STATS 116: Practice midterm exam

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

Name ____

SUNET ID ____

There are 4 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 12

Problem 3__________out of 12

Problem 4__________out of 12

Total__________out of 48

Problem 1

A certain baby cries if and only if she is hungry, tired, or both. Let C be the event that the baby is crying, H be the event that she is hungry, and T be the event that she is tired. Let $P(C) = c$, $P(H) = h$, and $P(T) = t$, where none of c, h, t are equal to 0 or 1. Suppose that H and T are independent.

(a) Find c, in terms of h and t. (Simplify fully.)

Note $C = H \cup T$, so by inclusion-exclusion and independence of H and T, we have

$$c = P(C) = P(H \cup T) = P(H) + P(T) - P(H \cap T) = \boxed{h + t - ht}$$

(b) Find $P(H \mid C)$, $P(T \mid C)$, and $P(H, T \mid C)$ (in terms of c, h, t; simplify fully).

By the definition of conditional probability, we have

$$\begin{aligned} P(H \mid C) &= \frac{P(H \cap C)}{P(C)} = \frac{P(H)}{P(C)} = \boxed{\frac{h}{c}} \\ P(T \mid C) &= \frac{P(T \cap C)}{P(C)} = \frac{P(T)}{P(C)} = \boxed{\frac{t}{c}} \\ P(H, T \mid C) &= \frac{P(H, T, C)}{P(C)} = \frac{P(H, T)}{P(C)} = \boxed{\frac{ht}{c}} \end{aligned}$$

(c) Are H and T conditionally independent given C? Justify your answer in two ways: algebraically using the quantities from (b), and with a brief but clear intuitive explanation in words.

If H and T were conditionally independent given C, then we'd have $P(H, T \mid C) = P(H \mid C)P(T \mid C)$. However, by the previous part we have

$$P(H, T \mid C) = \frac{ht}{c} \neq \frac{ht}{c^2} = \frac{h}{c}\frac{t}{c} = P(H \mid C)P(T \mid C)$$

since $c < 1$. Thus H and T are not conditionally independent given C. Intuitively, note that the knowledge that the baby is crying tells you the baby could be hungry, tired, or both. But then additionally knowing the baby is not hungry (e.g. since you just fed them) tells you for certain that the baby is tired. Thus, conditional on crying, hunger status tells you something about tiredness. Note this problem precisely corresponds to Berkson's paradox from Homework 2.

Problem 2

A publishing company employs two proofreaders, Prue and Frida. When Prue is proofreading a book, for each typo she has probability p of catching it and $q = 1 - p$ of missing it, independently. When Frida is proofreading a book, for each typo she has probability f of catching it and $g = 1 - f$ of missing it, independently.

(a) A certain book draft has n typos. The company randomly assigns it to one of the two proofreaders, with equal probabilities. Specify the distribution of the number of typos that get detected.

Let N be the number of typos. Clearly $\text{supp}(N) = \{0, 1, \dots, n\}$. Let F be the event that Frida is assigned. Then by LOTP, for each $k \in \text{supp}(N)$ we have

$$P(N = k) = P(N = k \mid F)P(F) + P(N = k \mid F^c)P(F^c) = \frac{1}{2}\left(\binom{n}{k} f^k g^{1-k} + \binom{n}{k} p^k q^{1-k}\right)$$

as the number of typos caught given Frida is assigned is $\text{Bin}(n, f)$ and the number of typos caught given Prue is assigned in $\text{Bin}(n, p)$, by the story of the Binomial. Note that N itself does NOT follow a Binomial distribution unless $p = f$.

(b) Another book is being written. When a draft of the book is complete, it will have a $\text{Pois}(\lambda)$ number of typos, and will be assigned to Prue to proofread. Find the probability that Prue catches exactly k typos, given that she misses exactly m typos (assume k and m are nonnegative integers). Hint: First compute the probability that Prue catches exactly k typos *and* she misses exactly m typos, and note this expression is valid for all nonnegative integers k and m.

