Rensselaer Polytechnic Institute
Department of Electrical, Computer, and Systems Engineering
ECSE-2500: Engineering Probability, Spring 2023
Exam 1 Solutions

1. (20 points.) We roll a fair 4-sided die labeled {1, 2, 3, 4} twice. Let A be the number on the first roll, B be the number on the second roll, and D be $A - B$.

   (a) (10 points.) What is the sample space S for the random variable D? How many possible outcomes are there?

   We can make a table of the possible outcomes for A, B, and D:

		B		
	1	2	3	4
1	0	-1	-2	-3
A 2	1	0	-1	-2
3	2	1	0	1
4	3	2	1	0

   From this we can see that the sample space for D is $S = \{-3, -2, -1, 0, 1, 2, 3\}$ and that there are 7 possible outcomes.
   A common mistake was to list the 16 possibilities for (A, B) for the sample space and say there were 16 outcomes.

   Grading criteria: 10 points in total
   -0 point: correct
   -2 point: minor error in sample space or possible outcomes
   -5 point: large errors in sample space or possible outcomes
   -10 point: incorrect

   (b) (10 points.) Are the events "the first roll is less than 3" and "the absolute value of D equals 1" independent? Clearly explain your reasoning.

   Let the event G be "the first roll is less than 3" and the event H be "the absolute value of D equals 1". Since the die is fair, $P(G) = P(A = 1 \text{ or } 2) = \frac{1}{2}$. The die being fair also means that all 16 entries in the table for D have the same probability of $\frac{1}{16}$, so $P(H) = \frac{6}{16} = \frac{3}{8}$.
   Now we have to compute $P(G \cap H)$, which corresponds to the possibilities $(A, B) = (2, 1)$, $(A, B) = (1, 2)$, and $(A, B) = (2, 3)$. Since there are 3 outcomes, $P(G \cap H) = \frac{3}{16}$.
   By the formal definition of independence, since $P(G \cap H) = \frac{3}{16} = \frac{1}{2} \cdot \frac{3}{8} = P(G)P(H)$, we can conclude that the two events are independent.

   Grading criteria: 10 points in total
   -0 point: correct
   -2 point: prob of the first event is missing or incorrect
   -2 point: prob of the second event is missing or incorrect

-2 point: prob of the intersection of two events is missing or incorrect
-2 point: the equation of independence check is missing or incorrect

2. (20 points.) Kendall Roy regularly attempts to take over his father's company. On each attempt, he has probability $\frac{1}{5}$ of succeeding, and every attempt is independent. Let X be the number of attempts required for him to finally take over the company.

   (a) (2 points.) What type of random variable is X?

   This is a geometric random variable (i.e., counting the number of independent trials until success).

   Grading criteria: 2 points in total
   -0 point: correct
   -2 point: incorrect

   (b) (5 points.) Compute $P(X > 3)$. Your answer should be a simple fraction or decimal.

$$\begin{aligned}
P(X > 3) &= \sum_{k=4}^{\infty} p_X(k) \\
&= \sum_{k=4}^{\infty} \left(\frac{4}{5}\right)^{k-1} \frac{1}{5} \\
&= \frac{1}{5}\left(\frac{4}{5}\right)^3 \sum_{k=0}^{\infty} \left(\frac{4}{5}\right)^k \\
&= \frac{1}{5}\left(\frac{4}{5}\right)^3 \frac{1}{1-\frac{4}{5}} \\
&= \left(\frac{4}{5}\right)^3 \\
&= \frac{64}{125} = 0.512
\end{aligned}$$

   An easier approach is to compute

$$\begin{aligned}
P(X > 3) &= 1 - P(X = 1) - P(X = 2) - P(X = 3) \\
&= 1 - \frac{1}{5} - \frac{4}{25} - \frac{16}{125}
\end{aligned}$$

   A common mistake was to use the wrong formula for the geometric random variable; since we have to make at least one takeover attempt, we need to use the version that starts at 1. Another common error was to calculate $P(X > 3) = 1 - P(X < 3)$, which ignores the possibility of $P(X = 3)$.

