STATS 116: Midterm exam

Monday, July 24, 2023 from 10:30 am to 11:20 am PDT

Name ____

SUNET ID ____

There are 3 problems on this examination, worth a total of 48 points. The point value of each problem is given below. Subparts within a problem may not carry equal weight. You are not expected to completely solve all of the problems within the time limit, so do your best.

This examination is closed book, with the exception of one standard size (8.5 inch by 11 inch) sheet of paper which may contain any information placed on it prior to the start of the examination. As per the Honor Code and syllabus, any collaboration with other students or any other individuals is strictly prohibited, as is the use of electronic devices during the examination (other than to check the time).

If a problem subpart depends on the answer to a previous subpart, you may receive full credit for this subpart without solving the previous subpart by expressing your answer in terms of the answer to the previous subpart. You may use the back page of each problem if you need more space; please indicate this on the main page with the problem if you do so, to reduce the probability that it is missed during grading. Unless you are explicitly asked to simplify, you may leave your answer in terms of binomial coefficients. Even if you are asked to simplify, you will receive most of the credit with a correct answer that you are unable to simplify either algebraically or via a "story." Good luck!

Problem 1	__________	out of 12
Problem 2	__________	out of 24
Problem 3	__________	out of 12
Total	__________	out of 48

Problem 1

In the card game Dominion, you start with a deck of 7 Copper cards (each worth \$1) and 3 Estate cards (each worth \$0), randomly shuffled. For the first turn, you draw the top five cards of this deck. For the second turn, you draw the remaining five cards in the deck.

(a) The coveted 2-5 split occurs when a player draws 2 Coppers on the first turn (thereby drawing 5 Coppers on the second turn), or vice versa (5 Coppers on the first turn and 2 Coppers on the second turn). What is the probability of getting a 2-5 split? Simplify.

There are $\binom{10}{5}$ equally likely combinations of 5 cards that make up the first turn (treating all 10 cards as distinguishable, but assuming order does not matter). Of these, $\binom{7}{2} \cdot \binom{3}{3}$ contain 2 Coppers and 3 Estates. By symmetry, the probability of 2 Coppers in the first turn is the same as the probability of 2 Coppers on the second turn, which is in fact that probability of 5 Coppers in the first turn. Hence by the naive definition, the desired probability is

$$2 \cdot \frac{\binom{7}{2} \cdot \binom{3}{3}}{\binom{10}{5}} = \frac{2 \cdot \frac{7!}{2!5!}}{\frac{10!}{5!5!}} = \frac{7!5!}{10!} = \frac{120}{8 \cdot 9 \cdot 10} = \boxed{\frac{1}{6}}.$$

(b) After turn 2, your starting deck is shuffled along with two Silver cards, each worth \$2, to create a 12-card deck. In the third turn, you draw the top five cards of this new shuffled deck. On average, what is the total dollar value of the cards you draw in turn 3? Simplify.

Let $I_1, \ldots, I_7$ be the indicators that each copper is drawn in turn 3 and J_1, J_2 be the indicators that each silver is drawn in turn 3. Each card has probability $5/12$ of being drawn in turn 3, so $\mathbb{E}[I_i] = \mathbb{E}[J_j] = 5/12$ for each i and j. The total dollar value of the cards in turn 3 is $I_1 + \cdots + I_7 + 2J_1 + 2J_2$, so by linearity the answer is $11 \cdot 5/12 = \boxed{55/12}$.

Problem 2

Jack and Jimmy are playing a short game of ping pong. Jack wins each point with probability $p \in (0,1)$ with $p \neq 1/2$, independently of the results of all other points. The first person to win 3 points wins the game.

(a) What is the probability that the game requires 5 points to be played?

Note that 5 points are played if and only if Jack wins exactly 2 of the first 4 points. The number of points Jack wins out of the first 4 is $\text{Bin}(4, p)$ by the story of the Binomial. So by the Binomial PMF, the desired probability is

$$\binom{4}{2} \cdot p^2 \cdot (1-p)^2 = \boxed{6p^2(1-p)^2}$$

(b) Given that the game requires 5 points to be played, what is the probability that Jack wins the game?

Given that the game requires 5 points, Jack wins the game if and only if Jack wins the 5th point. Thus the desired probability is

$$P(\text{Jack wins 5th point} \mid \text{Jack wins 2 of the first 4 points}) = P(\text{Jack wins 5th point}) = \boxed{p}$$

since the outcome of the 5th point is independent of the outcome of the first 4 points.

(c) What is the overall probability that Jack wins the game?

Let X be the number of points played and W be the event that Jack wins the game. By LOTP

$$P(W) = P(W, X = 3) + P(W, X = 4) + P(W \mid X = 5)P(X = 5)$$

The event $W, X = 3$ is equivalent to Jack winning the first 3 points, which has probability p^3. The event $W, X = 4$ is equivalent to Jack winning the 4th point and 2 of the first three points, which happens with probability $p \cdot \binom{3}{2} p^2(1-p) = 3p^3(1-p)$ by independence of the 4th point and the first three points. With $P(W \mid X = 5) = p$ and $P(X = 5) = 6p^2(1-p)^2$ from the previous parts, we conclude

$$P(W) = \boxed{p^3 + 3p^3(1-p) + 6p^3(1-p)^2}$$

(d) Is the indicator random variable of the event that Jack wins the game independent of the number of points played? Why or why not?

