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

Answer: $\boxed{=}$. Since (X, Y, Z) are i.i.d., they are exchangeable. Thus (X, Y, Z), (Y, Z, X), and (Z, X, Y) have the same (joint) distribution. Applying the function $g(x, y, z) = x/(x + y + z)$ to each then shows that $X/(X + Y + Z)$, $Y/(X + Y + Z)$, and $Z/(X + Y + Z)$ have the same distribution, hence the same expectation. But by linearity

$$\mathbb{E}\left(\frac{X}{X+Y+Z}\right) + \mathbb{E}\left(\frac{Y}{X+Y+Z}\right) + \mathbb{E}\left(\frac{Z}{X+Y+Z}\right) = \mathbb{E}\left(\frac{X+Y+Z}{X+Y+Z}\right) = 1$$

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

Answer: $\boxed{\geq}$. Note X and $1/(Y + Z)$ are independent, hence uncorrelated. Then by Jensen

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

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

Answer: $\boxed{\leq}$. We have

$$P(X + Y \geq 2) = P((X + Y)^2 \geq 4) \leq \frac{\mathbb{E}[(X + Y)^2]}{4}$$

by Markov's inequality. But

$$\begin{aligned}\mathbb{E}[(X + Y)^2] &= \mathbb{E}[X^2] + \mathbb{E}[2XY] + \mathbb{E}[Y^2] \text{ by linearity} \\ &= \mathbb{E}[X^2] + 2(\mathbb{E}[X])^2 + \mathbb{E}[X^2] \text{ since } X, Y \text{ are i.i.d.} \\ &\leq 4\mathbb{E}[X^2] \text{ since } (\mathbb{E}[X])^2 \leq \mathbb{E}[X^2] \text{ by Jensen}\end{aligned}$$

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

Answer: $\boxed{\geq}$. Note $X \leq 1$ and $Y \leq 1$ implies $X + Y \leq 2$. Hence

$$P(X + Y \leq 2) \geq P(X \leq 1, Y \leq 1) = P(X \leq 1)P(Y \leq 1) = (P(X \leq 1))^2$$

where the last two inequalities use the fact that X and Y are i.i.d.

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

Let $Y = |X_1|$. For each $y > 0$ we have

$$\begin{aligned} P(|Y| \le y) &= P(-y \le X_1 \le y) \\ &= \frac{1}{2b} \int_{-y}^{y} \exp\left(-\frac{|x|}{b}\right) dx \\ &= \frac{1}{b} \int_{0}^{y} \exp\left(-\frac{x}{b}\right) dx \\ &= \frac{1}{b} \left[-b \exp\left(-\frac{x}{b}\right)\right] \Big|_{x=0}^{x=y} = 1 - \exp\left(-\frac{y}{b}\right) \end{aligned}$$

Thus the PDF of Y is given by $f_Y(y) = \frac{1}{b} \exp\left(-\frac{y}{b}\right)$ for all $y > 0$, which matches the Expo($1/b$) PDF.

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

Note the PDF is even ($f(x) = f(-x)$) so $\mathbb{E}(X_1) = 0$. Hence using part (a) and the known mean and variance of the Expo(b) distribution, we conclude

$$\text{Var}(X_1) = \mathbb{E}(X_1^2) = \mathbb{E}(|X_1|^2) = \text{Var}(|X_1|) + (\mathbb{E}(|X_1|))^2 = \boxed{2b^2}$$

(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.

The probability each X_i is outside $[-\log(2), \log(2)]$ is $P(|X_i| \geq \log(2)) = \exp(-\log(2)) = 1/2$ since $|X_i| \sim \text{Expo}(1)$ by part (a). Let N be the number of X_i outside $[-\log(2), \log(2)]$; we have $N \sim \text{Bin}(900, 1/2)$. By the Normal approximation of the Binomial via the CLT, we know N is approximately $\mathcal{N}(450, 225)$. Including the continuity correction, for $Z \sim \mathcal{N}(0, 1)$

$$P(N \leq 420) \approx P(-0.5 \leq 15Z + 450 \leq 420.5) = P\left(-450.5/15 \leq Z \leq -29.5/15\right)$$
$$= \boxed{\Phi\left(-\frac{29.5}{15}\right) - \Phi\left(-\frac{450.5}{15}\right)}$$

The second term is not necessary for the approximation since it is so small.

(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.

The probability each X_i is outside $[-6, 6]$ is $P(|X_i| \geq 6) = \exp(-6)$ since $|X_i| \sim \text{Expo}(1)$ by part (a). Since this is a rare event, a Poisson approximation is more accurate than a Normal approximation. That is, with N now the number of X_i outside $[-6, 6]$, we have N is approximately $\text{Pois}(900 \cdot \exp(-6))$. Then the desired probability is

$$P(N \geq 2) = 1 - P(N = 0) - P(N = 1) \approx \boxed{1 - (1 + 900\exp(-6))\exp(-900\exp(-6))}$$

by the Poisson PMF.

