STATS 116: Homework 3

Due: Thursday, July 20, 2023 at 10:00 pm PDT on Gradescope

There are 7 problems on this assignment, each worth 8 points, although subparts within a problem may not be equally weighted. Credit will be assigned primarily based on reasoning and work, not the final answer. You do not need to simplify arithmetic expressions unless otherwise noted. While you may discuss the problems on this assignment other students, you must write up your own solutions. As per the syllabus, you may occasionally use the Internet or other public resources to clarify concepts with citation when this information is used as part of your own solution to a homework problem. However, you may not search for direct solutions to any problems assigned for homework or exams. For example, you can ask ChatGPT to clarify a particular concept from lecture that may be related to a problem, but you cannot feed it any part of a course assignment or a substantively similar version.

1. (BH 3.3) Let X be a random variable with CDF F, and $Y = \mu + \sigma X$, where μ and σ are real numbers with $\sigma > 0$. Find the CDF F_Y of Y, in terms of F.

   Fix $y \in \mathbb{R}$. We have

$$F_Y(y) = P(Y \leq y) = P(\mu + \sigma X \leq y) = P\left(X \leq \frac{y-\mu}{\sigma}\right) = F\left(\frac{y-\mu}{\sigma}\right)$$

2. (BH 3.20) Suppose that a lottery ticket has probability p of being a winning ticket, independently of other tickets. A gambler buys 3 tickets, hoping this will triple the chance of having at least one winning ticket.

   (a) What is the distribution of how many of the 3 tickets are winning tickets?

   By the story of the Binomial, the distribution is $\text{Bin}(3, p)$.

   (b) Show that the probability that at least 1 of the 3 tickets is winning is $3p - 3p^2 + p^3$ in two different ways: by using inclusion-exclusion, and by taking the complement of the desired event and then using the PMF of a certain named distribution.

   Let A_i be the event that the i-th ticket wins, for $i = 1, 2, 3$. By inclusion-exclusion and symmetry, we have

$$P(A_1 \cup A_2 \cup A_3) = 3P(A_1) - \binom{3}{2} P(A_1 \cap A_2) + P(A_1 \cap A_2 \cap A_3) = 3p - 3p^2 + p^3$$

Alternatively, using the binomial PMF directly, we have

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{3}{0} p^0 (1-p)^3 = 1 - (1-p)^3 = 3p - 3p^2 + p^3$$

(c) Show that the gambler's chances of having at least one winning ticket do not quite triple (compared with buying only one ticket), but that they do approximately triple if p is small.

By the union bound (Homework 1), we know

$$P(A_1 \cup A_2 \cup A_3) \leq 3P(A_1) = 3p$$

Alternatively we can directly argue that $3p - (3p - 3p^2 + p^3) = 3p^2 - p^3 \geq 0$ for all $p \in [0, 1]$. For small p, p^2 and p^3 are negligible compared to p, so the probability does approximately triple for small p.

3. Suppose X and Y are random variables for which $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$ for all real x and y. In class, we mentioned that this implies X and Y are independent, i.e. $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for all subsets A and B of $\mathbb{R}$. In this problem, we will show a weaker conclusion. For simplicity, you may assume X and Y are discrete and both supported on the set of integers $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$, though this is not actually necessary for these statements to hold.

(a) Show that $P(X > x, Y > y) = P(X > x)P(Y > y)$ for any x, y.

Fix real x, y. We have by LOTP that

$$\begin{aligned}
P(X > x, Y > y) = P(X > x) - P(X > x, Y \leq y) &= (1 - P(X \leq x)) - (P(Y \leq y) - P(X \leq x, Y \leq y)) \\
&= 1 - P(X \leq x) - P(Y \leq y) + P(X \leq x)P(Y \leq y) \\
&= (1 - P(X \leq x))(1 - P(Y \leq y)) \\
&= P(X > x)P(Y > y)
\end{aligned}$$

where the second equality uses the given condition $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$.

(b) Show that $P(X = x, Y = y) = P(X = x)P(Y = y)$ for any x, y.

