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 $\mathbb{P}(X < Y)$.

Solution.

$$P(X < Y) = \sum_{k=1}^{\infty} \sum_{l=k+1}^{\infty} P(X = k, Y = l). \tag{1}$$

Since random variables X and Y are independent,

$$P(X = k, Y = l) = P(X = k)P(Y = l) = p(1-p)^{k-1}q(1-q)^{l-1}. \tag{2}$$

Note that for any $k \geq 1$

$$\sum_{l=k+1}^{\infty} q(1-q)^{l-1} = (1-q)^k, \tag{3}$$

therefore

$$P(X < Y) = \sum_{k=1}^{\infty} \sum_{l=k+1}^{\infty} P(X = k, Y = l) \tag{4}$$

$$= \sum_{k=1}^{\infty} \sum_{l=k+1}^{\infty} p(1-p)^{k-1}q(1-q)^{l-1} \tag{5}$$

$$= \sum_{k=1}^{\infty} p(1-p)^{k-1}(1-q)^k \tag{6}$$

$$= p(1-q)\frac{1}{1-(1-p)(1-q)} \tag{7}$$

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

(a) (10 points) Find the CDF of Y.

(b) (5 points) Is Y discrete, continuous, or neither? If discrete, find the p.m.f. If continuous, find the density. If neither, explain why.

Solution.

(a) Compute the CDF $F_Y(t)$ using the definition

$$F_Y(t) = P(Y \leq t) = P(X^2 \leq t) = P(-\sqrt{t} \leq X \leq \sqrt{t}). \tag{8}$$

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 $t \in (0, 1)$,

$$P(-\sqrt{t} \leq X \leq \sqrt{t}) = \frac{2\sqrt{t}}{3}. \tag{9}$$

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

$$P(-\sqrt{t} \leq X \leq \sqrt{t}) = P(-\sqrt{t} \leq X \leq 1) = \frac{1+\sqrt{t}}{3}. \tag{10}$$

Finally,

$$F_Y(t) = \begin{cases} 0, & t < 0, \\ \frac{2\sqrt{t}}{3}, & t \in [0,1), \\ \frac{1+\sqrt{t}}{3}, & t \in [1,4), \\ 1, & t \geq 4. \end{cases} \tag{11}$$

(b) The CDF is continuous, therefore Y is a continuous random variable. We compute its PDF by differentiating the CDF

$$f_Y(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{3\sqrt{t}}, & t \in (0,1), \\ \frac{1}{6\sqrt{t}}, & t \in (1,4), \\ 0, & t > 4. \end{cases} \tag{12}$$

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

Solution. Denote by N_0, N_1, N_2, N_3 the digits of the random number N, so that

$$N = N_3 \cdot 10^3 + N_2 \cdot 10^2 + N_1 \cdot 10 + N_0 \cdot 1. \tag{13}$$

Random variables N_0, N_1 and N_2 are uniformly distributed on the set $\{0, 1, \dots 9\}$, while N_3 is uniformly distributed on $\{0, 1, 2, 3, 4\}$. Moreover, $X = N_0 + N_1 + N_2 + N_3$. Using the linearity of the expectation

$$E(X) = E(N_0) + E(N_1) + E(N_2) + E(N_3) = 4.5 + 4.5 + 4.5 + 2 = 15.5\,. \tag{14}$$

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 $\mathrm{Cov}(X,Y)$.
(b) Determine if X and Y are independent.

Solution.

(a) The point (X,Y) is uniformly distributed on the triangle T with area $1/2$, therefore the joint density of (X,Y) is given by

$$f_{X,Y}(x,y) = 2\mathbb{1}_T(x,y), \tag{15}$$

where $\mathbb{1}_T$ is the indicator function of the set T. Using the joint density, we compute $E(X)$ and $E(Y)$

$$E(X) = \int_0^1 \int_x^1 2x \, dydx = \int_0^1 2x(1-x)\, dx = \frac{1}{3}, \tag{16}$$

$$E(Y) = \int_0^1 \int_x^1 2y \, dydx = \int_0^1 (1-x^2)\, dx = \frac{2}{3}. \tag{17}$$

Finally, in order to compute $\mathrm{Cov}(X,Y)$, we need $E(XY)$

$$E(XY) = \int_0^1 \int_x^1 2xy\, dy dx = \int_0^1 x(1-x^2)\, dx = \frac{1}{4}, \tag{18}$$

which gives

$$\mathrm{Cov}(X,Y) = \frac{1}{4} - \frac{1}{3} \cdot \frac{2}{3} = \frac{1}{36}. \tag{19}$$

(b) If two random variables are independent, then their covariance is equal to zero. Since $\mathrm{Cov}(X,Y) \neq 0$, we conclude that random variables X and Y are not independent.

