STATS 116: Final exam

Saturday, August 19, 2023 from 8:30 am to 11:30 am PDT, Hewlett 102

Name ____

SUNET ID ____

There are 8 problems on this examination, worth a total of 100 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 two standard size (8.5 inch by 11 inch) sheets 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 or arithmetic expressions (but not sums or integrals, unless otherwise stated). 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 5__________out of 12
Problem 2__________out of 16	Problem 6__________out of 12
Problem 3__________out of 12	Problem 7__________out of 12
Problem 4__________out of 12	Problem 8__________out of 12

Total __________ out of 100

Problem 1

Let X, Y, and Z be i.i.d. random variables with $P(X > 0) = 1$. Write the most appropriate of $\leq$, $\geq$, $=$, or ? in each blank (where "?" means that no relation holds in general). Justification determines most of the credit in each part.

(a) $\mathbb{E}\left(\frac{X}{X+Y+Z}\right)$ _____ $1/3$

(b) $\mathbb{E}\left(\frac{X}{Y+Z}\right)$ _____ $1/2$

(c) $P(X + Y \geq 2)$ _____ $\mathbb{E}(X^2)$

(d) $P(X + Y \leq 2)$ _____ $(P(X \leq 1))^2$

Problem 2

The *Laplace* distribution with parameter $b > 0$ is a continuous distribution with PDF

$$f(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right), \quad x \in \mathbb{R}$$

Suppose $X_1, \ldots, X_{900}$ are i.i.d. Laplace(b) random variables for some $b > 0$.

(a) Show that $|X_1| \sim \text{Expo}(1/b)$.

(b) Compute $\text{Var}(X_1)$ (in terms of b). Simplify.

(c) If $b = 1$, give as accurate an approximation as you can of the probability that no more than 420 of the X_i are outside $[-\log(2), \log(2)]$ in terms of e and/or the Standard Normal CDF Φ.

(d) If $b = 1$, give as accurate an approximation as you can of the probability that at least 2 of the X_i are outside $[-6, 6]$ in terms of e and/or the Standard Normal CDF Φ.

Problem 3

From his extensive studies, Kevin knows that he has three pills that, when ingested by any rat, will each cure memorylessness in that rat with probabilities 0.5, 0.3, and 0.1, respectively.

(a) Because Kevin himself suffers from memorylessness, he has lost track of which pill is which, and so simply picks one pill at random to give to his first rat. Given that the first rat is cured, what is the probability the *second* rat, which is given one of the two remaining pills at random, receives the least effective pill?

(b) Kevin's third rat gets the last pill. What is the *unconditional* mean and variance of the number of rats (out of 3) that are cured (i.e. do not condition on the event that the first rat is cured, as in part (a))?

Problem 4

Kerrie the Archer throws a dart at a uniformly random point in the unit circle (the circle with radius 1 centered at (0,0) in the coordinate plane).

(a) Let R be the distance from the point where the dart lands to the origin. Find $\mathbb{E}(R)$ and $\text{Var}(R)$. Hint: You might consider starting by deriving the CDF of R.

(b) Now suppose Kerrie throws another dart uniformly at random on the circle, independent of the previous dart. Show that on average, the two darts are no more than 1 unit apart. Hint: It may be helpful to first compute the expected *squared* distance between the darts.

Problem 5

Jimmy is playing the game Ensemble Stars. His goal is to win a treasure chest. To do so, he can pull a lever where each pull will win him the treasure chest with probability 0.01, independently of the other pulls. Let $X \sim \mathrm{FS}(0.01)$ be the number of pulls Jimmy needs to win the treasure chest.

(a) Show that X satisfies the following memoryless property: For all nonnegative integers t and s, we have

$$P(X > t + s \mid X > t) = P(X > s)$$

(b) Now suppose Jimmy has a power-up that guarantees he will get the chest on the 300th pull if he hasn't received it already. With this power-up, what is the expected number of pulls Jimmy needs until he gets the chest (including the pull where he receives the chest)?

