) &= \frac{13}{52} = \frac{1}{4} \\ \mathbb{P}(\text{first Heart and first Spade}) &= 0. \end{aligned}$$
>
> So $\mathbb{P}(\text{first Heart and first Spade}) \neq \mathbb{P}(\text{first Heart})\,\mathbb{P}(\text{first Spade})$, and the events are not independent.

(c) "The first card is a Heart" and "The first card is an Ace".

> *Solution.* We have
>
> $$\begin{aligned} \mathbb{P}(\text{first Heart}) &= \frac{13}{52} = \frac{1}{4} \\ \mathbb{P}(\text{first Ace}) &= \frac{4}{52} = \frac{1}{13} \\ \mathbb{P}(\text{first Heart and first Ace}) &= \mathbb{P}(\text{first Ace of Hearts}) = \frac{1}{52}. \end{aligned}$$
>
> So $\mathbb{P}(\text{first Heart and first Ace}) = \mathbb{P}(\text{first Heart})\,\mathbb{P}(\text{first Ace})$, and the events are independent.

(d) "The first card is a Heart" and "The second card is a Heart".

Solution. We have

$$\begin{aligned}\mathbb{P}(\text{first Heart}) &= \frac{13}{52} = \frac{1}{4}\\ \mathbb{P}(\text{second Heart}) &= \frac{13}{52} = \frac{1}{4}\\ \mathbb{P}(\text{first Heart and second Heart}) &= \frac{13 \times 12}{52 \times 51} = \frac{1}{17}\end{aligned}$$

So $\mathbb{P}(\text{first Heart and second Heart}) \neq \mathbb{P}(\text{first Heart})\,\mathbb{P}(\text{second Heart})$, and the events are not independent.

(e) "The first card is a Heart" and "The second card is an Ace".

Solution. We have

$$\begin{aligned}\mathbb{P}(\text{first Heart}) &= \frac{13}{52} = \frac{1}{4}\\ \mathbb{P}(\text{second Ace}) &= \frac{4}{52} = \frac{1}{13}\\ \mathbb{P}(\text{first Heart and second Ace}) &= \frac{12 \times 4 + 1 \times 3}{52 \times 51} = \frac{51}{52 \times 51} = \frac{1}{52}\end{aligned}$$

Here, the 12×4 counted "a non-Ace Heart, followed by an Ace", while the 1×3 counted "the Ace of Hearts, followed by a non-Heart Ace". So $\mathbb{P}(\text{first Heart and second Ace}) = \mathbb{P}(\text{first Heart})\,\mathbb{P}(\text{second Ace})$, and the events are independent.

A2. Consider rolling two dice. Let A be the event that the first roll is even, let B be the event that the second roll is even, and let C be the event that the total score is even. You may assume the dice rolls are independent; so, in particular, events A and B are independent.

(a) Are A and C independent?

Solution. Let us first note that $\mathbb{P}(A) = \mathbb{P}(B) = \frac{3}{6} = \frac{1}{2}$. It's also the case that $\mathbb{P}(C) = \frac{18}{36}$, for example by counting the even outcomes out of the 36 equally likely possibilities.

We need to test is $\mathbb{P}(A \cap C) = \mathbb{P}(A)\,\mathbb{P}(C) = \frac{1}{4}$ or not. By counting from the 36 possibilities, we see that indeed $\mathbb{P}(A \cap C) = \frac{9}{36} = \frac{1}{4}$. Alternatively, we could note that $\mathbb{P}(C \mid A) = \mathbb{P}(B) = \frac{1}{2}$, since if the first dice is even, the second must be also even to get an even total. Then $\mathbb{P}(A \cap C) = \mathbb{P}(A)\,\mathbb{P}(C \mid A) = \frac{1}{4}$.

So the events are indeepdent.

(b) Are B and C independent?

Solution. Yes. The solution is essentially identical to part (a).

(c) Is it true that $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C)$?

Solution. By checking the 36 possibilities, one sees that

$$\mathbb{P}(A \cap B \cap C) = \frac{9}{36} = \frac{1}{4} \neq \frac{1}{8} = \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C).$$

Alternatively, note that the total being even is certain if both dice rolls are certain, so

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B)\,\mathbb{P}(C \mid A \cap B) = \frac{1}{4} \times 1 = \frac{1}{4},$$

to get the same result.

