AUBURN UNIVERSITY

Stat 5600/6600 - Probability and Statistics for Data Science

Midterm Exam #2 November 16, 2023

Instructions:

1. Write your name and the class you enrolled (i.e., 5600 or 6600) by your name.
2. Highest possible score in this exam is 100. You must always **explain your answers** and **show your work** to receive **full credit** or proper **partial credit**. If necessary, you can use both sides of the sheets, but make sure you have indicated doing so.
3. There are 4 questions. Please make sure you have the file with the 4 questions.
4. You are permitted both sides of a 8.5-11 sheet containing formulas, notes, or anything from the course as deemed helpful. You will turn in your exam. You may keep your note/formula sheet. No other aids of any kind are permitted.
5. Calculators are permitted but must be supplied by you. Cell phones are not permitted and should be turned off and put away before the exam begins.
6. You must always **explain your answers** and **show your work** to receive **full credit** or proper **partial credit**. If necessary, you can use the back of these pages, but make sure you have indicated doing so.
7. Q1(c) and Q2(d) are optional for STAT 5600 students, so they don't need to turn in the solution/answer for it.

Name: ____________________________________

Signature: ____________________________________

Question	1	2	3	4	5	Total (out of 100)
Score						

Note: The place-holder for Q5 above is for the take-home component of the exam.

Questions

Q1. (20 points) Let X and Y be two indicator random variables (r.v.) with

$$p_x = P(X = 1) \quad \text{and} \quad p_y = P(Y = 1).$$

(a) Show that $W = XY$ is also an indicator r.v.. Determine when W is 1 and when it is 0.

(b) What condition(s) (on X and Y) are required for $S = X + Y$ to be an indicator random variable? Express $p_s = P(S = 1)$ in terms of p_x and p_y when X and Y are independent.

(c) Suppose X and Y are the indicator r.v.'s described in the question and also are **independent**. Are the statements below true or false. Justify your answers:

(i) if $E(XY) = 0$ then $E(X + Y) \leq 1$

(ii) if $E(X + Y) \leq 1$ then $E(XY) = 0$

Q2. (20 points)
(a) Suppose we are (independently) tossing a coin, which has probability of heads $0 < p < 1$, k times (where k is fixed). Let N_k denote the number of heads in the k tosses. Determine $\mathrm{Var}(N_k)$. Is it increasing or decreasing in k (i.e., does $\mathrm{Var}(N_k)$ increase or decrease as k increases)?

(b) Let R_k denote the proportion of heads among the same k tosses in part (a). Derive $\mathrm{Var}(R_k)$. Is it increasing or decreasing in k?

(c) Suppose two independent random variables X and Y, have expectations p and q, respectively, and variances r and s, respectively. Find $\mathrm{Cov}(X, XY)$ in terms of p, q, r and s.

(d) Find $\mathrm{Var}\,(X + XY)$ for the independent random variables X and Y in part (c). (**Hint:** You may use the following without deriving them: (i) For any two random variables U and V, $\mathrm{Var}(U + V) = \mathrm{Var}(U) + \mathrm{Var}(V) + 2\,\mathrm{Cov}(U, V)$ and (ii) For the random variables X and Y in part (c), $\mathrm{Var}(XY) = rs + rq^2 + sp^2$ without deriving it.)

Q3. (20 points)

(a) Say a game consists of rolling a die until the player accumulates 15 dots (i.e. sum of the dots are ≥ 15). Let X be the number of rolls until the player accumulates 15 dots. Write the support of distribution of X. Does X have a geometric distribution? Why or why not?

(b) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ population. Using the MGF of normal distribution and properties of MGF, show that the sample mean $\bar{X}_n$ is normally distributed as $N(\mu, \sigma^2/n)$. **Hints:** For $X \sim N(\mu, \sigma^2)$, the MGF is $M_X(t) = \exp(\mu t + \sigma^2 t^2/2)$. Also, it might help to find the MGF of sum $S = \sum_{i=1}^{n} X_i$ first.

Q4. (20 points) Suppose the random variable X has density (i.e. pdf)

$$f_X(x) = \begin{cases} c\,x & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the value of c that makes this density a valid pdf.

(b) The hazard function for a random variable X is defined to be $f_X(x)/[1 - F_X(x)]$. Derive the hazard function $h_X(x)$ for the random variable X for $0 < x < 1$.

(c) Find the variance $\text{Var}(X)$.

(d) Find the conditional probability $P(X < 3/4 \mid X > 1/4)$.