
Instructions

(1) There are 4 independent problems. The point total is 25. The exam is due on gradescope on Friday November 10 at 9:00pm (grace period until Sunday November 12 at 9:00pm)

(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 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).

Do not use websites that provide answers to questions or any type of AI-bot.

(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: (6 pts) We have 15 blank cards marked 1 to 15 in black. A printing machine takes this batch of cards and print a number chosen uniformly at random in $\{1,\ldots,50\}$, in red, on each card, independently. Recall that a pair of cards is a set of two cards. A triple of cards is a set of 3 cards: $\{2,5\}$ and $\{5,2\}$ are the same pair of cards; $\{6,7,9\},\{7,9,6\},...$ are the same triple of cards.

(a-2pts) Give an exact formula for the probability p that no pairs of cards carry the same red number.

To compute this probability, we inspect the batch in order. For no two cards to carry the same red number, the number on the first card can be any number in $\{1,\ldots,50\}$, the number on the second card must be different from that on the first card, the number on the third card must be different from the first two and so on through the entire batch of cards. This give

$$p = 1 \times \frac{49}{50} \times \frac{48}{50} \times \cdots \times \frac{36}{50}.$$

(b-2pts) Let N be the number of distinct triples of cards sharing their red number. Can you compute $E(N)$ exactly? If yes, do it; if no, explain why.

Inspect the $\binom{15}{3}$ triples of cards and set $X_{ijk} = 1$ if cards $i < j < k$ carry the same red number, and $X_{ijk} = 0$ otherwise. Then $P(X_{ijk} = 1) = \frac{1}{2500}$ and $N = \sum_{i<j<k} X_{ijk}$ and

$$E(N) = \sum_{i<j<k} P(X_{ijk} = 1) = \binom{15}{3} \times \frac{1}{2500} \approx .18.$$

(c-2pts) Thinking of a possible reasonable approximation, what is your estimate of the probability $P(N = 0)$ that no triples of cards are assigned the same red number?

As above, inspect the $\binom{15}{3}$ triples of card and set $X_{ijk} = 1$ if cards $i < j < k$ carry the same number, and $X_{ijk} = 0$ otherwise. The Bernoulli random variables X_{ijk} are no independent but we are going to pretend they are (they are no too far from being independent, for instance, two of them are independent if they have no card in common). Based on this idea, we are going to use the Poisson approximation for the number N of triples with cards haring the same red number as if this random variable N was binomial. The mean of N, $E(N)$ equals $\binom{15}{3} \times \frac{1}{2500}$ and we thus model N using a Poisson random variable with that same mean. So $P(N = 0) \approx e^{-\binom{15}{3} \times \frac{1}{2500}} \approx e^{-.182} \approx .83.$

Problem 2: (7 pts) The pair (X, Y) has joint density function $f = f_{(X,Y)}$ given by

$$f(x, y) = x^2 e^{-x(y+1)} \mathbf{1}_{(0,+\infty)}(x) \mathbf{1}_{(0,+\infty)}(y), \quad (x, y) \in \mathbb{R}^2,$$

where $\mathbf{1}_A(z) = \begin{cases} 1 \text{ if } z \in A \\ 0 \text{ otherwise.} \end{cases}$

(a-3pts) Find the density function of the second marginal, Y.

We have $f_Y(y) = \int f(x, y)dx = \int_0^{+\infty} x^2 e^{-x(y+1)}dx$ if $y > 0$ and 0 otherwise. For $y > 0$, this integral, multiplied by $(y + 1)$ is the second moment of an exponential random variable with parameter $\lambda = y + 1$. This gives $f_Y(y) = 2(1 + y)^{-3}\mathbf{1}_{(0,+\infty)}(y)$.

(b-4pts) Find the cumulative distribution and the density function of $Z = X(1 + Y)$.

Note that $P(Z \leq 0) = 0$. For $z > 0$, define the region

$$R_z + \{(x, y) \in (0, +\infty)^2 : x(1 + y) \leq z\} = \{(x, y) : 0 < x < z; 0 < y \leq (z/x) - 1\}$$

and write $F_Z(z) = P(Z \leq z) = P(X(1 + Y) \leq z) = \iint_{R_z} x^2 e^{x(y+1)}dxdy$. Now,

$$\begin{aligned} \iint_{R_z} x^2 e^{x(y+1)}dxdy &= \int_0^z xe^{-x} \left(\int_0^{(z/x)-1} xe^{-xy}dy \right) dx \\ &= \int_0^z xe^{-x}(1 - e^{-((z/x)-1)x})dx = \int_0^z (xe^{-x} - xe^{-z})dx. \end{aligned}$$