Group feedback: This shows that just because events are "pairwise independent", it does not mean they are "mutually independent".

A3. Consider the random variable X with the following PMF:

x	-1	0	0.5	1	2
$p(x)$	0.1	0.3	0.3	0.2	0.1

Find the expectation and variance of X.

Solution. For the expectation,

$$\mathbb{E}X = -1 \times 0.1 + 0 \times 0.3 + 0.5 \times 0.3 + 1 \times 0.2 + 2 \times 0.1 = 0.45.$$

For the variance, we start with

$$\mathbb{E}X^2 = (-1)^2 \times 0.1 + 0^2 \times 0.3 + 0.5^2 \times 0.3 + 1^2 \times 0.2 + 2^2 \times 0.1 = 0.775.$$

Then, using the computational formula,

$$\mathrm{Var}(X) = \mathbb{E}X^2 - \mu^2 = 0.775 - 0.45^2 = 0.5725.$$

A4. Consider the random variable X with the following PMF:

x	1	2	4	5	a
$p(x)$	0.1	0.2	0.1	b	0.1

This random variable has $\mathbb{E}X = 4.3$. Find the values of a and b.

Solution. First, a PMF must sum to 1, so

$$1 = 0.1 + 0.2 + 0.1 + b + 0.1,$$

so $b = 0.5$.

Second, the expectation is

$$\mathbb{E}X = 1 \times 0.1 + 2 \times 0.2 + 4 \times 0.1 + 5b + 0.1a = 3.6 + 0.1a = 4.3.$$

So $a = 7$.

A5. A temperature T_C measured in degrees Celsius can be converted to a temperature T_F in degrees Fahrenheit using the formula $T_F = \frac{9}{5}T_C + 32$.

The average daily maximum temperature in Leeds in July is 19.0 °C. The variance of the daily maximum temperature measured in degrees Celsius is 10.4.

(a) What is the average daily maximum temperature in degrees Fahrenheit?

Solution. By linearity of expectation,

$$\mathbb{E}T_F = \mathbb{E}\left(\tfrac{9}{5}T_C + 32\right) = \tfrac{9}{5}\mathbb{E}T_C + 32.$$

So the answer is $\frac{9}{5} \times 19.0 + 32 = 66.2$ °F.

(b) What is the variance of the daily maximum temperature when measured in degrees Fahrenheit?

Solution. For the variance,

$$\operatorname{Var}(T_F) = \operatorname{Var}\left(\tfrac{9}{5}T_C + 32\right) = \left(\tfrac{9}{5}\right)^2 \operatorname{Var}(T_C) = \tfrac{81}{25}\operatorname{Var}(T_C).$$

So the answer is $\frac{81}{25} \times 10.4 = 33.7$.

B: Long questions

B1. Suppose A and B are independent events. Show that A and B^{c} are also independent events.

Solution. We need to show that

$$\mathbb{P}(A \cap B^{\mathsf{c}}) = \mathbb{P}(A)\,\mathbb{P}(B^{\mathsf{c}}). \tag{$*$}$$

Note that

$$A = (A \cap B) \cup (A \cap B^{\mathsf{c}}),$$

and the union is disjoint, so by Axiom 3,

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^{\mathsf{c}}).$$

Hence, the left-hand side of $(*)$ is

$$\begin{aligned}\mathbb{P}(A \cap B^{\mathsf{c}}) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) - \mathbb{P}(A)\,\mathbb{P}(B) \\ &= \mathbb{P}(A)\big(1 - \mathbb{P}(B)\big),\end{aligned}$$

where, in the second line, crucially we used the fact that A and B are independent to replace $\mathbb{P}(A \cap B)$ by $\mathbb{P}(A)\,\mathbb{P}(B)$.

The right-hand side of $(*)$ is

$$\mathbb{P}(A)\,\mathbb{P}(B^{\mathsf{c}}) = \mathbb{P}(A)\big(1 - \mathbb{P}(B)\big),$$

where we've used the complement rule $\mathbb{P}(B^{\mathsf{c}}) = 1 - \mathbb{P}(B)$.

Hence, we've shown the left- and right-hand sides of $(*)$ are equal, and we are done.

::::: {.myq} **B2.** You are dealt a hand of 13 cards from a 52-card deck. Let $E_{\mathrm{A}}, E_{\mathrm{K}}, E_{\mathrm{Q}}, E_{\mathrm{J}}$ respectively be the events that your hand contains the Ace, King, Queen and Jack of Spades.

