Stat 88 Spring 2021 MIDTERM EXAM SOLUTIONS

TIME AND CONDITIONS: You have **80 minutes** to complete the exam and **10 minutes** to upload your submission to Gradescope. The exam is open-book, open-Jupyter notebook, open-calculator, but **no other materials are allowed**. Note that you should not *need* a calculator/Jupyter.

QUESTIONS AND ANSWERS

• There are 7 questions. The first question is related to the honor code and your signature for the parts therein. If your response to this question is missing, **we will not grade the rest of the exam**.

• **Give brief explanations or show calculations in each question** unless the question says this is not required. You may use, without proof, any result proved or used in lecture, the textbook, and homework, unless the question asks for a proof.

• Please leave answers as **unsimplified arithmetic or algebraic expressions including finite sums** unless the question asks for a simplification.

GRADING

• The exam is worth 36 points. Each problem is worth 6 points.

• Please commit yourself to a single answer for each question. If you give multiple answers (such as both True and False) then please don't expect credit, even if the right answer is among those that you gave.

• **It is your responsibility to complete and submit the exam on time.** A late submission will not be accepted under any circumstances.

FORMAT

• **You must answer each question on a separate page for a (minimum) total of 7 pages.**

• Writing solutions on an iPad or tablet is acceptable.

• **You must select the correct page** associated for each subpart of each question when submitting to Gradescope. If you do not, you will get a zero for that question on the exam.

2. My dog Bella loves to play. If we throw a frisbee, the chance that Bella will catch the frisbee is $p_B = 0.3$, independently of all the other times.

(a) If I throw the frisbee 10 times, what is the chance that Bella will catch it more than once? (2 points)

Let B be the event that Bella catches the frisbee more than once in 10 tries. Then $P(B) = 1 - P(X = 0) - P(X = 1) = 1 - 0.7^{10} - 10 \times 0.3^1 \times 0.7^9$.

Points:

- 1 point for recognizing that B is binomial with correct parameters and PMF
- 1 point for correct summation indices (k = 2 to 10 or k = 0 to 1 for the complement)

(b) Bella's friend Annie (also a Labrador retriever), is much better than Bella at catching a frisbee - her probability of catching it is $p_A = 0.8$, independently of all other times. Suppose that Annie comes over to play with Bella, and I throw the frisbee for both Annie and Bella. Note that now the events that they each catch the frisbee are not necessarily independent since they both chase every throw and can get in each other's way. Can you compute the probability now that Annie catches the frisbee more than once out of 10 throws AND Bella catches the frisbee more than once out of 10 throws? If you cannot compute this probability, state why, and then provide upper and lower bounds for it. (4 points)

Let B be the event in part (a) and let A be the event that Annie catches the frisbee more than once in ten throws. We are thus asked to bound or calculate $P(AB)$. Similar to $P(B)$,

$$P(A) = 1 - P(X = 0) - P(X = 1) = 1 - 0.2^{10} - 10 \times 0.8^1 \times 0.2^9 \qquad \text{(1 point)}$$

Since event A and event B are not independent, nor can we infer $P(A|B)$, $P(B|A)$, or $P(A \cup B)$ from the question, the best we can do is to bound $P(AB)$. (1 point)

$$\max\{P(A) + P(B) - 1, 0\} \leq P(AB) \leq \min\{P(A), P(B)\}$$

(using $P(A) + P(B) - 1 \leq P(AB) \leq P(B)$) (1 point for upper bound, 1 for lower bound)

3. Bella loves treats and her humans have bought seven different kinds of treats for her (Beef Jerky, Sweet Potato, Peanut Butter, Pumpkin, Milk, Apple, and Chicken flavors). Each day, she gets one treat after dinner. They let her choose the flavor. She likes them all equally, so she is equally likely to choose any one of them, independently of all other times and choices. Note that in the following, you may assume that we will never run out of any flavor of treat.

(a) The company that makes these treats had narrowed their options to these after doing focus group tests. They began with 12 different flavors of treats, and tested these 12 flavors on 10 dogs (all of whom picked one flavor at random, independently of the other choices). What is the expected number of flavors picked by at least one dog? (3 points)

`Let` X `be the total number of flavors picked by at least one dog. Let` I_k `indicate whether flavor` k `was picked by at least one dog where` $1 \leq k \leq 12$. `From our definition,` $X = \sum_{k=1}^{12} I_k$

$$E[X] = E[\sum_{k=1}^{12} I_k] = \sum_{k=1}^{12} E[I_k]$$
$$= 12 \cdot P(\texttt{flavor k picked by at least one dog}) = 12 \cdot (1 - (1 - \frac{1}{12})^{10})$$

Points:

