MATH 4710 **PRELIM 2** **FALL 2021**

Instructions

(1) There are 4 independent problems. The point total is 30. The exam is due on gradescope on Thursday November 11 at 8:00pm (grace period until 12:00pm)

(2) You can write your answers on any reasonable media that is convenient to you as long as you can produce a clean pdf to upload on gradescope. Write your name and Cornell NetID on the top of the first page before you begin.

(3) Write clearly using a black or blue pen or pencil. When you are done, create a pdf file of your work and upload it on gradescope. Make sure to assign problems to pages on gradescope.

(4) **Always provide logical reasons for your answers and explain your computations**. Do not refer to homework problems to justify your solutions. You will receive partial credit for well explained steps in the right direction even if you are not able to provide a complete solution. For numerical answers, give either a simplified fraction or a decimal answer, which ever comes more easily. For instance, $3 \times 7^2/(15)^9$ is very acceptable (much better than the expanded version) but $8/12$ is not very good. You can use a scientific calculator to compute decimal answers but only after writing explicitly what it is that you compute.

(5) You can use your notes, our canvas website including all documents provided there and the book. Do not use other websites or the internet (except for Desmos or a simple electronic calculator). Do not discuss prelim problems with other students. Do not discuss prelim problems with anyone except Pr. Saloff-Coste (ask Professor Saloff-Coste privately, by email, or in office hours if you have questions. It is OK to do so)..

(6) Academic integrity is expected of all Cornell University students at all times, whether in the presence or absence of members of the faculty.

Problem 1: (9 pts)

(a-2pts) We have a rack with 8 slots numbered 1 to 8. We have eight balls numbered 1 to 8. The balls are placed in a random order in the slots, one per slot. What is the expected number of balls for which the slot number and the ball number coincide?

For each $i = 1, \dots, 8$, set $X_i = 1$ if ball i is in slot i, and $X_i = 0$ otherwise. With this notation, the total number of balls placed in the slot with same number as the ball is $X = \sum_1^8 X_i$. It follows that $E(X) = \sum_1^8 E(X_i) = \sum_1^8 P(X_i = 1)$. For each ball the probability that it is placed in its own slot is $1/8$ so that $E(X) = 1$. On average, we expect that one ball will occupy its own slot.

(b-3pts) To put money aside for my next vacation, each evening, I play the following game. First, I flip a fair coin. If it comes Heads, I roll a fair six-faced die until I get a 1. If the coin comes Tails, I roll the die until I get 3 or 6. In both cases, I then put aside as many dollars as the number of times I rolled the die. What is the expected amount of money that I put aside each evening?

Let X be the random variable representing the money I put aside on a given evening. We need to compute $P(X = k)$, $k = 1, 2, \dots$. Let $Y \in \{H, T\}$ be the result of the coin flip. Given the result of the coin flip, the number of rolls is a geometric random variable with parameter $p = 1/6$ if $Y = H$ and parameter $p = 1/3$ if $Y = T$.

In other words, we have

$$PX = k|Y = H) = (1/6)(5/6)^{k-1} \text{ and } P(X = k|Y = T) = (1/3)(2/3)^{k-1}.$$

It follows that

$$P(X = k) = \frac{1}{2}\left(P(X = k|Y = H) + P(X = k|Y = T)\right) = \frac{1}{2}\left((1/6)(5/6)^{k-1} + (1/3)(2/3)^{k-1}\right).$$

Now,

$$E(X) = \sum_{k=1}^{\infty} kP(X = k) = \frac{1}{2}\left(\frac{1}{6}\sum_{k=1}^{\infty} k(5/6)^{k-1} + \frac{1}{3}\sum_{k=1}^{\infty} k(2/3)^{k-1}\right)$$

The expectation of a geometric random variable with parameter p is $1/p$. It follows that

$$\frac{1}{6}\sum_{k=1}^{\infty} k(5/6)^{k-1} = 6 \quad \text{and} \quad \frac{1}{3}\sum_{k=1}^{\infty} k(2/3)^{k-1} = 3.$$

This gives $E(X) = (6 + 3)/2 = 4.5$.

