University of Toronto Scarborough

Department of Computer & Mathematical Sciences

STAB52H3 Introduction to Probability

December 2020 Final Examination

Instructors: Leonard Wong and Sotirios Damouras
Duration: 2hours

Examination aids allowed: Open notes/books, non-programmable scientific calculator

Last Name: ______________________________
First Name: ______________________________
Student #: ______________________________

Instructions:

- Read the questions carefully and answer only what is being asked.
- Show your intermediate work, and write clearly and legibly.
- **TOTAL POINTS ADD TO 120, BUT YOU WILL BE GRADED OUT OF 100 (i.e. there is 20% Extra Credit)**

Question:	1	2	3	4	5	6	7	8	Total
Points:	12	12	18	18	12	12	18	18	120
Score:									

1. Consider a binary communication channel for exchanging messages encoded in *bits*, i.e. sequences of 0's and 1's. The channel is "noisy", in the sense that 0's have a 10% chance of being flipped to 1's during transmission, and 1's have a 5% chance of being flipped to 0's, *independently* of other bits in the message.

   (a) (4 points) If you send the message $(1, 0, 1)$ through the channel, find the probability it is received correctly.

   (b) (8 points) If you receive the message $(1, 1)$, find the probability that this was the actual message that was sent through the channel. For you answer, assume that all four 2-bit messages, namely (0, 0), (0, 1), (1, 0), (1, 1), are equally likely to have been sent through the channel.

2. An exam consists of 4 questions, where each question has one of two possible variations.

   (a) (2 points) How many distinct exams are there? (exams are distinct if at least one question has different variations.)

   (b) (4 points) Each student is randomly assigned one of the above versions of the exam. Find the probability that two students get *identical* exams (i.e. all 4 questions have the same variations).

   (c) (6 points) Find the probability that two students get exams which differ by exactly two questions (i.e. have 2 questions with the same variations, and 2 questions with different variations).

3. Due to natural variation the weight of an apple, denoted by W, follows a distribution (not necessarily Normal) with a mean of 250grams (gr) and a standard deviation of 20gr.

   (a) (6 points) Find an *upper bound* on the probability that a randomly chosen apple weighs 300gr or more; make the bound as tight as possible by using all of the available information.

   (b) (6 points) Assume you pick two apples whose weights (W_1, W_2) are *negatively correlated*, with correlation coefficient $\rho = -0.5$. Find an upper bound on the probability that their *average* weight $(\frac{W_1+W_2}{2})$ is greater or equal to 300gr. Again, make the bound as tight as possible by using all of the available information.

   (c) (6 points) Finally, consider a box containing $n = 36$ apples, where all apple weights are *independent*. Calculate an *approximate* probability that the whole box weighs more than 9,120gr, making use of the Central Limit Theorem.
   Below are values of the Standard Normal CDF $\Phi(z) = P(Z \leq z)$ (where $Z \sim N(0, 1)$):

$$\begin{Bmatrix} \Phi(.5) = 0.6914625 & \Phi(1) = 0.8413447 & \Phi(1.5) = 0.9331928 & \Phi(2) = 0.9772499 \\ \Phi(2.5) = 0.9937903 & \Phi(3) = 0.9986501 & \Phi(3.5) = 0.9997674 & \Phi(4) = 0.9999683 \end{Bmatrix}$$

4. In a board game, a pawn starts at the origin (0,0) and moves randomly along a square lattice, according to a sequence of *independent, fair* coin tosses: every time you get Heads you move to the right (i.e. increase x-coordinate by $+1$), and every time you get Tails you move up (i.e. increase y-coordinate by $+1$); e.g. after tossing (H, T, H, H), your pawn will be at position (3,1). Define the random variable D_n to be the *square distance* of your pawn from the origin after n tosses; e.g. $D_4(H, T, H, H) = 3^2 + 1^2 = 10$.

   (a) (3 points) Find the probability that after 4 tosses, your pawn is at position (2,1).

   (b) (3 points) Find the probability that after 4 tosses, your pawn is at position (2,2).

(c) (6 points) Find the PMF of D_4.

(d) (6 points) Find the expected value of D_n, as a function of n.
(*Hint*: Use moments of the Binomial distribution.)

5. (12 points) Let Y_1 and Y_2 be i.i.d. Geometric(p) random variables in terms of the number of failures before the first success. Explicitly, you are given that

$$\mathbb{P}(Y_i = y) = (1-p)^y p, \quad y = 0, 1, \ldots.$$

Derive the probability mass function of $X = Y_1 + Y_2$. (Hint: This is a discrete version of convolution.)

6. (12 points) Let X and Y be independent exponential random variables with rates 3 and 1 respectively (so $X \sim \mathrm{Exp}(3)$ and $Y \sim \mathrm{Exp}(1)$). Find the probability $\mathbb{P}(X \leq cY)$ for $c > 0$.

7. Let $X_1, X_2, \ldots$ be i.i.d. random variables such that $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$.

(a) (4 points) Show carefully that the moment generating function of X_i is given by $M(t) = \frac{1}{2}(e^t + e^{-t}), \quad t \in \mathbb{R}$.

(b) (8 points) You are given that for $t \in \mathbb{R}$ we have the inequality $\frac{1}{2}(e^t + e^{-t}) \leq e^{\frac{t^2}{2}}$. Define $Z_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$. Use (a) and the above inequality to show that for any $z > 0$ and $t > 0$ we have the inequality $\mathbb{P}(Z_n \geq z) \leq e^{-tz + \frac{t^2}{2}}$.
(Hint: Apply Markov's inequality to e^{tZ_n}.)

(c) (6 points) Let $z > 0$ be fixed. By (b), since $\mathbb{P}(Z_n \geq x) \leq e^{-tx - \frac{t^2}{2}}$ for all $t > 0$, we have $\mathbb{P}(Z_n \geq z) \leq \min_{t>0} e^{-tz + \frac{t^2}{2}}, \quad z > 0$. By finding the minimum value of the right hand side, show that for $z > 0$ we have the upper bound $\mathbb{P}(Z_n \geq z) \leq e^{\frac{-z^2}{2}}$ that holds for all $n \geq 1$.

8. Let U_1 and U_2 be i.i.d. Uniform$(0, 3)$ random variables. (Note that the interval is $[0, 3]$, not $[0, 1]$.) Consider the random variable V defined by

$$V = (U_1 - U_2) 1_{\{U_1 \geq U_2\}} = \begin{cases} U_1 - U_2, & \text{if } U_1 \geq U_2; \\ 0, & \text{otherwise.} \end{cases}$$

(a) (6 points) Find $\mathbb{P}(U_1 > 2 | U_1 \geq U_2)$. (Hint: Draw a picture and read off the areas. You do not have to integrate explicitly.)

(b) (3 points) Write down the joint density of (U_1, U_2) and specify the domain in $\mathbb{R}^2$ on which the joint density is positive.

(c) (9 points) Find the expected value of V. (Hint: You only need to integrate on the region corresponding to the condition $\{U_1 \geq U_2\}$.)