Fix real x, y. If either is not an integer, both sides of the equality are trivially 0, so we can assume x and y are integers. Now

$$\begin{aligned}
P(X = x, Y = y) &= P(X = x, Y \leq y) - P(X = x, Y \leq y - 1) \\
&= (P(X \leq x, Y \leq y) - P(X \leq x - 1, Y \leq y)) \\
&- (P(X \leq x, Y \leq y - 1) - P(X \leq x - 1, Y \leq y - 1)) \\
&= (P(X \leq x) - P(X \leq x - 1))P(Y \leq y) - P(Y \leq y - 1)(P(X \leq x) - P(X \leq x - 1)) \\
&= P(X = x)P(Y \leq y) - P(Y \leq y - 1)P(X = x) \\
&= P(X = x)P(Y = y)
\end{aligned}$$

where the third equality uses the given condition $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$.

4. (BH 3.30) A company with n women and m men as employees is deciding which employees to promote.

   (a) Suppose for this part that the company decides to promote t employees, where $1 \leq t \leq n + m$, by choosing t random employees (with equal probabilities for each set of t employees). What is the distribution of the number of women who get promoted?

   By the story of the Hypergeometric, the distribution is HGeom(n, m, t) (the women are "white balls" and the men are "black balls").

   (b) Now suppose that instead of having a predetermined number of promotions to give, the company decides independently for each employee, promoting the employee with probability p. Find the distributions of the number of women who are promoted, the number of women who are not promoted, and the number of employees who are promoted.

   Let W be the number of women who are promoted and T be the number of employees who are promoted. By the story of the Binomial, $W \sim \text{Bin}(n, p)$ women are promoted, $n - W \sim \text{Bin}(n, 1 - p)$ women are not promoted, and $T \sim \text{Bin}(n + m, p)$ employees are promoted.

5. (BH 4.26) Suppose Nick is flipping a nickel with probability p_1 of Heads and Penny is (independently) flipping a penny with probability p_2 of Heads. Let $X_1, X_2, \ldots,$ be Nick's results and $Y_1, Y_2, \ldots$ be Penny's results, with $X_i \sim \text{Bern}(p_1)$ and $Y_j \sim \text{Bern}(p_2)$.

   (a) Find the distribution and expected value of the first time at which they are simultaneously successful, i.e., the smallest n such that $X_n = Y_n = 1$. Hint: Define a new sequence of Bernoulli trials (coin flips).

Let N be the time this happens. Then $N \sim \text{FS}(p_1 p_2)$ by the story of the First Success, and $\mathbb{E}[N] = \boxed{\dfrac{1}{p_1 p_2}}$.

(b) Find the expected time until at least one has a success (including the success). Hint: Define a new sequence of Bernoulli trials.

Each flip, the probability neither is successful is $(1 - p_1)(1 - p_2)$, independently of the other flips. Thus if T is the time until at least one has a success, we have $T \sim \text{FS}(1-(1-p_1)(1-p_2))$ and hence $\mathbb{E}[T] = \boxed{\dfrac{1}{1-(1-p_1)(1-p_2)}}$.

(c) For $p_1 = p_2$, find the probability that their first successes are simultaneous, and use this to find the probability that Nick's first success precedes Penny's.

Let T_1 and T_2 be the first times at which Nick and Penny are successful, respectively. Let $p = p_1 = p_2$ and $q = 1 - p$. Then noting T_1 and T_2 are independent $\text{FS}(p)$ random variables we have (recalling the formula for the sum of a geometric series) that

$$P(T_1 = T_2) = \sum_{n=1}^{\infty} P(T_1 = n, T_2 = n) = \sum_{n=1}^{\infty} q^{n-1} p \cdot q^{n-1} p = \frac{p^2}{1-q^2} = \frac{p}{2-p}$$

Alternatively, consider the first trial where at least one of Nick and Penny succeeded. Their first successes are simultaneous if and only if they both succeed at this trial. So the probability is

$$P(\text{both succeed} \mid \text{ at least one succeeds}) = \frac{p^2}{1-q^2} = \frac{p}{2-p}$$