Many of you use $E(X) = (1/2)E(\text{Geom}(1/6)) + (1/2)E(\text{Geom}(1/3))$ without explanations. Many who try to explain do not quite succeed. Here is another clever explanation proposed by one of you (I add a few details): Suppose every night we roll two dice (blue and red): the blue one until we get a 1, the red one until we get a 3 or a 6. Let B and R the number of rolls of each die. Now, independently, we flit a coin and set $Z = 1$ if Heads and $Z = 0$ if Tails. Then we can represent X (from the question) as $X = BZ + R(1 - Z)$ and compute using linearity of expectation $E(X) = E(BZ) + E(R(1 - Z))$. By independence, of B and Z, and R and Z, $E(BZ) = E(B)E(Z)$ and $E(R(1 - Z)) = E(R)E(1 - Z)$. From there you get $E(X) = (1/2)E(B) + (1/2)E(R) = ...$

(b-4pts) An urn has n red balls and m blue balls. The balls are removed from the urn, one by one, choosing uniformly at random which ball from those left in the urn is removed each time until the urn is empty. On average, how many times is a red ball followed by a blue ball in this entire removal process.

Imagine that the red balls are marked from 1 to n. For each $i \in \{1, \dots, n\}$ set $X_i = 1$ if the removal of red ball number i is followed by a blue ball in the removal process, and set $X_i = 0$ otherwise. Let X be the number of times a red ball is followed by a blue ball in this entire removal process. Then

$E(X) = \sum_1^n E(X_i) = \sum_1^n P(X_i = 1)$. Now, fix i, and let $T \in \{1, \dots, n+m\}$ be the step at which red ball number i is removed. We have

$$P(X_i = 1) = \sum_{k=1}^{n+m-1} P(T = k \text{ and the next ball removed is blue})$$

Red ball number i is equally likely to be removed at any step so that $P(T = k) = 1/(n+m)$, $k = 1, \dots, n+m$. Given that red ball number i has been removed at step number k with $k \le n+m-1$, each of the other $n+m-1$ balls is equally likely to be the next ball removed, that is, for each $k \le n+m-1$,

$$P(\text{the } k+1 \text{ ball removed is blue}|T = k) = \frac{m}{n+m-1}.$$

It follows that

$$P(T = k \text{ and the next ball removed is blue}) = \frac{1}{n+m}\frac{m}{n+m-1}$$

and

$$\begin{aligned} P(X_i = 1) &= \sum_{k=1}^{n+m-1} P(T = k \text{ and the next ball removed is blue}) \\ &= (n+m-1)\frac{1}{n+m}\frac{m}{n+m-1} = \frac{m}{n+m}. \end{aligned}$$

Finally, $E(X) = nm/(n+m)$.

Alternate solution: for $i = 1, \dots, n+m-1$, set $Y_i = 1$ if the ball removed at step i is red and a blue ball is removed next, and $Y_i = 0$ otherwise. The $E(X) = \sum_1^{n+m-1} E(Y_i) = \sum_1^{n+m-1} P(Y_i = 1)$. The probability that a red ball is removed at step i is $n/(n+m)$. Given that a red ball is removed at step $i \le n+m-1$, the probability that a blue ball is removed next is $m/(n+m-1)$ because the ball removed at step $i+1$ is equally likely to be any of the $n+m-1$ other balls. This gives $E(X) = nm/(n+m)$.

Alternate solution: Imagine the red balls are also marked 1 to n. Let $Z_i = 1$ if red ball numbered i is followed by a blue ball. Then $X = \sum_1^n Z_i$ and $E(X) = nE(Z_1)$ (by linearity and the fact that the expectation is the same for all i). We need to compute $E(Z_1) = P(Z_1 = 1)$. This is the probability that red ball number 1 be followed by a blue ball. For that to happen, red ball number 1 must not be the last ball. If it is not the last ball then the probability that it is followed by a blue ball is $\frac{m}{n+m-1}$. This gives $P(X_1 = 1) = \frac{n+m-1}{n+m}\frac{m}{n+m-1} = \frac{m}{n+m}$. Finally, $E(X) = \frac{nm}{n+m}$.