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?

Let G be the event the first rat is cured, and let A_i, B_i, C_i be the events that the i-th rat got the pills with effectiveness probabilities 0.5, 0.3, and 0.1, respectively, which we label A, B, and C, respectively. By Bayes' rule and LOTP we compute

$$P(A_1 \mid G) = \frac{P(G \mid A_1)P(A_1)}{P(G \mid A_1)P(A_1) + P(G \mid B_1)P(B_1) + P(G \mid C_1)P(C_1)} = \frac{0.5 \cdot 1/3}{0.5 \cdot 1/3 + 0.3 \cdot 1/3 + 0.1 \cdot 1/3} = \frac{5}{9}$$

$$P(B_1 \mid G) = \frac{P(G \mid B_1)P(B_1)}{P(G)} = \frac{0.3 \cdot 1/3}{(0.5 + 0.3 + 0.1)(1/3)} = \frac{1}{3}$$

$$P(C_1 \mid G) = 1 - P(A_1 \mid G) - P(B_1 \mid G) = \frac{1}{9}$$

Now by LOTP with extra conditioning

$$\begin{aligned} P(C_2 \mid G) &= P(C_2 \mid A_1, G)P(A_1 \mid G) + P(C_2 \mid B_1, G)P(B_1 \mid G) + P(C_2 \mid C_1, G)P(C_1 \mid G) \\ &= \frac{1}{2} \cdot \frac{5}{9} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{9} \\ &= \boxed{\frac{4}{9}} \end{aligned}$$

(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))?

Let I_A, I_B, and I_C be the indicator r.v.'s for the events that the rats receiving pills A, B, and C, respectively, are cured. The number of rats that are cured can be written as $I_A + I_B + I_C$, and so by linearity

$$\mathbb{E}(I_A + I_B + I_C) = \mathbb{E}(I_A) + \mathbb{E}(I_B) + \mathbb{E}(I_C) = 0.5 + 0.3 + 0.1 = \boxed{0.9}$$

Next note I_A, I_B, and I_C are independent, so

$$\mathrm{Var}(I_A + I_B + I_C) = \mathrm{Var}(I_A) + \mathrm{Var}(I_B) + \mathrm{Var}(I_C) = 0.5(1-0.5) + 0.3(1-0.3) + 0.1(1-0.1) = \boxed{0.55}$$

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.

Let (X, Y) be the coordinates of the point where the dart lands. Then $R = \sqrt{X^2 + Y^2}$ and one could proceed via multivariate LOTUS as $f_{XY}(x, y) = 1/\pi$ for (x, y) in the unit circle. Alternatively, one could use the fact that for all $B \subseteq \mathbb{R}^2$ contained in the unit circle, $P((X, Y) \in B) = \frac{|B|}{\pi}$ since (X, Y) are uniform on the unit circle, which has area π. For each $0 \leq r \leq 1$, the event $R \leq r$ is equivalent to (X, Y) lying in the circle of radius r centered at the origin, which has area πr^2 . Thus

$$P(R \leq r) = \frac{\pi r^2}{\pi} = r^2, \quad 0 \leq r \leq 1$$

so the PDF of R is $f_R(r) = 2r$ for $r \in (0, 1)$. Then

$$\mathbb{E}(R) = \int_0^1 r f_R(r) dr = \int_0^1 2r^2 dr = \boxed{\frac{2}{3}}$$

$$\mathbb{E}(R^2) = \int_0^1 r^2 f_R(r) dr = \in_0^1 2r^3 dr = \frac{1}{2}$$

$$\text{Var}(R) = \mathbb{E}(R^2) - (\mathbb{E}(R))^2 = \boxed{\frac{1}{18}}$$

(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.

Let (X_1, Y_1) and (X_2, Y_2) be the coordinates of the two darts and $D = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$ be the distance between the darts. We first compute

$$\begin{aligned}
\mathbb{E}[D^2] &= \mathbb{E}[(X_2 - X_1)^2 + (Y_2 - Y_1)^2] \\
&= \mathbb{E}(X_1^2 + Y_1^2) + \mathbb{E}(X_2^2 + Y_2^2) - 2\mathbb{E}(X_1 X_2) - 2\mathbb{E}(Y_1 Y_2) \\
&= 2\mathbb{E}(R^2) \\
&= 1 \text{ since } \mathbb{E}(R^2) = 1 \text{ as computed above}
\end{aligned}$$

where the third equality uses the fact that as X_1 and X_2 are i.i.d. mean 0 — and same with Y_1 and Y_2, by symmetry. By Jensen we conclude $(\mathbb{E}(D))^2 \leq \mathbb{E}(D^2) = 1$, as desired.