   Grading criteria: 5 points in total
   -0 point: correct
   -2 point: Wrong formula for geometric RV
   -2 point:missing P(x=3)
   -2 point: incorrect answer

(c) (8 points.) Sketch and determine the values of the conditional PMF $p_{X \mid X \le 3}(k)$.

From the above, we can compute that $P(X \le 3) = 1 - P(X > 3) = \frac{61}{125} = 0.488$. We can see that the conditional PMF only has three non-zero values at $k = \{1,2,3\}$ which are computed as

$$\begin{aligned}
p_{X \mid X\le 3}(1) &= \tfrac{p_X(1)}{P(X\le 3)} &= \tfrac{1/5}{61/125} &= \tfrac{25}{61} &= 0.410 \\
p_{X \mid X\le 3}(2) &= \tfrac{p_X(2)}{P(X\le 3)} &= \tfrac{4/25}{61/125} &= \tfrac{20}{61} &= 0.328 \\
p_{X \mid X\le 3}(3) &= \tfrac{p_X(3)}{P(X\le 3)} &= \tfrac{16/125}{61/125} &= \tfrac{16}{61} &= 0.262
\end{aligned}$$

A very common problem was to just give the raw values of $p_X(1)$, etc. instead of normalizing to compute the conditional probability. Remember, all the values of the conditional PMF have to sum to 1! You also got dinged if you didn't draw your sketch with delta functions (arrows).

Grading criteria: We give full mark of Problem 2(c-d) to all students, since they are not covered in lectures. For students who already did any of them correctly in Exam 1, we will give a bonus 5 points.

(d) (5 points.) Compute $P(X \le 2 \mid X \le 3)$.

We simply add up the values for $k = 1,2$ in the conditional PMF above to obtain

$$P(X \le 2 \mid X \le 3) = P(X = 1 \mid X \le 3) + P(X = 2 \mid X \le 3) = \frac{25}{61} + \frac{20}{61} = \frac{45}{61} = 0.738$$

Note that this doesn't have anything to do with the memoryless property of the geometric RV. While it's true that $P(X > 4 \mid X > 3) = P(X > 4 - 3) = P(X > 1)$, it doesn't work when we have less-than's.

Grading criteria: We give full mark of Problem 2(c-d) to all students, since they are not covered in lectures. For students who already did any of them correctly in Exam 1, we will give a bonus 5 points.

3. (20 points.) Paxton Hall-Yoshida decides to sell his car to a mechanic at an initial price of \$1500. The mechanic agrees, but will reduce the price by \$400 for every problem she finds. Suppose that the number of problems is modeled by a Poisson random variable with mean 2.

(a) (10 points.) Compute the probability that Paxton makes money by selling the car (i.e., that the sale price is greater than \$0).

Let Y be the sale price of the car. We can compute a table of X, the corresponding value of Y, and the PMF. This looks like a truncated Poisson distribution since after we find 4 errors, the sale price is \$0.

# errors	sale price	probability		
0	\$1500	$\frac{2^0}{0!}e^{-2}$	$= e^{-2}$	$= 0.135$
1	\$1100	$\frac{2^1}{1!}e^{-2}$	$= 2e^{-2}$	$= 0.271$
2	\$700	$\frac{2^2}{2!}e^{-2}$	$= 2e^{-2}$	$= 0.271$
3	\$300	$\frac{2^3}{3!}e^{-2}$	$= \frac{4}{3}e^{-2}$	$= 0.180$
4	\$0	$\dots$		$= 0.143$

We can compute that the probability that the sale price is above \$0 is the sum of the first 4 values of the Poisson PMF, equal to $\frac{19}{3}e^{-2} = 0.857$.
Many people forgot to include either the possibilities $X = 0$ or $X = 3$. There were also some misconceptions about what the correct PMF was (e.g., $\alpha = \frac{1}{2}$ instead of 2).

Grading criteria: 10 points in total
-0 point: correct
-2 point: prob of 0 error (price=1500) is missing or incorrect
-2 point: prob of 1 error (price=1100) is missing or incorrect
-2 point: prob of 2 error (price=700) is missing or incorrect
-2 point: prob of 3 error (price=300) is missing or incorrect
-2 point: final prob (sum of the four probs above) is missing or incorrect

(b) (10 points.) Compute the expected sale price. (Once the sale price hits 0, it stays at 0... Paxton won't pay the mechanic to take the car!)