Not for general p. Using the notation from the previous part, we have $P(I_W = 1 \mid X = 3) = \frac{p^3}{p^3+(1-p)^3}$ by the definition of conditional probability, since $X = 3$ if and only if Jack wins all 3 points or Jimmy wins all 3 points. However, $P(I_W = 1 \mid X = 5) = p$ by part (b). If I_W were independent of X then we'd have both $P(I_W = 1 \mid X = 3)$ and $P(I_W = 1 \mid X = 5)$ equal to $P(I_W = 1) = P(W)$. But we have $P(I_W = 1 \mid X = 3) \neq P(I_W = 1 \mid X = 5)$ since $p \neq 1/2$; thus I_W cannot be independent of X.

For the remainder of the problem, suppose that the game is now played until Jack wins 3 points.

(e) What is the mean and variance of the number of points played?

The number of points Jimmy wins by the time Jack wins 3 points follows a $\text{NBin}(3, p)$ distribution. Hence if X is the total number of points played, then $X - 3 \sim \text{NBin}(3, p)$. The mean and variance of a $\text{NBin}(3, p)$ random variable are $3(1-p)/p$ and $3(1-p)/p^2$, respectively. Hence

$$\mathbb{E}[X] = 3 + \mathbb{E}[X-3] = 3 + 3\frac{1-p}{p} = \boxed{\frac{3}{p}}, \quad \text{Var}(X) = \text{Var}(X-3) = \boxed{\frac{3(1-p)}{p^2}}$$

(f) Now suppose instead that $p = 1/2$. Given that the game lasted a total of 10 points (including the 3 points Jack won), what is the probability that Jack never won two points in a row? Simplify.

With $p = 1/2$, all possible sequences of winners for the 10 points are equally likely. With the given information, our sample space is the set of all such sequences where Jack wins the last point and 2 of the first 9 points. This is $\binom{9}{2} = 36$ sequences. Of these, the sequences where Jack never won 2 points in a row are those where the 9th point went to Jimmy and out of the first 8 points, Jack's 2 points were not consecutive. There are $\binom{8}{2} = 28$ sequences for the first 8 points with 2 wins for Jack; of these, 7 have Jack's 2 wins consecutive. Thus the final answer is $\boxed{21/36}$ by the naive definition.

Problem 3

Martha has 300 red balls and 500 green balls in a jar. She randomly removes balls from the jar one at a time. Let X be the number of green balls she draws before getting her first red ball.

(a) Let F be the CDF of X. Compute $F(x)$ for all real x. Your answer can be in terms of the floor function: let $\lfloor x \rfloor$ denote the greatest integer less than or equal to x. Hint: since X must take on integer values, $P(X > k) = P(X \geq k+1)$ for any integer k.

We begin by noting $\operatorname{supp}(X) = \{0, 1, \ldots, 500\}$. For each $k \in \operatorname{supp}(X)$ we have

$$P(X \leq k) = 1 - P(X > k) = 1 - P(X \geq k+1) = 1 - \frac{\binom{500}{k+1}}{\binom{800}{k+1}}$$

since the event $X \geq k+1$ is equivalent to the first $k+1$ balls drawn being green (note $\binom{500}{501} := 0$). Then for arbitrary real x we have

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - \frac{\binom{500}{\lfloor x \rfloor + 1}}{\binom{800}{\lfloor x \rfloor + 1}} & 0 \leq x < 500 \\ 1 & x \geq 500 \end{cases}$$

(b) Find $\mathbb{E}[X]$ (simplify).

This problem performs the same computation as the example in Lecture 12 where we computed the expected number of cards before drawing the first ace. Number the green balls from 1 to 500. For $i = 1, \ldots, 500$, let I_i be the indicator r.v. for the event A_i that green ball number i appears before any of the red balls. Then $X = I_1 + \cdots + I_{500}$. Since $P(A_i) = 1/301$ by symmetry, by linearity and the "fundamental bridge" we have

$$\mathbb{E}[X] = \sum_{i=1}^{500} P(A_i) = \boxed{\frac{500}{301}}$$

Partial credit is given by deriving the PMF from the answer in part (a) and then writing the expectation as a sum using the definition of expectation.

Table of distributions

Below is some information about some named distribution families that might be useful. Note: below, the letter q denotes the quantity $1 - p$.

Name	Param.	PMF	Mean	Variance
Bernoulli	p	$P(X=1) = p, P(X=0) = q$	p	pq
Binomial	n, p	$\binom{n}{k} p^k q^{n-k}$, for $k \in \{0, 1, \dots, n\}$	np	npq
FS	p	pq^{k-1}, for $k \in \{1, 2, \dots\}$	$1/p$	q/p^2
Geom	p	pq^k, for $k \in \{0, 1, 2, \dots\}$	q/p	q/p^2
NBinom	r, p	$\binom{r+n-1}{r-1} p^r q^n, n \in \{0, 1, 2, \dots\}$	rq/p	rq/p^2
HGeom	w, b, n	$\frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$, for $k \in \{0, 1, \dots, n\}$	$\mu = \frac{nw}{w+b}$	$(\frac{w+b-n}{w+b-1})\mu(1 - \frac{\mu}{n})$
Poisson	λ	$\frac{e^{-\lambda}\lambda^k}{k!}$, for $k \in \{0, 1, 2, \dots\}$	λ	λ