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 $X \sim \text{FS}(0.01)$ 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

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

Note that for any nonnegative integer t, the event $X > t$ is equivalent to the event that the first t pulls failed, which has probability 0.99^t. Then using the definition of conditional probability we have

$$P(X > t + s \mid X > t) = \frac{P(X > t + s, X > t)}{P(X > t)} = \frac{P(X > t + s)}{P(X > t)} = \frac{0.99^{t+s}}{0.99^t} = 0.99^s = P(X > s)$$

for any nonnegative integers t, 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)?

We can let $Y = \min(X, 300)$ be the number of pulls Jimmy needs to get the chest with the power-up. By LOTE we have

$$\mathbb{E}(Y) = \mathbb{E}(Y \mid Y < 300)P(Y < 300) + 300P(Y = 300) = \mathbb{E}(X \mid X < 300)P(X < 300) + 300P(X \geq 300)$$

To solve for $\mathbb{E}(X \mid X < 300)$ we apply LOTE again to note

$$\begin{aligned} 100 = \mathbb{E}(X) &= \mathbb{E}(X \mid X < 300)P(X < 300) + \mathbb{E}(X \mid X \geq 300)P(X \geq 300) \\ &= \mathbb{E}(X \mid X < 300)P(X < 300) + \mathbb{E}(X \mid X > 299)P(X > 299) \\ &= \mathbb{E}(X \mid X < 300)P(X < 300) + (299 + \mathbb{E}(X))P(X > 299) \\ &= \mathbb{E}(X \mid X < 300)P(X < 300) + 399(0.99)^{299} \end{aligned}$$

where the third equality uses the memoryless property from part (a) to note $X - 299 \mid X > 299 \sim \text{FS}(0.01)$. Thus we conclude

$$\mathbb{E}(Y) = 100 - 399(0.99)^{299} + 300(0.99)^{299} = \boxed{100 - 99(0.99)^{299}}$$

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 $i = 1, \ldots, n$ in Jessica's sample is missing. Suppose that Jessica works at the national bank and so she knows the bank account balance $X_1, \ldots, X_n$ 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 $(M_1, X_1, Y_1), \ldots, (M_n, X_n, Y_n)$ 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.

Given bank account balance, missingness is independent of income. That is, among within individuals with the same bank account balance, those with higher income have the same missingness rates as those with lower income.

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

$e(X_i)$ is known and the term $Y_i(1 - M_i)$ is 0 for all missing observations ($M_i = 1$) and just the observation Y_i for all non-missing observations ($M_i = 0$). Thus $\bar{Y}$ is a function of only known quantities. We compute

$$\begin{aligned}
\mathbb{E}[\bar{Y}] &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\frac{Y_i(1-M_i)}{1-e(X_i)}\right] \text{ by linearity} \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}\left[\frac{Y_i(1-M_i)}{1-e(X_i)} \mid X_i\right]\right] \text{ by iterated expectation} \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[(1-e(X_i))^{-1}\mathbb{E}[(1-M_i) \mid X_i]\mathbb{E}[Y_i \mid X_i]\right] \text{ by Takeout and the cond. indep. assumption} \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\mathbb{E}[Y_i \mid X_i]] \text{ as } \mathbb{E}[1-M_i \mid X_i] = P(M_i = 0 \mid X_i) = 1 - e(X_i) \\
&= \mu \text{ by iterated expectation again}
\end{aligned}$$

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

One solution is to simply note $\bar{Y}$ is a sample average of the i.i.d. random variables $Z_i = Y_i(1 - M_i)/(1 - e(X_i))$ which we know have finite mean by the previous part. Thus by the strong law of large numbers, $\bar{Y}$ converges to μ almost surely, which implies the weak law of large numbers as stated in lecture. Note this argument does not require the knowledge that $\mathrm{Var}(Y) < \infty$.

Alternatively, to apply the weak law of large numbers directly we need to show the Z_i have finite variance. With $e(X_i) \leq \delta$ with probability 1, we know $1/(1 - e(X_1)) \leq \delta^{-1}$ with probability 1, so

$$\mathrm{Var}(Z_1) \leq \mathbb{E}(Z_1^2) \leq \delta^{-2}\mathbb{E}(Y_1^2(1 - M_1)^2) \leq \delta^2\mathbb{E}(Y_1^2) = \delta^2(\sigma^2 + \mu^2) < \infty$$

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.

