MATH 180A - INTRODUCTION TO PROBABILITY
PRACTICE FINAL

FALL 2020

Name (Last, First): ______________________________

Student ID: ______________________________

REMEMBER THIS EXAM IS GRADED BY A HUMAN BEING. WRITE YOUR SOLUTIONS NEATLY AND COHERENTLY, OR THEY RISK NOT RECEIVING FULL CREDIT.

THIS EXAM WILL BE SCANNED. MAKE SURE YOU WRITE ALL SOLUTIONS ON THE PAPER PROVIDED. DO NOT REMOVE ANY OF THE PAGES.

THE EXAM CONSISTS OF N QUESTIONS. YOU ARE ALLOWED TO USE RESULTS FROM THE TEXTBOOK, HOMEWORK, AND LECTURE.

1. Suppose that $X \sim \text{Geom}(p)$ and $Y \sim \text{Geom}(q)$ are independent random variables. Find the probability $\mathbb{P}(X < Y)$.

Proof. We decompose the desired probability as

$$\mathbb{P}(X < Y) = \sum_{k=1}^{\infty} \mathbb{P}(X < Y, X = k) = \sum_{k=1}^{\infty} \mathbb{P}(X = k, Y > k) = \sum_{k=1}^{\infty} \mathbb{P}(X = k)\mathbb{P}(Y > k).$$

Since $X \sim \text{Geom}(p)$, we have $\mathbb{P}(X = k) = p(1-p)^{k-1}$. Similarly, since $Y \sim \text{Geom}(q)$, we have $\mathbb{P}(Y > k) = (1-q)^k$. So,

$$\begin{aligned}
\mathbb{P}(X < Y) &= \sum_{k=1}^{\infty} \mathbb{P}(X = k)\mathbb{P}(Y > k) \\
&= \sum_{k=1}^{\infty} p(1-p)^{k-1}(1-q)^k \\
&= p(1-q)\sum_{k=1}^{\infty} ((1-p)(1-q))^{k-1} \\
&= p(1-q)\frac{1}{1-(1-p)(1-q)}.
\end{aligned}$$

$\square$

2. Suppose that $X \sim \text{Unif}[-2,1]$. Let $Y = X^2$.

(a) Find the CDF of Y.

Proof. Compute the CDF $F_Y(t)$ using the definition

$$F_Y(t) = P(Y \leq t) = P(X^2 \leq t) = P(-\sqrt{t} \leq X \leq \sqrt{t}).$$

Note that $Y = X^2 \in [0,4]$, therefore $F_Y(t) = 0$ for $t < 0$, and $F_Y(t) = 1$ for $t \geq 4$. For any $t \in (0,1)$,

$$P(-\sqrt{t} \leq X \leq \sqrt{t}) = \frac{2\sqrt{t}}{3}.$$

If $t \in [1,4)$, then

$$P(-\sqrt{t} \leq X \leq \sqrt{t}) = P(-\sqrt{t} \leq X \leq 1) = \frac{1+\sqrt{t}}{3}.$$

Finally,

$$F_Y(t) = \begin{cases} 0, & t < 0, \\ \frac{2\sqrt{t}}{3}, & t \in [0,1), \\ \frac{1+\sqrt{t}}{3}, & t \in [1,4), \\ 1, & t \geq 4. \end{cases}$$

□

(1) Is Y discrete, continuous, or neither? If discrete, find the p.m.f. If continuous, find the density. If neither, explain why.

Proof. Y is continuous since its CDF is continuous. Its density is

$$f_Y(t) = F_Y'(t) = \begin{cases} \frac{1}{3\sqrt{t}} & \text{if } t \in (0,1); \\ \frac{1}{6\sqrt{t}} & \text{if } t \in (1,4); \\ 0 & \text{else.} \end{cases}$$

□

3. Suppose that we choose a number N uniformly at random from the set $\{0, \ldots, 4999\}$. Let X denote the sum of its digits. For example, if $N = 123$, then $X = 1 + 2 + 3 = 6$. Determine $\mathbb{E}[X]$.