Problem 6

Jessica wants to understand the distribution of income in a large city. She does this by attempting to survey n people in the city, chosen at random. Unfortunately, not everyone likes reporting their income to strangers so many of her observations are missing. Let M_i be the indicator r.v. of the event that the income Y_i of individual $i = 1, \dots, n$ in Jessica's sample is missing. Suppose that Jessica works at the national bank and so she knows the bank account balance $X_1, \dots, X_n$ for all individuals she attempts to survey. Further assume that $P(M_i = 1 \mid X_i) = e(X_i)$ for some known "missingness propensity function" e with $e(X) \leq 1 - \delta$ with probability 1 for some $\delta > 0$, and also that the triples $(M_1, X_1, Y_1), \dots, (M_n, X_n, Y_n)$ are i.i.d. with M_i is conditionally independent of Y_i given X_i. Let $\mu = \mathbb{E}(Y_1)$ and $\sigma^2 = \mathrm{Var}(Y_1)$, both finite.

(a) Briefly explain in words what it means for M_i to be conditionally independent of Y_i given X_i in the context of this problem.

(b) Let $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n \frac{Y_i(1-M_i)}{1-e(X_i)}$. Explain why $\bar{Y}_n$ can be computed using known quantities, and show that $\mathbb{E}(\bar{Y}) = \mu$.

(c) Argue that $\bar{Y}_n$ satisfies a weak law of large numbers as n gets large, i.e. that $P(|\bar{Y}_n - \mu| > \epsilon) \to 0$ as $n \to \infty$, for any $\epsilon > 0$.

Problem 7

Suppose Z_1 and Z_2 are i.i.d. Standard Normal random variables. Let $X_1 = Z_1$ and $X_2 = aZ_1 + \sqrt{1-a^2}Z_2$ for some $a \in (0,1)$.

(a) Compute the probability that X_1 and X_2 differ (in absolute value) by at least 1 in terms of Φ, the Standard Normal CDF.

(b) Compute $\text{Cor}(X_1, X_2)$.

(c) Find the joint PDF of (X_1, X_2), and conclude that X_1 and X_2 are exchangeable.

Problem 8

San Flan is a dangerous square-shaped city measuring 7 miles by 7 miles, divided into $14^2 = 196$ equally sized square neighborhoods, each measuring 0.5 miles by 0.5 miles. Suppose various crimes occur at locations uniformly distributed within the boundaries of San Flan, independently of previous crimes, and that the total number of crimes that occur on a given day in San Flan follows a Pois(100) distribution.

(a) What is the mean and standard deviation in the number of crimes that occur on the day in a particular neighborhood, Catpatch?

(b) On average, how many neighborhoods would we expect to record at least one crime?

(c) What is the correlation between the number of crimes on the day in Catpatch and the total number of crimes that occur in all of San Flan?

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 or PDF	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
NBin	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
Uniform	$a < b$	$\frac{1}{b-a}$, for $x \in (a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	μ, σ^2	$\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2
Log-Normal	μ, σ^2	$\frac{1}{x\sigma\sqrt{2\pi}}e^{-(\log x-\mu)^2/(2\sigma^2)}, x > 0$	$\theta = e^{\mu+\sigma^2/2}$	$\theta^2(e^{\sigma^2}-1)$
Expo	λ	$\lambda e^{-\lambda x}$, for $x > 0$	$1/\lambda$	$1/\lambda^2$
Gamma	a, λ	$\Gamma(a)^{-1}(\lambda x)^a e^{-\lambda x}x^{-1}$, for $x > 0$	a/λ	a/λ^2
Beta	a, b	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$, for $0 < x < 1$	$\mu = \frac{a}{a+b}$	$\frac{\mu(1-\mu)}{a+b+1}$
Chi-Square	n	$\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$, for $x > 0$	n	$2n$

The function Γ is given by

$$\Gamma(a) = \int_0^\infty x^a e^{-x}\frac{dx}{x}$$

for all $a > 0$. For any $a > 0$, $\Gamma(a+1) = a\Gamma(a)$. We have $\Gamma(n) = (n-1)!$ for n a positive integer, and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.