We have $\int_0^z xe^{-x}dx = [-xe^{-x}]_0^z + \int_0^z e^{-x}dx = -ze^{-z} + 1 - e^{-z}$ and $\int_0^z xe^{-z}dx = \frac{z^2}{2}e^{-z}$. This gives

$$F_Z(z) = 1 - e^{-z} - ze^{-z} - \frac{1}{2}z^2e^{-z} \text{ for } z > 0.$$

The density function of Z is

$$f_Z(z) = F'_Z(z) = \frac{1}{2}z^2e^{-z} \text{ for } z > 0$$

and 0 otherwise.

Problem 3: (7 pts) People enter a building at the ground floor and take a single elevator leading to all $N \geq 1$ floors of the building. For each person, the floor at which they exit the elevator is a uniform random variable in $\{1, \dots, N\}$ and the choices made by different people are independent. The elevator stops at a given floor only if at least one person exits there.

(a-1pts) If k people enter the building, what is the chance that the elevator does not make a stop at floor 1?

The probability that none of the k people exits at floor 1 is $(1 - 1/N)^k$.

(b-2pts) If k people enter the building, what is the expected number of floors at which the elevator stops?

For each floor $i \in \{1, \dots, N\}$, let $X_i = 1$ if the elevator stops and 0 otherwise. Let X be the number of floors at which the elevator stops. We have $X = \sum_1^N X_i$ and $E(X) = \sum_1^N E(X_i) = N(1 - (1 - 1/N)^k)$.

Remark: The final result says that $E(X)$ is the same as if X was a binomial random variable with parameter N and $1 - (1 - 1/N)^k$, but X is not a binomial random variable because the X_is are not independent: Suppose that only one person enter the building, obviously, if $X_1 = 1$ then all the other X_is are 0.

(c-2pts) We now assume that the number K of people entering the building is a Poisson random variable with mean 10.

(c1) Let X be the number of floors at which the elevator stops. What is $\sum_{n=1}^N nP(X = n|K = k)$?

By definition, this is the expectation of X when we know that k people enter the building. It is assumed implicitly that the random variable K is independent of the choices made by the individuals. The given quantity equals $N(1 - (1 - 1/N)^k)$. Note that this would not be true if K and the choices made by the individuals are dependent. Imagine for instance that the first $N - 1$ floors are occupied by doctor offices but the last floor is occupied by an art school whose students arrive from nearby schools by bus carrying groups of 50 students...

(c2) What is the expected number of floors at which the elevator stops?

Let X be the number of floors at which the elevator stops. We have

$$E(X) = \sum_{n=1}^{N} nP(X = n) = \sum_{n=1}^{N} \sum_{k=0}^{+\infty} nP(X = n \& K = k) = \sum_{k=0}^{+\infty} \sum_{n=1}^{N} nP(X = n|K = k)P(K = k).$$

Now, (c1) tells us the value of the inner sum $\sum_{n=1}^N nP(X = n|K = k) = N(1 - (1 - 1/N)^k)$ and $P(K = k) = e^{-10}\frac{10^k}{k!}$. It follows that

$$\begin{aligned} E(X) &= Ne^{-10}\sum_0^{\infty}(1 - (1 - 1/N)^k)\frac{10^k}{k!} \\ &= N\left(1 - e^{-10}\sum_0^{\infty}\frac{(10(1 - 1/N))^k}{k!}\right) = N(1 - e^{-10}e^{10(1-1/N)}) = N(1 - e^{-10/N}). \end{aligned}$$

(d-2pts) There are 4 floors and 108 people enter the building. What is the approximate probability rounded to two decimals that the number of people exiting at the third floor is either 28 or 29? This question is independent of the previous questions.

For this question, we use the central limit theorem with continuity correction. Let Y be the number of people exiting at the third floor. It is a binomial random variable with $n = 108$ and $p = 1/4$ so that $np = 27$ and $np(1 - p) = 81/4$. We are asked to compute $P(28 \leq Y \leq 29) = P(27.5 < Y < 29.5)$. Next, $P(27.5 < Y < 29.5) = P(1/9 < \frac{Y-27}{9/2} < 5/9) \approx \Phi(5/9) - \Phi(1/9) \approx 0.71 - 0.54 \approx 0.17$.