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 $\mathbb{E}[N]$.

Solution. Denote by A_i the event that the i-th roll gives an odd number and that all previous rolls did not give a six. Then we can represent N in terms of the indicators of the events A_i

$$N = \sum_{i=1}^{\infty} \mathbb{1}_{A_i}. \tag{20}$$

From the linearity of expectation,

$$E(N) = \sum_{i=1}^{\infty} E(\mathbb{1}_{A_i}) = \sum_{i=1}^{\infty} P(A_i). \tag{21}$$

For any $i \geq 1$,

$$P(A_i) = \frac{1}{2} \cdot \left(\frac{5}{6}\right)^{i-1}. \tag{22}$$

Plugging (22) into (21) gives

$$E(N) = \sum_{i=1}^{\infty} \frac{1}{2} \cdot \left(\frac{5}{6}\right)^{i-1} = \frac{1}{2} \cdot 6 = 3. \tag{23}$$

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

(a)

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

(b)

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

Solution.

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

$$\lim_{n\to\infty} P\Big(-a \le \frac{S_n}{\sqrt{n}} \le a\Big) = \Phi(a) - \Phi(-a), \tag{24}$$

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

$$P\Big(-\frac{n}{4} \le S_n \le \frac{n}{2}\Big) = P\Big(-\frac{\sqrt{n}}{4} \le \frac{S_n}{\sqrt{n}} \le \frac{\sqrt{n}}{2}\Big). \tag{25}$$

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

$$P\Big(-\frac{\sqrt{n}}{4} \le \frac{S_n}{\sqrt{n}} \le \frac{\sqrt{n}}{2}\Big) \ge P\Big(-a \le \frac{S_n}{\sqrt{n}} \le a\Big), \tag{26}$$

and this inequality holds after taking the limit

$$\lim_{n\to\infty} P\Big(-\frac{\sqrt{n}}{4} \le \frac{S_n}{\sqrt{n}} \le \frac{\sqrt{n}}{2}\Big) \ge \lim_{n\to\infty} P\Big(-a \le \frac{S_n}{\sqrt{n}} \le a\Big) = \Phi(a) - \Phi(-a). \tag{27}$$

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

$$\lim_{n\to\infty} P\Big(-\frac{\sqrt{n}}{4} \le \frac{S_n}{\sqrt{n}} \le \frac{\sqrt{n}}{2}\Big) \ge 1. \tag{28}$$

We conclude that

$$\lim_{n\to\infty} P\Big(-\frac{n}{4} \le S_n \le \frac{n}{2}\Big) = 1. \tag{29}$$

(b) Fix a small number $\varepsilon > 0$. From the Central Limit Theorem

$$\lim_{n\to\infty} P\Big(-\varepsilon \le S_n/\sqrt{n} \le \varepsilon\Big) = \Phi(\varepsilon) - \Phi(-\varepsilon). \tag{30}$$

Since for any $n \ge 1$

$$P(S_n = 0) \le P\Big(-\varepsilon \le S_n/\sqrt{n} \le \varepsilon\Big), \tag{31}$$

the inequality also holds after taking the limit $n \to \infty$

$$\lim_{n\to\infty} P(S_n = 0) \le \lim_{n\to\infty} P\Big(-\varepsilon \le S_n/\sqrt{n} \le \varepsilon\Big) = \Phi(\varepsilon) - \Phi(-\varepsilon). \tag{32}$$

The value $\Phi(\varepsilon) - \Phi(-\varepsilon)$ can be made arbitrarily small by taking ε small enough. We thus conclude that

$$\lim_{n\to\infty} P(S_n = 0) = 0. \tag{33}$$