STAT 88 - Instructor: Zhiyi You

Midterm

SOLUTIONS

Friday, July 16, 2021

Print your name: ______________________________

SID Number: ______________________________

Exam Information and Instructions:

- You will have 100 minutes to solve all five questions in this midterm, plus 10 minutes to scan and upload your work.
- We will be using Gradescope to grade this exam. Write any work you want graded on the front of each page, in the space below each question. If you need more space, please show the grader explicitly where the rest of your work is.
- Provide calculations or brief reasoning in every answer. The final answer only worth very few points without proper explanation.
- Unless stated otherwise, you may leave answers as unsimplified numerical and algebraic expressions. Finite sums are fine, but simplify any infinite sums.
- This exam is open book, open notes, and you may use calculators or python notebook. But you are not required to use them to give your answer.
- You should NOT communicate with anyone other than the instructor or the GSI during the exam.
- The exam will be scored out of 50 points, and there are 6 questions including the honor code as question 1.

GOOD LUCK!

SID: ____________________________

1. (1 pt) Statistics and the entire academic enterprise are based on one quality – integrity. We are all part of a community that doesn't fabricate evidence, doesn't fudge data, doesn't present other people's work as our own, doesn't lie and cheat. You trust that we will treat you fairly and with respect. We trust that you will treat us and your fellow students fairly and with respect. **Please transcribe the UC Berkeley's Honor Code below AND sign your name next to it:**

   "I certify that all solutions will be entirely my own and that I will not consult or share information with other people during the exam. I promise I will act with honesty, integrity, and respect for others."

2. Rainy Weekend!

Suppose you're heading off for a long weekend (Friday, Saturday, and Sunday) somewhere and the weather report for your destination says:

- Chance of rain on Friday: 85%
- Chance of rain on Saturday: 75%
- Chance of rain on Sunday: 65%

In each part below, find the chance exactly if it can be found using no further assumptions (unless specified). If it can't be found, then (again using no further assumptions) find the best lower bound and upper bound that you can.

(a) (3 pts) the chance that it rains in your destination sometime during the long weekend

$85\% \sim 100\%$

(b) (3 pts) the chance that it rains in your destination on all three days of the long weekend

$25\% \sim 65\%$

(c) (3 pts) now suppose that it rains independently on Friday and Sunday, repeat (b) the chance that it rains in your destination on all three days of the long weekend

$85\% * 65\% + 75\% - 100\% \sim 85\% * 65\%$

SID: ____________________________

3. Collecting digits

   I have a fair, 10-sided die, with numbers 0-9. Find:

   (a) (2 pts) the chance that I see the numbers 3, 9 and 7 from my first three rolls (does not have to be in the same order);

   $$\frac{3!}{10^3}$$

   (b) (2 pts) the chance that I see three different numbers from my first three rolls;

   $$\frac{10*9*8}{10^3}$$

   (c) (2 pts) the chance that I see the numbers 3, 9 and 7 as the first three different numbers from my rolls (e.g. 339397 is one of the possibilities); hint: since we only care about different numbers we see, you may consider this as if we are sampling without replacement from numbers 0-9.

   $$\frac{3*2*1}{10*9*8}$$

   (d) (2 pts) suppose I need X rolls after the first to see a second different number, and Y more rolls to see a third different one (e.g. in 224246, $X = 2$ and $Y = 3$), name the distribution of Y and give its parameter(s);

   $$Geom(\frac{8}{10})$$

   (e) (2 pts) the expected number of rolls you need to see three different numbers (e.g. in 224246, you need 6 rolls). hint: use (d).

   $$1 + \frac{10}{9} + \frac{10}{8}$$

4. Bootstrap sampling

Suppose that the Stats department has $N = 100$ Graduate students and faculties, including Adam, Brian, Caroline and Dan. Make a simple random sampling (sampling without replacement) of size $n = 12$.

(a) (2 pts) Suppose that the simples are ordered, i.e. we label them with 1 to $n = 12$. What is the chance that the 4-th sample is Adam and the 11-th sample is Brian?

$$\frac{1}{99 * 100}$$

(b) (2 pts) What is the chance that we sampled all the four of Adam, Brian, Caroline and Dan?

$$\frac{\binom{4}{4}\binom{96}{8}}{\binom{100}{12}}$$

(c) (2 pts) Suppose that the sample does contain all the four of Adam, Brian, Caroline and Dan. Draw a bootstrap sample from the sample ($n = 12$ draws at random with replacement). Find the probability that none of Adam, Brian, Caroline and Dan is chosen in the bootstrap sample. Be careful about your use of the multiplication rule.

$$\left(\frac{8}{12}\right)^{12}$$

(d) (2 pts) Provide a python expression that evaluates to the numerical answer to part (c). You may run it, but only your python codes (not the outcome) will be graded. Please limit your codes to no more than three lines.

(8/12) ** 12

or

stats.binom.pmf(0, 12, 4/12)

(e) (2 pts) Find an exponential approximation to the answer in Part (c) when both N and n are large.

$$e^{-4}$$

SID: ______________________

5. Sum of two Poisson random variables

   Let $X_1 \sim Poisson(1)$ and $X_2 \sim Poisson(2)$ and they are independent. $S := X_1 + X_2$ is their sum.

   (a) (3 pts) Find the conditional distribution of $S|X_1 = 2$;

   $$P(S = s|X_1 = 2) = P(X_2 = s - 2) = e^{-2}\frac{2^{s-2}}{(s-2)!}, \quad s = 2, 3, 4, \cdots$$

   (b) (3 pts) Show that $(X_1|S = 4) \sim$ Binom $(4, p)$ and determine the parameters p;

   $$\begin{aligned} P(X_1 = k|S = 4) &= \frac{P(X_1 = k, X_2 = 4-k)}{P(S = 4)} \\ &= \frac{\frac{e^{-1}}{k!} \cdot \frac{e^{-2}2^{4-k}}{(4-k)!}}{\frac{e^{-3}3^4}{4!}} \\ &= \binom{4}{k}\frac{2^{4-k}}{3^4}, \quad k = 0, 1, 2, 3, 4, \end{aligned}$$

   which is the pmf of Binom$(4, 1/3)$.

   (c) (2 pts) Find $E(X_1|S = 4)$.

   $$p = 4/3$$

   (d) (2 pts) As an analogue to conditional expectation, we define conditional variance to be variance of the conditional distribution. In other words, you may treat the conditional distribution as if it was a regular distribution and calculate its regular variance. Find the conditional variance $Var(X_1|S = 4)$.

   $$8/9$$

SID: ________________________

6. Four players are playing cards using a standard deck of 52 cards. Each one is dealt with 13 cards so that no card is left.

   (a) (5 pts) Find the chance that one of the players has exactly three Aces in hand;

$$4 \times \frac{\binom{4}{3}\binom{48}{10}}{\binom{52}{13}}$$

   (b) (5 pts) There are 13 different numbers of cards in the deck. We say someone is holding a 'three of a kind' if he has exactly three cards with the same number (e.g. AAA, but not AAAA). A player may hold more than one 'three of a kind' at the same time. What is the expected number of 'three of a kind' in all four players' hands combined? (e.g. player 1 is holding exactly three Aces, player 3 is holding exactly three 7's and three 9's, while no other 'three of a kind' appears, then the number is 3)

$$13 \times 4 \times \frac{\binom{4}{3}\binom{48}{10}}{\binom{52}{13}}$$

Scratch Paper