Proof. Note that $X = Y_1 + Y_2 + Y_3 + Y_4$, where Y_i is the ith digit of the number that is drawn. In other words,

$$N = 1000Y_1 + 100Y_2 + 10Y_3 + Y_4.$$

Next, we see that

$$Y_1 \sim \text{Unif}\{0, 1, \ldots 4\}, \qquad Y_2, Y_3, Y_4 \sim \text{Unif}\{0, 1, \ldots, 9\}.$$

This means that

$$\mathbb{E}[Y_1] = 2, \qquad \mathbb{E}[Y_2] = \mathbb{E}[Y_2] = \mathbb{E}[Y_3] = 4.5.$$

So,

$$\mathbb{E}[X] = \sum_{i=1}^{4} \mathbb{E}[Y_i] = 2 + 3(4.5) = 15.5.$$

□

4. Let T be the triangle in $\mathbb{R}^2$ with vertices $(0,0)$, $(0,1)$, and $(1,1)$ (including the interior). Suppose that $P = (X,Y)$ is a point chosen uniformly at random inside of T.

(a) What is the joint density function of (X,Y)? Use this to compute $\mathrm{Cov}(X,Y)$.

Proof. The joint density of (X,Y) is

$$f_{(X,Y)} = \begin{cases} \dfrac{1}{\text{Area}(T)} = 2 & \text{if } (x,y) \in T; \\ 0 & \text{if } (x,y) \notin T. \end{cases}$$

We compute the covariance using the formula $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$:

$$\begin{aligned} \mathbb{E}[XY] &= \int_0^1 \int_x^1 2xy\,dydx \\ &= \int_0^1 \left(xy^2\right)\Big|_x^1 dy \\ &= \int_0^1 x - x^3\,dy \\ &= \left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Big|_0^1 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}. \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{E}[X] &= \int_0^1 \int_x^1 2x\,dydx \\ &= \int_0^1 2x - 2x^2\,dx \\ &= \left(x^2 - \frac{2x^3}{3}\right)\Big|_0^1 = 1 - \frac{2}{3} = \frac{1}{3} \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[Y] &= \int_0^1 \int_x^1 2y\,dydx \\ &= \int_0^1 \left(y^2\Big|_x^1\right) dx \\ &= \int_0^1 1 - x^2\,dx \\ &= \left(x - \frac{x^3}{3}\right)\Big|_0^1 = 1 - \frac{1}{3} = \frac{2}{3}. \end{aligned}$$

So, $\mathrm{Cov}(X,Y) = \frac{1}{4} - \frac{1}{3}\frac{2}{3} = \frac{1}{4} - \frac{2}{9} > 0$. ∎

(b) Determine if X and Y are independent.

Proof. X and Y are not independent because $\mathrm{Cov}(X, Y) \neq 0$. □

5. Suppose that we roll a fair six-sided die until we roll a 6, at which point we stop. Let N be the number of times that we rolled an odd number before we stopped. For example, we could have the sequence of rolls $(1, 3, 4, 1, 2, 6)$, in which case $N = 3$. Compute the expectation $\mathbb{E}[N]$.

Proof. Note that

$$N = \sum_{i=1}^{\infty} \mathbb{1}_{E_i},$$

where $\mathbb{1}_{E_i}$ is the indicator function on the event

$$E_i = \{\text{odd number on the } i\text{th roll and no roll of 6 beforehand}\}.$$

By independence, we have

$$\mathbb{P}(E_i) = \left(\frac{5}{6}\right)^{i-1} \frac{1}{2}.$$

By the linearity of expectation, we have

$$\mathbb{E}[N] = \sum_{i=1}^{\infty} \mathbb{E}[\mathbb{1}_{E_i}] = \sum_{i=1}^{\infty} \left(\frac{5}{6}\right)^{i-1} \frac{1}{2} = \frac{1}{2} \frac{1}{1 - \frac{5}{6}} = \frac{1}{2} 6 = 3.$$

□

6. Suppose that we have i.i.d. random variables $X_1, X_2, \dots$ with mean zero $\mathbb{E}[X_1] = 0$ and unit variance $\mathrm{Var}(X_1) = 1$. Determine the following limits with precise justifications.