The expected value is computed as

$$E(Y) = (\$1500)(e^{-2}) + (\$1100)(2e^{-2}) + (\$700)(2e^{-2}) + (\$300)(\frac{4}{3}e^{-2}) = \$5500e^{-2} = \$744$$

An extremely common error here was to assume that since $E(X) = 2$, that we can simply compute $E(Y) = \$1500 - \$400E(X) = \$700$. The error in thinking here is that $E(X)$ takes into account the probabilities that we have 4 or more errors. That is, you're not using the information that once we have 4 or more errors, the car is worth \$0. Instead, you could be potentially "selling" the car for negative dollars!
Please don't ask for points back if you made this mistake!

Grading criteria: 10 points in total
-0 point: correct
-3 point: expectation formula is incorrect
-3 point: final result is incorrect or missing
-5 point: completely wrong or missing expectation formula

-10 point: blank or totally incorrect

4. (15 points.) An EDM DJ's record case contains singles with beats per minute (BPM) that have the following probability mass function, where X is the random variable representing the BPM of a randomly chosen record.

x	120	130	140	150	160
$p_X(x)$	0.04	0.40	0.28	0.16	0.12

(a) (5 points.) Compute $E(X)$.

We directly apply the formula:

$$\begin{aligned} E(X) &= \sum_{x\in S} x p_X(x) \\ &= 120\cdot 0.04 + 130\cdot 0.4 + 140\cdot 0.28 + 150\cdot 0.16 + 160\cdot 0.12 \\ &= 4.8 + 52 + 39.2 + 24 + 19.2 \\ &= 139.2 \end{aligned}$$

Grading criteria: 5 points in total
-0 point: correct
-2 point: the expectation formula is incorrect or missing
-2 point: the answer is incorrect or missing
-5 point: blank or totally incorrect

(b) (5 points.) Compute $\mathrm{Var}(X)$.

First we compute

$$\begin{aligned} E(X^2) &= \sum_{x\in S} x^2 p_X(x) \\ &= 120^2\cdot 0.04 + 130^2\cdot 0.4 + 140^2\cdot 0.28 + 150^2\cdot 0.16 + 160^2\cdot 0.12 \\ &= 576 + 6760 + 5488 + 3600 + 3072 \\ &= 19496 \end{aligned}$$

Then

$$\begin{aligned} \mathrm{Var}(X) &= E(X^2) - (E(X))^2 \\ &= 19496 - (139.2)^2 \\ &= 119.36 \end{aligned}$$

Grading criteria: 5 points in total
-0 point: correct
-2 point: the variance formula is incorrect or missing
-2 point: the answer is incorrect or missing

-5 point: blank or totally incorrect

(c) (5 points.) Suppose the heart rate of a techno dancer is related to the record's BPM by the formula $Y = 1.2X - 40$. Compute $E(Y)$.

We directly use the linearity of expectation:

$$\begin{aligned} E(Y) &= E(1.2X - 40) \\ &= 1.2E(X) - 40 \\ &= (1.2)(139.2) - 40 \\ &= 127.04 \end{aligned}$$

Grading criteria: 5 points in total
-0 point: correct
-2 point: the linearity of expectation formula is missing or incorrect
-2 point: the final answer is missing or incorrect
-5 point: blank or totally incorrect

5. (25 points.) The safe containing the answers to this exam is protected by a 4-digit numerical code where none of the numbers are repeated. A student has one chance to guess the code and gets an ordered color-coded response where green means "this digit in this position is correct", yellow means "this digit is incorrect but appears in another position", and gray means "this digit is not in the code at all." For example, if the code was 1234 and the student guessed 4238, the response would be yellow-green-green-gray. (This is just like Wordle with numbers.)

(a) (5 points.) How many possible safe combinations are there?

There are 10 possibilities for the first number in the combination, 9 possibilities for the second number, 8 possibilities for the third number, and 7 possibilities for the last number. Thus there are $10 \cdot 9 \cdot 8 \cdot 7 = 5040$ possible safe combinations.