Let X be the number of typos Prue catches and Y be the number of typos Prue misses. Note $X + Y \sim \text{Pois}(\lambda)$. Then

$$P(X = k, Y = m) = P(X = k, X + Y = k + m) = P(X = k \mid X + Y = k + m)P(X + Y = k + m)$$

by the definition of conditional probability. But

$$P(X = k \mid X + Y = k + m) = \binom{k+m}{k} p^k q^m$$

by the Binomial PMF, and hence

$$P(X = k, Y = m) = \binom{k+m}{k} p^k q^m \cdot \exp(-\lambda)\lambda^{k+m}/(k+m)! = \exp(-\lambda)\frac{1}{k!m!}(\lambda p)^k(\lambda q)^m$$

Then by LOTP

$$P(Y = m) = \sum_{k=0}^{\infty} P(X = k, Y = m) = \exp(-\lambda)\frac{(\lambda q)^m}{m!}\sum_{k=0}^{\infty}\frac{(\lambda p)^k}{k!} = \exp(-\lambda q)\frac{(\lambda q)^m}{m!}$$

so by the definition of conditional probability

$$P(X = k \mid Y = m) = \frac{P(X = k, Y = m)}{P(Y = m)} = \frac{\exp(-\lambda)\frac{1}{k!m!}(\lambda p)^k(\lambda q)^m}{\exp(-\lambda q)\frac{(\lambda q)^m}{m!}} = \boxed{\exp(-\lambda p)\frac{(\lambda p)^k}{k!}}$$

Problem 3

The United States Senate has 100 senators: 2 from each of the 50 states. Each state has a junior senator and a senior senator (based on which of them has served longer). A committee of size 20 is formed randomly, with all sets of 20 senators equally likely.

(a) Specify the distribution of the number of junior senators on the committee.

The number of junior senators follows a $\boxed{\text{HGeom}(50, 50, 20)}$ distribution (the junior senators are the "white" balls, the senior senators are the "black" balls, and we take a sample without replacement of size 20).

(b) Find the expected number of junior senators on the committee (simplify).

Let $J \sim \text{HGeom}(50, 50, 20)$ be the number of junior senators on the committee. By part (a) and the given table of distributions, we have

$$\mathbb{E}[J] = \frac{20 \cdot 50}{50 + 50} = \boxed{10}$$

(c) Find the expected number of states such that both senators from that state are on the committee (simplify).

Let $I_i, i = 1, \ldots, 50$ be the indicator for the event A_i that both senators from state i are on the committee. Then the number of states for which both senators are on the committee is $X = I_1 + \cdots + I_{50}$. By the naive definition of probability, we have

$$\mathbb{E}[I_i] = P(A_i) = \frac{\binom{98}{18}}{\binom{100}{20}} = \frac{\frac{98!}{18!80!}}{\frac{100!}{20!80!}} = \frac{98!}{100!} \cdot \frac{20!}{18!} = \frac{20 \cdot 19}{100 \cdot 99} = \frac{19}{495}$$

(let the sample space be all possible collections of senators). Then by linearity

$$\mathbb{E}[X] = \sum_{i=1}^{50} \mathbb{E}[I_i] = \frac{50 \cdot 19}{495} = \boxed{\frac{190}{99}}$$

Problem 4

Let $N \sim \mathrm{FS}(p)$, with $\frac{1}{2} < p < 1$.

(a) Find $\mathbb{E}[2^N]$ (simplify).

By LOTUS we have

$$\mathbb{E}[2^N] = \sum_{n=1}^{\infty} 2^n (1-p)^{n-1} p = \frac{p}{1-p} \sum_{n=1}^{\infty} [2(1-p)]^n = \frac{p}{1-p} \cdot \frac{2(1-p)}{1-2(1-p)} = \boxed{\frac{2p}{2p-1}}$$

(b) Explain in words why the answer to (a) makes sense in the extreme cases where p is very close to $1/2$ and where p is very close to 1. (You can do this part even if you haven't solved (a), by giving an intuitive argument for what you think the answer should look like in these extreme cases.)

When p is very close to $1/2$, $\mathbb{E}[2^N]$ approaches infinity. This makes sense as in that case, 2^N is the payoff in a game where your payoff doubles for every consecutive tail you get in flips of a (roughly) fair coin. Each flip has probability about $1/2$ of landing tails, so the weighted infinite sum of payoff times probability of that payoff does not have decaying terms.

When p approaches 1, $\mathbb{E}[2^N]$ approaches $\frac{2 \cdot 2}{2-1} = 2$. This makes sense as with p near 1, we have $N = 1$ (and hence $2^N = 2$) with very high probability.

(c) Let F be the CDF of the indicator random variable for the event $N > 3$. Find F, being sure to specify $F(x)$ for all real x.

We can think of N as the number of flips of a coin landing heads with probability p until the first heads are flipped (including that last head). Thus the event $N > 3$ is the same as the event that the first three coins land tails, i.e. $P(N > 3) = (1-p)^3$. This could be also computed by summing the PMF. We conclude that $I_{N>3} \sim \mathrm{Bern}((1-p)^3)$ and hence its CDF is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - (1-p)^3 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

by our knowledge of what the CDF of a Bernoulli distribution looks like from lecture.

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