(a) What is $\mathbb{P}(E_{\mathrm{A}})$, the probability that your hand contains the Ace of Spades?

Solution. There are 52 cards of which 13 will end up in my hand, so $\mathbb{P}(E_{\mathrm{A}}) = \frac{13}{52}$.

(b) Explain why $\mathbb{P}(E_{\mathrm{K}} \mid E_{\mathrm{A}}) = \frac{12}{51}$.

Solution. Given I have the Ace of Spades, there are $52 - 1 = 51$ cards left available, of which $13 - 1 = 12$ will end up in my hand, so $\mathbb{P}(E_{\mathrm{K}} \mid E_{\mathrm{A}}) = \frac{12}{51}$.

(c) Using the chain rule, calculate the probability that your hand contains all four of the Ace, King, Queen and Jack of Spades.

Solution. Continuing the logic of part (b), we have

$$\mathbb{P}(E_{\mathrm{Q}} \mid E_{\mathrm{A}} \cap E_{\mathrm{K}}) = \frac{11}{50} \qquad \mathbb{P}(E_{\mathrm{J}} \mid E_{\mathrm{A}} \cap E_{\mathrm{K}} \cap E_{\mathrm{Q}}) = \frac{10}{49}.$$

Using the chain rule,

$$\begin{aligned} P(E_{\mathrm{A}} \cap E_{\mathrm{K}} \cap E_{\mathrm{Q}} \cap E_{\mathrm{J}}) &= \mathbb{P}(E_{\mathrm{A}})\,\mathbb{P}(E_{\mathrm{K}} \mid E_{\mathrm{A}})\,\mathbb{P}(E_{\mathrm{Q}} \mid E_{\mathrm{A}} \cap E_{\mathrm{K}})\,\mathbb{P}(E_{\mathrm{J}} \mid E_{\mathrm{A}} \cap E_{\mathrm{K}} \cap E_{\mathrm{Q}}) \\ &= \frac{13}{52} \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49} = \frac{13 \times 12 \times 11 \times 10}{52 \times 51 \times 50 \times 49}, \end{aligned}$$

which is the same as we got in lectures.

(d) Check that your answer agrees with the answer we found by classical probability methods in Example 6.4 in Lecture 6. Which method do you prefer?

Solution. Personally, I slightly prefer this answer – it seems more obvious how the answer relates to the method, whereas in lectures a lot of terms in a ratio of binomial coefficients "magically" cancelled out. Your mileage may vary.

B3. Soldiers are asked about their use of illegal drugs, using a so-called "randomised survey". Each soldier is handed a deck of three cards, picks one of the three cards at random, and responds according to what the card says. The three cards say:

1. "Say 'Yes.' "
2. "Say 'No.' "
3. "Truthfully answer the question 'Have you taken any illegal drugs in the past 12 months?' "

(a) What are some advantages or disadvantages of performing the experiment this way?

Solution. The main advantage is that it seems likely that a soldier might want to lie in answer to a "straight question", given that if their superiors discovered they had taken illegal drugs, there could be very serious consequences. This method allows a certain "plausible deniability": just because the soldier answers "Yes", we cannot know for sure whether they have taken illegal drugs or merely picked the "Yes" card. Thus we might hope to get more honest answers this way. Perhaps you can think of other advantages.

There could be disadvantages. The complicated set-up of the experiment could lead to the subjects (or experimenters) making an error. The scientists good be "lulled into a false sense of security" of thinking they get fully honest answers, when soldiers picking card 3 might still choose to lie. Perhaps you can think of other disadvantages.

(b) Suppose that 40% of soldiers respond "Yes". What is the likely proportion of soldiers who have taken illegal drugs in the past 12 months.

Solution. Let C_1, C_2, C_3 be the events that a soldier picks cards 1, 2, or 3 respectively, which have probabilities $\mathbb{P}(C_1) = \mathbb{P}(C_2) = \mathbb{P}(C_3) = \frac{1}{3}$ and make up a partition. Let Y be the event that the soldier answers yes. We know that $\mathbb{P}(Y \mid C_1) = 1$, $\mathbb{P}(Y \mid C_2) = 0$ and $\mathbb{P}(Y \mid C_3) = \mathbb{P}(D)$, where $\mathbb{P}(D)$, which we want to find, is the proportion of soldiers who have taken illegal drugs in the past 12 months. We are also told that $\mathbb{P}(Y) = 0.4$.