Note $X_1 - X_2 = (1-a)Z_1 - \sqrt{1-a^2}Z_2 \sim \mathcal{N}(0, (1-a)^2 + (1-a^2)) = \mathcal{N}(0, 2(1-a))$, by recalling that the sum of independent Normals is Normal. Thus $(X_1 - X_2)/(\sqrt{2(1-a)}) \sim \mathcal{N}(0,1)$ and

$$\begin{aligned} P(|X_1 - X_2| \geq 1) &= P(X_1 - X_2 \geq 1) + P(X_1 - X_2 \leq -1) \\ &= P\left(\frac{X_1 - X_2}{\sqrt{2(1-a)}} \geq \frac{1}{\sqrt{2(1-a)}}\right) + P\left(\frac{X_1 - X_2}{\sqrt{2(1-a)}} \leq -\frac{1}{\sqrt{2(1-a)}}\right) \\ &= \boxed{2\Phi\left(-\frac{1}{\sqrt{2(1-a)}}\right)} \end{aligned}$$

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

We have

$$\text{Cov}(X_1, X_2) = \text{Cov}(Z_1, aZ_1 + \sqrt{1-a^2}Z_2) = a\text{Cov}(Z_1, Z_1) = a$$

by bilinearity and independence of Z_1, Z_2 (which implies they are uncorrelated). With $\text{SD}(X_1) = \text{SD}(X_2) = 1$ (note $X_2 \sim \mathcal{N}(0,1)$), we conclude $\text{Cor}(X_1, X_2) = \boxed{a}$ as well.

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

We derive the joint PDF of (X_1, X_2) via change of variables. Let $g(z_1, z_2) = (z_1, az_1 + \sqrt{1-a^2}z_2) \equiv (x_1, x_2)$. Solving we get $z_1 = x_1$, $z_2 = (x_2 - ax_1)/\sqrt{1-a^2}$, so

$$\left|\frac{\partial(z_1, z_2)}{\partial(x_1, x_2)}\right| = \left|\begin{bmatrix} 1 & 0 \\ -\frac{a}{\sqrt{1-a^2}} & \frac{1}{\sqrt{1-a^2}} \end{bmatrix}\right| = \frac{1}{\sqrt{1-a^2}}$$

Thus

$$\begin{aligned} f_{X_1X_2}(x_1, x_2) &= f_{Z_1Z_2}(z_1, z_2) \left| \frac{\partial(z_1, z_2)}{\partial(x_1, x_2)} \right| \\ &= \frac{1}{2\pi\sqrt{1-a^2}} \exp\left(-\frac{x_1}{2}\right) \exp\left(-\frac{(x_2 - ax_1)^2}{2(1-a^2)}\right) \\ &= \frac{1}{2\pi\sqrt{1-a^2}} \exp\left(-\frac{x_1^2 + x_2^2 - 2ax_1x_2}{2(1-a^2)}\right) \end{aligned}$$

Note $f_{X_1X_2}(x_1, x_2) = f_{X_1X_2}(x_2, x_1)$, showing exchangeability.

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?

Each crime independently has probability $1/196$ of landing in Catpatch. Thus, by the chicken-egg story, the number of crimes in Catpatch follows a Pois$(100/196)$ distribution, so the mean is $100/196 = \boxed{25/49}$ and the standard deviation is $\sqrt{100/196} = \boxed{5/7}$. Alternatively, you could use iterated expectations and law of total variance, conditioning on the number of crimes N that occurred in the city as a whole and noting that the number of crimes in Catpatch given $N = n$ follows a Bin$(n, 1/196)$ distribution.

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

Each neighborhood has probability $1 - \exp(-25/49)$ of recording at least one crime, by the Poisson PMF. Creating an indicator r.v. for the event that each neighborhood recorded at least one crime, we conclude the desired expectation is $\boxed{196(1 - \exp(-25/49))}$.

(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?

Let X be the number of crimes in Catpatch and Y be the number of crimes in San Flan *outside* Catpatch. By the chicken-egg story, we know X and Y are independent, hence uncorrelated. Then

$$\text{Cov}(X, X + Y) = \text{Cov}(X, X) = \text{Var}(X) = 25/49$$

and then

$$\text{Cor}(X, X + Y) = \frac{100/196}{\text{SD}(X)\text{SD}(X + Y)} = \frac{25/49}{5/7 \cdot 10} = \boxed{\frac{1}{14}}$$

Alternatively, use iterated expectations to compute $\mathbb{E}[X(X + Y)]$ by conditioning on $X + Y$.

Table of distributions

Below is some information about some named distribution families and the Gamma function 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 $\Gamma(\frac{1}{2})=\sqrt{\pi}$.