Stanford Stats 117 Midterm Examination

Closed book and single sheet of paper
Duration: 120 minutes

Fall 2024

Name:[1] ______________________________

Student ID Number: ______________________________

By taking this exam, you agree to be bound by the Stanford Honor Code, meaning specifically in this context that you

- will not give or receive aid in examinations;
- will do your share and take an active part in seeing to it that others as well as yourself uphold the spirit and letter of the Honor Code.

Signature: ______________________________

[1] It might be good to write your name on each page of your exam

Question M.1 (True/False): Mark each question as true or false. No explanation needed.

(a) **(4pts)** For any random variable X, the expected value $\mathbb{E}[X]$ is one that X may take.

(b) **(4pts)** For any random variable X, there is always one most probable value of X.

(c) **(4pts)** There are discrete random variables X that may take on infinitely many values.

(d) **(4pts)** If X and Y are independent random variables, then for any sets A and B,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

(e) **(4pts)** In a round-robin tournament of n players (each player plays each other exactly once), there are $n(n-1)$ games.

Question M.2 (**10pts**): There are 24 people at tennis team tryouts one evening in April, where the sun is getting low in the sky in the west; the tennis courts are aligned so one side faces west and the other east. Each player is paired with one other player in a match. How many ways are there for players to be matched, assuming that in each game it *does matter* who faces the sun? *Hint.* If there are $2n$ players, assign each a number in $1, \ldots, 2n$. Then pair the players off.

Question M.3: Let X and Y be discrete random variables, each taking on values $\{a, b, c\}$. Mark each of the following as "guaranteed to hold," meaning that the statement as written must be true; "cannot hold," meaning that the statement as written must be false; or "not enough information to determine," meaning that the result may change depending on the particular joint probabilities. You may provide a one sentence or less justification for your answer. For shorthand, let $P(X = a, Y = a) = p_{aa}$, $P(X = a, Y = b) = p_{ab}$, and so on.

(a) **(6pts)** Let X and Y have joint p.m.f.

		$Y =$		
		a	b	c
$X =$	a	p_{aa}	p_{ab}	p_{ac}
	b	p_{ba}	p_{bb}	p_{bc}
	c	p_{ca}	p_{cb}	p_{cc}

Then X and Y are independent.

(b) **(6pts)** Let X and Y have joint p.m.f.

		$Y =$		
		a	b	c
$X =$	a	0	p_{ab}	0
	b	0	p_{bb}	0
	c	0	p_{cb}	0

Then X and Y are independent.

(c) **(6pts)** Let X and Y have joint p.m.f.

		$Y =$		
		a	b	c
$X =$	a	0	p_{ab}	p_{ac}
	b	p_{ba}	0	p_{bc}
	c	p_{ca}	p_{cb}	0

where each non-zero entry (i.e., p_{ab}, p_{ac}, p_{ba}, etc.) is positive. Then X and Y are independent.

Question M.4: Recall the St. Pertersburg paradox, where we play a game by flipping a (fair) coin until Tails appears, doubling our wealth at each iteration of the game. So if X is our wealth, then for $k = 0, 1, 2, \ldots$, $P(X = 2^k) = \frac{1}{2^{k+1}}$. Suppose you play this in a casino with a maximum payout M, meaning that if you hit wealth $X \geq M$, your payout is simply M (i.e., you get $\min\{X, M\}$).

(a) **(20pts)** Assuming that M is a power of 2, give

$$\mathbb{E}[\min\{X, M\}].$$

(b) **(2pts)** If $M = 2^{30} = 1073741824$ (about \$1 billion), how much would you pay to play this game? Explain your answer in 1 sentence.

Question M.5 (15pts): A student who took Stats117 last year comes to talk shop with you, and states the following: "Let's suppose $X \sim \mathsf{Binomial}(n, p)$ and $Y \sim \mathsf{Binomial}(n, q)$. Then X has p.m.f. p_X and Y has p.m.f. p_Y, and I can make a combination random variable Z with p.m.f.

$$p_Z(z) = P(Z = z) = \frac{1}{2}p_X(z) + \frac{1}{2}p_Y(z).$$

The cool thing is that this is still binomial and $Z \sim \mathsf{Binomial}(n, \frac{p+q}{2})$." Is this student correct? If so, explain why, and if not, give a counterexample.

Question M.6: Let X be the number of purchases that Professor Duchi will purchase from `lego.com` over his lifetime. Suppose that the p.m.f. of X is $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$, that is, $X \sim \mathsf{Poisson}(\lambda)$.

(a) **(10pts)** Find $P(X \geq 1)$ and $P(X \geq 2)$ *without* summing an infinite series.

(b) **(10pts)** Lego only knows about individuals who have bought at least one item from `lego.com` (someone who has never made a purchase does not appear in the customer database). When Lego computes the number of purchases for everyone in the database, the data are draws from X *given* that $X \geq 1$. Give the p.m.f. of X conditional on $X \geq 1$.

Question M.7 (20pts): Suppose we have two tests for the presence of a disease. We would like to know when positive results on both tests increases the probability of having the disease over a positive result on only a single test. Let D be the event that an individual has the disease, and let T_1 and T_2 be, respectively the events that test 1 and test 2 are positive. Show that

$$P(D \mid T_1, T_2) \geq P(D \mid T_1) \text{ if and only if } P(T_2 \mid T_1, D) \geq P(T_2 \mid T_1).$$

(You may assume all probabilities are positive.)

Question M.8* (5pts, Extra credit): In the setting of Question M.7, assume that T_1 and T_2 are independent conditional on disease status. Show that

$$P(D \mid T_1, T_2) \geq P(D \mid T_1) \text{ if and only if } P(T_2 \mid D) \geq P(T_2 \mid D^c).$$

Explain your result in a sentence or two.

SCRATCH PAGE

SCRATCH PAGE