The law of total probability tells us that

$$\mathbb{P}(Y) = \mathbb{P}(C_1)\,\mathbb{P}(Y \mid C_1) + \mathbb{P}(C_2)\,\mathbb{P}(Y \mid C_2) + \mathbb{P}(C_3)\,\mathbb{P}(Y \mid C_3).$$

With the information we have, we get

$$0.4 = \tfrac{1}{3} \times 1 + \tfrac{1}{3} \times 0 + \tfrac{1}{3}\,p = \tfrac{1}{3} + \tfrac{1}{3}\,p.$$

Solving this gives $p = \frac{1}{5} = 20\%$.

(c) If a soldier responds "Yes", what is the probability that the soldier has taken illegal drugs in the past 12 months.

Solution. This is asking for $\mathbb{P}(D \mid Y)$. Another one for Bayes theorem:

$$\mathbb{P}(D \mid Y) = \frac{\mathbb{P}(D)\mathbb{P}(Y \mid D)}{\mathbb{P}(Y)}.$$

From the question we know that $\mathbb{P}(Y) = 0.4$. From part (a) we know that $\mathbb{P}(D) = 0.2$. We also know that $\mathbb{P}(Y \mid D) = \frac{2}{3}$, as the soldier will answer Yes is they pick either cards 1 or 3. Hence

$$\mathbb{P}(D \mid Y) = \frac{0.2 \times \frac{2}{3}}{0.4} = \frac{1}{3}.$$

B4. A random variable X_n is said to follow the *discrete uniform distribution* on $\{1, 2, \dots, n\}$ if each of the n values in that set $\{1, 2, \dots, n\}$ is equally likely.

(a) Show that the expectation of X_n is $\mathbb{E}X_n = \dfrac{n+1}{2}$.

Solution. We have $p(x) = \frac{1}{n}$ for $x = 1, 2, \dots, n$. So the expectation is

$$\mathbb{E}X = \sum_{x=1}^{n} x\,\frac{1}{n} = \frac{1}{n}\sum_{x=1}^{n} x = \frac{1}{n}\,\frac{n(n+1)}{2} = \frac{n+1}{2}.$$

(b) Find the variance of X_n.

Solution. It turns out to be much easier to use the computational formual $\mathrm{Var}(X) = \mathbb{E}X^2 - \mu^2$.

First,
$$\mathbb{E}X^2 = \sum_{x=1}^{n} x^2\,\frac{1}{n} = \frac{1}{n}\sum_{x=1}^{n} x^2 = \frac{1}{n}\,\frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}.$$
Then using $\mu = (n+1)/2$ from part (a), we have
$$\mathrm{Var}(X) = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{(n+1)(4n+2-3n-3)}{12} = \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}$$

(c) Let Y be a discrete uniform distribution on $b-a+1$ values $\{a, a+1, a+2, \dots, b-1, b\}$, for integers a and b with $a < b$. Using parts (a) and (b), but without calculating any sums directly, find the expectation and variance of Y.

*[**Note:** "$b-a+1$ values" is correct, but this was wrong earlier.]*

Solution. If we take $n = b-a+1$, then Y has the same distribution as $X_n + (a-1)$. This is because $x = 1$ maps to $y = 1 + (a-1) = a$; $x = 2$ maps to $y = 2 + (a-1) = a+1$; and so on; up to $x = n$ mapping to $y = n + (a-1) = (b-a+1) + (a-1) = b$. So the ranges match up perfectly.

Thus we have
$$\mathbb{E}Y = \mathbb{E}\big(X_{b-a+1} + (a-1)\big) = \mathbb{E}X_{b-a+1} + (a-1) = \frac{b-a+1+1}{2} + (a-1) = \frac{a+b}{2}$$
and
$$\mathrm{Var}(Y) = \mathrm{Var}\left(X_{b-a+1} + (a-1)\right) = \mathrm{Var}(X_{b-a+1}) = \frac{(b-a+1)^2-1}{12}.$$
You can rearrange the variance a bit if you like, but it doesn't really get any nicer.

You may use without proof the standard results
$$\sum_{x=1}^{n} x = \frac{n(n+1)}{2} \qquad \sum_{x=1}^{n} x^2 = \frac{n(n+1)(2n+1)}{6}.$$

B5. A gambling game works as follows. You keep tossing a fair coin until you first get a Head. If the first Head comes up on the nth coin toss, then you win 2^n pounds.