Grading criteria: 5 points in total
-0 point: correct
-1 point: the amount of combination for the first number is missing or incorrect
-1 point: the amount of combination for the second number is missing or incorrect
-1 point: the amount of combination for the third number is missing or incorrect
-1 point: the amount of combination for the last number is missing or incorrect
-1 point: the final answer is missing or incorrect

(b) (5 points.) What is the probability of getting all green (i.e., the student guesses the combination exactly)?

There is only one guess that results in all green (i.e., the correct combination), so the probability is $\frac{1}{5040}$.

Grading criteria: 5 points in total
-0 point: correct
-2 point: the answer is partially incorrect
-5 point: blank or totally incorrect

(c) (5 points.) What is the probability that the student gets 2 greens, 1 yellow, and 1 gray (in any order)?

There are $\binom{4}{2} = 6 =$ places the two greens could be, and then 2 ways the yellow and gray could go in the remaining two slots. The probability of getting a fixed order of 2 greens, 1 yellow, and 1 gray (e.g., green-green-yellow-gray) is $\frac{1}{10} \cdot \frac{1}{9} \cdot \frac{1}{8} \cdot \frac{6}{7} = \frac{6}{5040}$; that is, the green slots are fixed and the yellow slot is also fixed (it must be the number in the gray position), and then there are 6 non-combination numbers that could go in the gray position. This the probability is $12 \cdot \frac{6}{5040} = \frac{1}{70}$.

Grading criteria: 5 points in total
-0 point: correct
-2 point: minor errors, e.g., missing "6", "6/5040"
-5 point: blank or totally incorrect

(d) (5 points.) What is the probability of getting all yellow (i.e., the student guesses all the right numbers but they are all in the wrong positions)?

This is a little trickier. There are $\binom{10}{4} = 210$ ways the student can guess the 4 numbers of the combination, and only one of these is correct. Then if they chose the correct set of numbers, there are $4! = 24$ ways of ordering them. However, not all of these 24 orderings results in 4 yellows (for example, if the first number is correct, they will get a green in the first slot). There are a couple ways of thinking about how many of the 24 combinations will result in 4 yellows; it turns out 9 of the 24 possibilities will work. One quick way is just to enumerate them; another way is to see that there are 3 choices for the first slot and that only 3 of the 6 ways to fill the remaining slots have no digits in the right place. That is, there are 3 places we could put the first number, which displaces some other number that also has three places to go. Then the remaining 2 numbers have to switch places to be yellow.

Thus the probability of getting 4 yellows is $\frac{1}{210} \cdot \frac{9}{24} = \frac{1}{560}$.

More generally, you may be interested to know that a permutation of n objects such that none of them are in the right positions is called a "derangement", and that the number of derangements of n objects is the closest integer to $n!$ divided by e (in this case, $n = 4$ and $4!/e = 8.829$ which is closest to 9). This number is also called the subfactorial of n, written $!n$.

A common wrong answer was to assume that there were 3 possible values in each slot that would result in a yellow, resulting in the wrong answer of $\frac{9}{560}$. The problem with this reasoning is that the 4 slots cannot be chosen independently. That is, once we choose one of the 3 numbers for the first slot, we have either 2 or 3 choices for the second slot, 2 or 3 choices for the third slot, and then the last slot is fully defined.

Grading criteria: 5 points in total
-0 point: correct
-1 point: missing the process deriving "1/210"
-2 point: missing the process deriving "9/24"
-5 point: blank or totally incorrect

(e) (5 points.) What is the probability that the student gets 3 greens and 1 yellow (in any order)?

The probability is 0! That is, we can't have 3 digits in the right place and 1 digit in the wrong place, since there aren't any more places left. This is a variation on an old probability paradox involving addressing envelopes: if we have letters to n people and n addressed envelopes, and we randomly assign letters to envelopes, there is no way to have exactly $n-1$ letters in the right envelopes.

Grading criteria: 5 points in total
-0 point: correct
-2 point: no further (detailed) and correct explanation
-5 point: blank or totally incorrect