Problem 2: (6 pts) In this problem, you need to explain carefully each steps you take, citing results that justify taking those steps. The exercise is about making reasonable assumptions based on material and concepts we have learned in class to arrive to a reasonable estimate given the limited information you have. Identify clearly the assumption you are making; provide justifications (we are not asking for "proofs" here) for these assumptions. If you use approximations, say clearly that you are doing so and explain why you think it is justified.

The owner of a used bookstore specializing in rare books has observed that in the last 1000 days there was five days on which no customers visited the store. What is a reasonable estimate for the expected number of visitors to the store on a given day? What is a reasonable estimate for the probability that exactly 4 or 5 people will visit the store next Monday?

Call X the number of customers visiting the store on a given day. The first step is to estimate $P(X = 0)$ using the observed frequency of that event in 1000 repeated independent trials, that is, $P(X = 0) \approx 5/1000 = .005$. This is justified heuristically by the (weak) law of large number.

Next, we are at a loss on how to compute the average number of visitors on a given day... We can think of the entire population of the city in which the store is situated. On a given day, each person has a very small probability to visit the store and it seems appropriate to model the number of people visiting the store using a Poisson distribution with some unknown parameter λ. This is because the Poisson distribution is appropriate when we count occurrences of a rare event. This means that we think that $P(X = k) = e^{-\lambda}\lambda^k/k!$.

Together with our prediction that $P(X = 0) \approx .005$, this gives $e^{-\lambda} \approx .005$ or $\lambda \approx \ln(200) \approx 5.3$.

According to the predictions and assumptions explained above, $E(X)$, the average number of customers visiting the store on a given day, is equal to $\lambda \approx 5.3$ because X is a Poisson λ random variable (or, more precisely, is well approximated by such).

To estimate the probability that 4 or 5 customers visit the shop next Monday, that is $P(X \in \{4, 5\})$, we use

$$P(X = 4) = e^{-5.3}(5.3)^4/4! \approx .164, \quad P(X = 5) = e^{-5.3}(5.3)^5/5! \approx .174,$$

and

$$P(X \in \{4, 5\}) \approx .34.$$

Additional justifications: We are counting events occurring in time and we discussed in class general hypotheses under which one can prove that the Poisson distribution is a good model for such counting. There are three hypotheses to verify. All three can be justified heuristically in this case. See the Notes, Section 3.9.2.

Problem 3: (7 pts) A group of twenty people including twin sisters Alice and Jess meet for dinner every Tuesdays at a restaurant where they seat around five tables in groups of four. The seating is done at random so that the four people at a given table are chosen uniformly at random among the twenty participants.

(a-2pts) On a given Tuesday, what is the probability that Alice and Jess seat at the same table?

Let us consider the table at which Alice seats that Tuesday. The group of three people seating at that table with Alice is chosen uniformly at random among the $\binom{19}{3}$ possibilities. Among those, there are $\binom{18}{2}$ groups including Jess. So the desired probability is

$$\frac{\binom{18}{2}}{\binom{19}{3}} = \frac{3}{19}.$$

Although this is a convincing argument, you should note that the sentence "the group of three people seating at that table with Alice is chosen uniformly at random among the $\binom{19}{3}$ possibilities" states something that is true but not entirely obvious. One (somewhat clever) way to explain it is symmetry. A more direct way is to compute the probability that guests $\{w, x, y, z\}$ seat at table 1 given that Alice seats at table 1. This is 0 is Alice is not one of w, x, y, z, and it is

$$\frac{1}{\binom{20}{4}} \times \frac{1}{5} = \frac{20 \times 19 \times 18 \times 17}{5 \times 4 \times 3 \times 2 \times 1} = \frac{1}{\binom{19}{3}}.$$

(b-2pts) What is the probability distribution for the number of times Alice and Jess will seat together over the next 2 years (we assume that any year has exactly 52 weeks)?