Next, by symmetry we have

$$1 = P(T_1 = T_2) + P(T_1 < T_2) + P(T_1 > T_2) = P(T_1 = T_2) + 2P(T_1 < T_2)$$

Solving gives

$$P(T_1 < T_2) = \frac{1}{2}\left(1 - \frac{p}{2-p}\right) = \boxed{\frac{1-p}{2-p}}$$

6. The game of **wizards** is played with a standard deck of 52 cards, augmented with two jokers and six "wizard" cards for a total of 60 cards. At the beginning of the first round, the cards are shuffled a single card is dealt to each of three players, including you. The cards are played in sequence, starting with the first player, which is you. The rule is that the highest ranked card of the same suit as the card played by the first player (you) wins the round (ranks go from 2 at the lowest up to Ace at the highest), with the following exceptions:

- The first player to play a wizard always wins
- Spades are a "trump" suit, meaning that if the card played by the first player (you) is a diamond, heart, or club, then other players can win if they play a spade. If there are spades (but no wizards), the highest spade wins
- Anyone who plays a joker cannot win the round

Each player must declare (without seeing the other players' cards) whether they think they will win. A player gets -10 points if they declare incorrectly. They get 20 points if they correctly declare they will win, and 10 points if they correctly declare they will lose. You are dealt the King of Hearts and you'd like to maximize your expected score. Should you declare that you will win the trick?

Parsing the rules, we see that we *lose* the trick if either of the other players plays one of the following cards: Ace of Hearts, any of the 13 spades, or any of the 6 wizards. We win if neither of the other players plays one of these cards. There are $\binom{59}{2}$ total choices for the cards that the other two players have. Out of these, $\binom{39}{2}$ lead to me winning (there are 39 cards that do not beat my king of hearts). Thus, by the naive definition, the probability I win the trick is

$$\frac{\binom{39}{2}}{\binom{59}{2}} = \frac{39 \cdot 38}{59 \cdot 58} \equiv p \approx 0.43$$

If I declare that I win, my expected score is $20 \cdot p - 10(1-p) = 30p - 10$. If I declare that I lose, my expected score is $10 \cdot (1-p) - 10 \cdot p = 10 - 20p$. We can check that the expected score from declaring winning is larger whenever $p > 0.4$, so indeed I should declare that I will win.

7. The probability generating function of a random variable X supported on the nonnegative integers is denoted by G_X and given by $G_X(s) = \mathbb{E}[s^X]$ for all values of s where the expectation exists.

(a) Suppose $X \sim \text{Bern}(p)$. Find $G_X(s)$ for any real s.

We compute via LOTUS

$$G_X(s) = \mathbb{E}[s^X] = s^0 \cdot (1-p) + s^1 \cdot p = \boxed{1 - p + sp}$$

(b) Now suppose $Y \sim \text{Pois}(\lambda)$. Find $G_Y(s)$ for any real s.

By LOTUS again, we have

$$G_Y(s) = \mathbb{E}[s^Y] = \sum_{y=0}^{\infty} s^y \exp(-\lambda) \frac{\lambda^y}{y!} = \exp(-\lambda) \sum_{y=0}^{\infty} \frac{(s\lambda)^y}{y!} = \exp(-\lambda)\exp(s\lambda) = \boxed{\exp(\lambda(s-1))}$$

(c) Show that $p_X(0) = G_X(0)$ where p_X is the PMF of X. Hint: Recall $0^0 = 1$.

By LOTUS, we have
$$G_X(0) = \mathbb{E}[0^X] = \sum_{x=0}^{\infty} 0^x \cdot p_X(x) = p_X(0)$$
since $0^x = 0$ for all $x = 1, 2, \ldots$.

(d) Show that $\mathbb{E}[X] = G'_X(1)$, where $G'_X(s)$ is the derivative of the function G_X evaluated at s.

We write
$$G_X(s) = \sum_{x=0}^{\infty} s^x p_X(x)$$
Taking the derivative with respect to s on both sides, we get
$$G'_X(s) = \sum_{x=0}^{\infty} x s^{x-1} p_X(x)$$
Plugging in $s = 1$ and noting $1^x = 1$ for all nonnegative integers x, we conclude
$$G'_X(1) = \sum_{x=0}^{\infty} x p_X(x) = \mathbb{E}[X]$$
by the definition of expectation.