Problem 4: (5 pts) We say that a random variable with finite expectation $E(X)$ is symmetric around its mean if $P(X - E(X) < -x) = P(X - E(X) > x)$ for all $x \geq 0$. Recall that if Z is a standard normal random variable, $P(-1.96 < Z < 1.96) \approx .95$. In this problem, you can use the statements in earlier questions when solving later questions even if you have not solved the earlier questions.

(a-3pts) The continuous random variable X has finite expectation $E(X)$, is symmetric around its mean, and has a positive continuous density on $(-\infty, +\infty)$.

(a1) What is $P(X < E(X))$? Show that there is exactly one value y such that $P(X < y) = 1/2$. What is this value?

Because X is a continuous random variable, $P(X = E(X)) = 0$. Hence, $1 = P(X < E(X)) + P(X > E(X)) = 2P(X < E(X))$. It follows that $P(X < E(X)) = 1/2$. Because $F_X(x) = P(X < x)$ and $F'_X(x)$ is continuous and positive, F_X is strictly increasing and continuous with limits at $-\infty$ equal to 0 and limit at $+\infty$ equal to 1, it takes all values in $(0, 1)$ (intermediate value theorem), exactly once. This shows that $P(X < y) = 1/2$ implies $y = E(X)$.

A uniform random variable on $[-5, -4] \cup [4, 5]$ is symmetric around its mean (0) but $P(X < y) = 1/2$ for all $y \in [-4, 4]$ also showing the median is not unique, in general.

(a2) Given a real a such that $P(X < a) < 1/2$, show that there exists exactly one real $b \neq a$ such that $P(X > b) = P(X < a)$.

By the reasoning in (a1), on $(-\infty, E(X))$, only $x = a$ satisfies $P(X < x) = P(X < a)$. Similarly, the function $G(t) = P(X > t)$ is strictly decreasing from $[E(X), +\infty)$ onto $(0, 1/2]$ so that there is only one real b such that $P(X > b) = P(X < a)$.

A uniform on $[-2, 2]$ is symmetric around its mean but $P(X < a) = P(X > c) = 0$ for all $a \leq -2$ and all $c \geq 2$.

(a3) Suppose we know that $P(X > b) = P(X < a) = .05$. What is the mean of X as a function of a, b?

Using (a2) and the symmetry hypothesis, given a such that $P(X < a) = \theta \leq 1./2$, there is only one b such that $P(X > b) = P(X < a)$ and

$$P(X < a) = P(X - E(X) < a - E(X)) = P(X - E(X) > -a + E(X)) = P(X > -a + 2E(X)).$$

It follows that $b = -a + 2E(X)$, that is, $E(X) = (b + a)/2$. Note that the value of $\theta \leq 1/2$, for us, $\theta = .05$, is not important.

The result is not true for a uniform on $[-12, -11] \cup [-9, 9] \cup [11, 12]$ because $P(X < -11) = .05$ and $P(X > 9) = .05$ but $E(X) = 0 \neq (-11 + 9)/2 = 1$.

(b-2pts) A club serves dinner to members only. They are seated at 12-seat tables. Each member decides to come with probability p and the decisions of the members are independent. The manager observes that, over a long period of times, 95% of the time there are between 6 and 8 full tables of members, and the remainder of the time the numbers are equally likely to fall above or below this range. How many members are there in the club and what is p, approximately?

Let n be the number of members. The number of members at one dinner is a Binomial random variable with parameter n and p which we approximate using a normal random variable X with mean $E(X) = np$ and variance $\sigma^2 = np(1 - p)$. Normal random variables are symmetric around their mean and have continuous positive density. Using this approximation and the hypothesis, we know that $P(X < 71.5) = P(X > 96.5)$. This is because 6 full tables means 72 members presents, 8 tables means 96 members. Using (a) this gives $E(X) = np = 168/2 = 84$. Also,

$$P(71.5 < X < 96.5) = P\left(-\frac{12.5}{\sqrt{np(1-p)}} < \frac{X - 84}{\sqrt{np(1-p)}} < \frac{12.5}{\sqrt{np(1-p)}}\right) = .95$$

So $\frac{12.5}{\sqrt{np(1-p)}} \approx 1.96$ or $np(1-p) \approx \left(\frac{12.5}{1.96}\right)^2 \approx 40.67$. This gives $84(1-p) = 40.67$, that is, $p \approx .52$ and $n \approx 162$.

If we do not use the continuity correction, we use $P(X < 72) = P(X > 96)$ which gives the same result $np = 84$. Then the deviation 12.5 is replaced by 12 and we get $p \approx .55$, $n \approx 153$. You should use the continuity correction!