The number of times Alice and Jess will be seating together at a table over the next 104 weeks is a random variable, call it X. This random variable counts the number of successes (say that seating together is a "success") in 104 repeated identical independent experiment (what happens each week). The probability of success in one of these experiment is $3/19$ (see answer to question a). It follows thta X is a binomial random variable with parameter $n = 104$ and $p = 3/19$. This means that $P(X = k) = (3/19)^k (16/19)^{104-k} \binom{104}{k}$, $k = 0, \ldots, 104$.

(c-3pts) How can we approximate the probability that Alice and Jess will seat together at least twenty times over the next 2 years, and why? What is good approximate value for this probability?

Given that X is binomial with $np(1-p) = 104 \times (3/19) \times (16/19) \approx 13.83 > 10$, it is reasonable to approximate $P(X \geq 20)$ using the central limit theorem, that is, using a normal distribution. To that end, we use the continuity correction (especially because $np(1-p)$ is not very large) and write. $P(X \geq 20) = P(X \geq 19.5)$. Next, we "center" and "normalize the variance" as follows:

$$P(X \geq 19.5) = P\left(\frac{X - E(X)}{\sqrt{np(1-p)}} \geq \frac{19.5 - E(X)}{\sqrt{np(1-p)}}\right).$$

We have $E(X) = np \approx 16.421053$ and

$$\frac{16.5 - E(X)}{\sqrt{np(1-p)}} \approx \frac{.3078947}{3.71864} \approx .83$$

Using a normal table to read the values of the cumulative distribution Φ, of a normal $N(0,1)$, this yields $P(X \geq 19.5) \approx 1 - \Phi(.83) \approx .203$.

Problem 4: (8 pts) The positive random variable X has cumulative distribution function

$$F(x) = \begin{cases} 1 - e^{-x^2/2} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

(a-2pts) What is the probability that $X \in (0.1, .2)$? (3 decimals)

We have $P(X \in (.1, .2)) = F(.2) - F(.1) = e^{-.005} - e^{-.02} \approx .015$

(b-3pts) What is the density function f of the random variable X? What is $E(X)$?

At any point where F has a continuous derivative, the density of X is the derivative $f = F'$ of F. It follows that $f(x) = xe^{-x^2/2}\mathbf{1}_{(0,+\infty)}(x)$. We then compute $E(X) = \int_0^{+\infty} x^2 e^{-x^2/2}dx$ by integration by part with $u(x) = x$, $v'(x) = xe^{-x^2/2}$, so that $E(X) = \int_0^{+\infty} u(x)v'(x)dx = [uv]_0^{+\infty} - \int_0^{+\infty} u'(x)v(x)dx$ where $v(x) = -e^{-x^2/2}$. This gives $E(X) = \frac{1}{2}\int_0^{+\infty} e^{-x^2/2}dx = \frac{\sqrt{2\pi}}{2} = \sqrt{\pi/2}$. Here we use the fact that $\int_{-\infty}^{+\infty} e^{-x^2/2}dx = \sqrt{2\pi}$.

An alternate proof is to use that $\int_{-\infty}^{+\infty} x^2 e^{-x^2/2}dx = \sqrt{2\pi}$. We know this fact because it is equivalent to the fact that the variance of a normal $(0, 1)$ random variable is 1.

(c-3pts) What is the density function of the random variable $Y = 3X^2$?

To compute the density of Y, f_Y, we first compute its cumulative distribution function $F_Y(y) = P(Y \le y)$. Note first that $P(Y \le y) = 0$ when ever $y \le 0$. For $y > 0$, because $Y = 3X^2$, we have

$$P(Y \le y) = P(3X^2 \le y) = P(X \le \sqrt{y/3}) = 1 - e^{-y/6}.$$

It follows that the density of Y is $f_Y(y) = \frac{1}{6}e^{-y/6}\mathbf{1}_{(0,+\infty)}(y)$. That is, Y is an exponential random variable with parameter $\lambda = 1/6$.