- `1 point for some attempt to use method of indicator and unambiguously defining the indicators`
- `1 point for recognizing the linearity of expectation and that all indicators are identically distributed`
- `1 point for correct` $P(\texttt{flavor k picked by at least one dog})$

(b) What is the expected number of days until Bella has picked all the seven flavors that we got for her? (3 points)

This is a classic coupon collector's problem. Let Y be the total number of days until Bella has picked all seven flavors that we got for her. Let Y_k be the number of days until Bella has picked her kth flavor after she has gotten k-1 flavors. From our definitions, $Y = \sum_{k=1}^{7} Y_k$. We also know that $Y_k \sim Geom(\frac{7-(k-1)}{7})$ because when she has k-1 different flavors, any one of the remaining 7-(k-1) flavors qualifies as the k-th flavor for her.

$$E[Y] = E[\sum_{k=1}^{7} Y_k] = \sum_{k=1}^{7} E[Y_k] = \sum_{k=1}^{7} \frac{7}{7-(k-1)} = \sum_{k=1}^{7} \frac{7}{8-k} = \sum_{k=0}^{6} \frac{7}{7-k}$$

Points:

- 1 point for dividing Y into 7 random variables or "stages" and unambiguously defining each stage or R.V.
- 1 point for recognizing that the waiting time for each stage follows some kind of Geometric distribution
- 1 point for correct parameter for every Geometric and correct expectation for each Geometric

4. Bella and her friend Kobie both live in households with four humans each. Each of the humans in Bella's household (and Kobie's) might take her for a walk that day. Let B be the number of walks that Bella gets on a randomly selected day, and K be the number of walks that Kobie goes on. Assume that B and K are iid random variables that have the uniform distribution on $\{1, 2, 3, 4\}$. Let M be the maximum of B and K.

(a) Write down the joint distribution of B and M. (3 points)

Since B and K each has 4 possible values, there are 16 possible value pairs of (B, K). We categorize those pairs based on the maximum of B and K.

- $M = 1$: (1,1)
- $M = 2$: (1,2), (2,1), (2,2)
- $M = 3$: (1,3), (2,3), (3,3), (3,2), (3,1)
- $M = 4$: (1,4), (2,4), (3,4), (4,4), (4,3), (4,2), (4,1)

	B = 1	B = 2	B = 3	B = 4
M = 1	1/16	0	0	0
M = 2	1/16	2/16	0	0
M = 3	1/16	1/16	3/16	0
M = 4	1/16	1/16	1/16	4/16

Points:

- 1 point for correct possible values of (B, M)
- 2 points for fully correct probabilities
- 1 point for almost correct probabilities

(b) Are B and M independent? (1 points)

No, some of the example reasons are:

- $P(B = 1|M = 1) = 1 \neq P(B = 1) = \frac{1}{4}$
- $P(B = 1, M = 1) = \frac{1}{16} \neq P(B = 1)P(M = 1) = \frac{1}{4} \cdot \frac{1}{16}$
- The possible values of B will change given different values of M.

(c) What is the expectation of M? (2 points)

$E[M] = 1 \cdot \frac{1}{16} + 2 \cdot \frac{3}{16} + 3 \cdot \frac{5}{16} + 4 \cdot \frac{7}{16} = \frac{25}{8}$

Points:

- 1 point for correct marginal probabilities of M
- 1 point for correct definition of expectation
- 2 points for some other correct but less intuitive approach with adequate justification, e.g. finding expectation by conditioning on M

5. Bella was born in 2011, and is n days old today. She has received little doggie treats every day, and has observed that the number of treats that she receives per day has a Poisson distribution with expectation 10.

(a) One day she noticed that she didn't receive *any* treats and was wondering if her humans were aware of this omission. What is the probability of her not getting treats on any given day? (2 points)

The expectation of the a Poisson distribution is its parameter. Thus, let $X \sim Poisson(10)$, we are asked to find $P(X=0) = e^{-10}\frac{10^0}{0!} = e^{-10}$

Points:

- 1 point for recognizing the parameter as 10
- 1 point for correct $P(X=0)$

(b) Now Bella is a curious dog and would like to know the probability that in n days she would have at least one day in which she received no treats. Can you approximate this probability? (4 points)

Let p be the probability that Bella receives no treats on a given day. From part (a), we know $p = P(X=0) = e^{-10}\frac{10^0}{0!} = e^{-10}$.

$$
\begin{aligned}
&P(\texttt{at least one day in which she received no treats}) \\
&= 1 - P(\texttt{she receives at least one treat every day}) \\
&= 1 - (1-p)^n \approx 1 - e^{-np}
\end{aligned}
$$

The last line uses the following exponential approximation:$\log(1-p)^n = n\log(1-p) \approx -np$ since $p = e^{-10} << 1$ and is very close to zero.