(a)

$$\lim_{n\to\infty} \mathbb{P}\left(-\frac{n}{4} \le X_1 + \cdots + X_n < \frac{n}{2} \right)$$

First proof. Note that

$$\begin{aligned} 1 &\ge \mathbb{P}\left(-\frac{n}{4} \le X_1 + \cdots + X_n < \frac{n}{2} \right) \\ &= \mathbb{P}\left(-\frac{1}{4} \le \frac{X_1 + \cdots + X_n}{n} < \frac{1}{2} \right) \\ &\ge \mathbb{P}\left(-\frac{1}{4} \le \frac{X_1 + \cdots + X_n}{n} \le \frac{1}{4} \right). \end{aligned}$$

By the law of large numbers,

$$\lim_{n\to\infty} \mathbb{P}\left(-\frac{1}{4} \le \frac{X_1 + \cdots + X_n}{n} \le \frac{1}{4} \right) = 1.$$

This implies that

$$\lim_{n\to\infty} \mathbb{P}\left(-\frac{n}{4} \le X_1 + \cdots + X_n < \frac{n}{2} \right) = 1.$$

□

Second proof. Denote $S_n := X_1 + X_2 + \cdots + X_n$ and fix $a > 0$. It follows from the Central Limit Theorem that

$$\lim_{n\to\infty} P\Big(- a \le \frac{S_n}{\sqrt{n}} \le a\Big) = \Phi(a) - \Phi(-a),$$

where $\Phi(\cdot)$ is the CDF of a standard normal random variable. Now,

$$P\Big(- \frac{n}{4} \le S_n \le \frac{n}{2}\Big) = P\Big(- \frac{\sqrt{n}}{4} \le \frac{S_n}{\sqrt{n}} \le \frac{\sqrt{n}}{2}\Big).$$

Take n_0 such that $\frac{\sqrt{n_0}}{4} > a$. Then for any $n > n_0$

$$P\Big(- \frac{\sqrt{n}}{4} \le \frac{S_n}{\sqrt{n}} \le \frac{\sqrt{n}}{2}\Big) \ge P\Big(- a \le \frac{S_n}{\sqrt{n}} \le a\Big),$$

and this inequality holds after taking the limit

$$\lim_{n\to\infty} P\Big(- \frac{\sqrt{n}}{4} \le \frac{S_n}{\sqrt{n}} \le \frac{\sqrt{n}}{2}\Big) \ge \lim_{n\to\infty} P\Big(- a \le \frac{S_n}{\sqrt{n}} \le a\Big) = \Phi(a) - \Phi(-a).$$

The above lower bound holds for any fixed $a > 0$. In particular, by taking $a > 0$ arbitrarily large, we have that

$$\lim_{n\to\infty} P\Big(- \frac{\sqrt{n}}{4} \le \frac{S_n}{\sqrt{n}} \le \frac{\sqrt{n}}{2}\Big) = 1.$$

□

(b)
$$\lim_{n\to\infty} \mathbb{P}(X_1 + \cdots + X_n = 0).$$

Proof. Note that
$$\begin{aligned}\mathbb{P}(X_1 + \cdots + X_n = 0) &= \mathbb{P}\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}} = 0\right) \\ &= \mathbb{P}\left(0 \leq \frac{X_1 + \cdots + X_n}{\sqrt{n}} \leq 0\right).\end{aligned}$$

By the central limit theorem,
$$\lim_{n\to\infty} \mathbb{P}\left(0 \leq \frac{X_1 + \cdots + X_n}{\sqrt{n}} \leq 0\right) = \Phi(0) - \Phi(0) = 0.$$

□