(a) What is the probability that the first Head is seen on the nth toss of the coin?

Solution. This happens if the first $n-1$ tosses are Tails, with probability $(\frac{1}{2})^{n-1}$, them the nth toss is Heads, with probability $\frac{1}{2}$. Altogether, this is $(\frac{1}{2})^{n-1} \times \frac{1}{2} = (\frac{1}{2})^n$.

(b) Show that the expected winnings from playing this game are infinite.

Solution. The expected winnings are
$$\sum_{n=1}^{\infty} 2^n \times \mathbb{P}(\text{first Head on } n\text{th toss}) = \sum_{n=1}^{\infty} 2^n \times \left(\tfrac{1}{2}\right)^n = \sum_{n=1}^{\infty} 1 = \infty$$

(c) The "St Petersburg paradox" refers to the observation that, despite the expected winnings from this game being infinite, few people would be prepared to play this game for, say, £100, and almost no one for £1000. Discuss a few possible "resolutions" to this paradox which could explain why people are unwilling to play this game despite seemingly having infinite expected winnings.

Discussion. One possibility is:

- The people are being irrational, and in fact *should* play the game for £1000.

but I'm not sure anyone *really* thinks that.

Some other possible explanations include:

- The expectation is only infinite if you really could win an extraordinarily large amount of money. Suppose that the person offering the game only has £2^{20}, or just over £1 million. In that case, if the first 20 tosses are all Tails, the opponent gives you all £2^{20} then declares bankruptcy and the game stops. In this more realistic case, your expected winnings are only

$$\sum_{n=1}^{20} 2^n \times \left(\tfrac{1}{2}\right)^n + 2^{20} \times \left(\tfrac{1}{2}\right)^{20} = \sum_{n=1}^{20} 1 + 1 = 21,$$

  or £21; a more reasonable price to pay to play the game.
- The amount of benefit (or "utility") one gets from winning a large amount of money might not be directly proportional to the amount. For example, £200 million might be very nice, but it's not *twice* as nice as £100 million – after all, what else could you really do with the second £100 million. Perhaps the utility of £m scales more logarithmically than linearly, like $\log_2 m$ in some appropriate "happiness units" In that case, the expected *utility* from the game is

$$\sum_{n=1}^{\infty} \log_2(2^n) \times \left(\tfrac{1}{2}\right)^n = \sum_{n=1}^{\infty} n \times \left(\tfrac{1}{2}\right)^n = 2,$$

  happiness units, and you might be willing to pay 2 happiness-units-worth of money to play.
- Normal advice to play games with positive expected winnings only really applies if you can play the game many times (or very similar games). For repeated games, the expected winnings can be interpreted as "the winnings you are likely to get in the long run". For one-off highly unusual games, this doesn't hold, so one needs a different criterion to decide whether to play. (If I was allowed to play this game a million times for £100 a round, but didn't have to settle the money until all one million games had finished, then I would strongly consider playing.)

You can probably come up with other explanations of your own too.

C: Assessed questions

C1. A computer spam filter is 98% effective at sending spam emails to my junk folder, but will also incorrectly send 1% of legitimate emails to my junk folder. Suppose that 1 in 10 emails are spam. What proportion of emails in my junk folder are actually legitimate emails? Explain your solution fully.

> *Hints.* Some general advice:
>
> - Give the events in the question names.
> - Note if any of the events make up a partition.
> - Write down what probabilities (or conditional probabilities) you are told in the question.
> - What probability (or conditional probability) do you want to find out?
> - Are any of the results we have learned in the course about conditional probability useful here?

C2. Let X be a random variable.

(a) Let $Y = aX$ be another random variable. What is $\mathbb{E}Y$, in terms of $\mu = \mathbb{E}X$?

> *Hint.* We have seen this in Lecture 10.

(b) Using part (a), show that $\operatorname{Var}(aX) = a^2 \operatorname{Var}(X)$.

> *Hint.* Use the definition of variance (or the computational formula), then use properties like the linearity of expectation. Don't actually compute any sums!

(c) Prove that $\operatorname{Var}(X + b) = \operatorname{Var}(X)$.

> *Hint.* If $Z = X + b$, what is $\mathbb{E}Z$ in terms of μ, to start with?