Points:

- 2 points for correct exact probability $1-(1-e^{-10})^n$
- 2 points for correct exponential approximation based on the exact probability they calculated

6. Bella has a wonderful veterinarian who has been in practice for many decades. He can usually examine a sick dog and narrow down the possible conditions to one of three, and then do a blood test. Suppose that in one instance, at first he assigns an equal probability to three conditions that a patient might have, call them c_1, c_2, and c_3. He then performs a blood test that will be positive with probability 0.8 if the patient has c_1, 0.6 if the patient has condition c_2, and 0.4 if the patient has c_3. If the blood test turns out to be positive, now what are the probabilities of the three conditions given this new information? (6 points)

Since prior probabilities are that they are equally likely, $P(c_1) = P(c_2) = P(c_3) = \frac{1}{3}$. Given that $P(+|c_1) = 0.8$, $P(+|c_2) = 0.6$, $P(+|c_3) = 0.4$.

$$\begin{aligned}
P(c_1|+) &= \frac{P(c_1) \cdot P(+|c_1)}{P(c_1) \cdot P(+|c_1) + P(c_2) \cdot P(+|c_2) + P(c_3) \cdot P(+|c_3)} \\
&= \frac{1/3 \cdot 0.8}{1/3 \cdot 0.8 + 1/3 \cdot 0.6 + 1/3 \cdot 0.4} = \frac{4}{9} \\
P(c_2|+) &= \frac{P(c_2) \cdot P(+|c_2)}{P(c_1) \cdot P(+|c_1) + P(c_2) \cdot P(+|c_2) + P(c_3) \cdot P(+|c_3)} \\
&= \frac{1/3 \cdot 0.6}{1/3 \cdot 0.8 + 1/3 \cdot 0.6 + 1/3 \cdot 0.4} = \frac{1}{3} \\
P(c_3|+) &= \frac{P(c_3) \cdot P(+|c_3)}{P(c_1) \cdot P(+|c_1) + P(c_2) \cdot P(+|c_2) + P(c_3) \cdot P(+|c_3)} \\
&= \frac{1/3 \cdot 0.4}{1/3 \cdot 0.8 + 1/3 \cdot 0.6 + 1/3 \cdot 0.4} = \frac{2}{9}
\end{aligned}$$

Equivalently, $P(c_3|+) = 1 - P(c_1|+) - P(c_2|+) = 1 - \frac{4}{9} - \frac{1}{3} = \frac{2}{9}$

Points:

- 1 point for recognizing $P(c_1) = P(c_2) = P(c_3) = \frac{1}{3}$
- 1 point for recognizing $P(+|c_1) = 0.8$, $P(+|c_2) = 0.6$, $P(+|c_3) = 0.4$
- 3 points for correctly applying Bayes' rule to at least one of $P(c_1|+)$, $P(c_2|+)$, and $P(c_3|+)$ with the denominator written in terms of likelihoods and priors, not merely $P(+)$
    - 1 point for correct tree diagram if the above rubric does not apply
- 1 point for correct $P(c_1|+)$, $P(c_2|+)$, and $P(c_3|+)$

7. Bella is participating in a TV reality show for dogs. She needs to get out of a maze, and her starting point is an intersection with three possible paths that she can take. The first path will get her out and to her favorite toy in an expected time of 2 minutes. The second one will have her running through the maze for an expected time of 5 minutes only to end back up at her starting position. The third path, like the second one, will also bring her back here to her starting position, and the expected time she will take on that path is 7 minutes. By the time she gets back, she will have forgotten that this is where she started, so the game starts over for her. She can faintly smell her toy, so she is twice as likely to choose the first path as either of the other two (which are equally likely). What is the expected number of minutes until she gets to her toy? (6 points)

Let X be the number of minutes until she gets to her toy.

$$\begin{aligned}
E[X] &= E[X|\texttt{first path}] \cdot P(\texttt{first path}) \\
&\quad + E[X|\texttt{second path}] \cdot P(\texttt{second path}) \\
&\quad + E[X|\texttt{third path}] \cdot P(\texttt{third path}) \\
E[X] &= 2 \cdot \frac{1}{2} + (5 + E[X]) \cdot \frac{1}{4} + (7 + E[X]) \cdot \frac{1}{4} \\
E[X] &= 8
\end{aligned}$$

Points:

- 1 points for finding expectation by conditioning on which path Bella takes
- 3 points for correct $E[X|\texttt{path } k]$ for k = 1, 2, 3
- 2 points for almost correct $E[X|\texttt{path } k]$ for k = 1, 2, 3
- 1 point for correct $P(\texttt{path } k)$ for k = 1, 2, 3
- 1 point for correct $E[X]$