

# Homework 2 - Applied Stochastic Processes

kipngeno koech - bkoech

September 30, 2024

## 1. Discrete Random Variables and Real Life Applications (16 Points)

- (a) A factory produces electronic components, with each component either passing a quality check or being rejected. Let  $X$  represent the number of components that pass out of 5 tested components in a day, where the probability of each component passing the quality check is  $p = 0.8$ .

- i. (a) (2 points): Define the probability mass function (PMF) for  $X$  as a binomial distribution. Show that the total probability is 1 by summing the PMF over all possible values of  $X$ . PMF for  $X$  as a binomial distribution is given by:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

where  $n = 5$  and  $p = 0.8$ . The total probability is 1:

$$\begin{aligned} \sum_{k=0}^5 P(X = k) &= \sum_{x=0}^5 \binom{5}{x} 0.8^x (1 - 0.8)^{5-x} = 1 \\ \binom{5}{0} 0.8^0 (1 - 0.8)^5 &+ \binom{5}{1} 0.8^1 (1 - 0.8)^4 + \binom{5}{2} 0.8^2 (1 - 0.8)^3 + \binom{5}{3} 0.8^3 (1 - 0.8)^2 + \binom{5}{4} 0.8^4 (1 - 0.8)^1 + \binom{5}{5} 0.8^5 (1 - 0.8)^0 = 1 \\ 0.0032 &+ 0.0064 + 0.0512 + 0.2048 + 0.4096 + 0.32768 = 1 \\ \mathbf{1} &= \mathbf{1} \end{aligned}$$

- ii. (3 points): Suppose the factory wants to predict the likelihood of a specific number of components passing the quality check. Find the expected value and variance of  $X$ , and explain how the factory can use this information to estimate daily production quality. The expected value of  $X$  is given by:

$$E(x) = np = \text{number of trials} \times \text{probability of success}$$

$$E(X) = np = 5 \times 0.8 = \mathbf{4}$$

The variance of  $X$  is given by:

$$Var(X) = np(1 - p) = \text{number of trials} \times \text{probability of success} \times \text{probability of failure}$$

$$Var(X) = np(1 - p) = 5 \times 0.8 \times 0.2 = \mathbf{0.8}$$

- iii. (3 points): Calculate the probability that exactly 3 out of 5 components pass the quality check using the PMF. Based on this, discuss how rare or frequent it is for the factory to encounter this scenario, and explain the implications for managing production quality.

$$P(X = 3) = \binom{5}{3} 0.8^3 (1 - 0.8)^2 = \mathbf{0.2048}$$

The probability is 0.2048. This scenario is relatively common, with a 20.48% chance of occurring. The factory can use this information to manage production quality by setting performance targets and identifying unusual trends in production.

- (b) A warehouse tracks the number of defective products returned by customers daily. Let  $X$  represent the number of defective products returned on a given day. The company has noticed that on average, 4 defective products are returned daily. Based on historical data, the number of defective products returned follows a Poisson distribution with parameter  $\lambda = 4$ , which represents the average number of returns per day.

- i. (2 points): Define the PMF for  $X$ , where  $X$  follows a Poisson distribution with parameter  $\lambda = 4$ . Verify that the sum of all probabilities over the range of  $X$  (from 0 to infinity) equals 1, proving that the PMF is valid. The PMF of  $X$  as a Poisson distribution is given by:

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \text{ where } X \sim \text{Poisson}(\lambda = 4)$$

where  $\lambda = 4$ . The sum of all probabilities over the range of  $X$  is 1:

$$\begin{aligned} \sum_{k=0}^{\infty} P(X = k) &= 1 \\ \sum_{k=0}^{\infty} \frac{e^{-4} 4^k}{k!} &= 1 \end{aligned}$$

since  $e$  is a constant, we can factor it out of the sum:

$$e^{-4} \sum_{k=0}^{\infty} \frac{4^k}{k!} = 1$$

an infinite sum of the Poisson distribution is equal to  $e$  raised to the power of the parameter  $\lambda$ :

$$e^{-4} \times e^4 = 1$$

$$\mathbf{1} = \mathbf{1}$$

- ii. (3 points): Find the expected value  $E(X)$  and the variance  $\text{Var}(X)$  of the number of defective products returned. Explain the significance of these values for the warehouse's operations and how they can be used to set performance targets or identify unusual trends in returns. The expected value of  $X$  is given by:

$$E(X) = \lambda = 4$$

The variance of  $X$  is given by:

$$\text{Var}(X) = \lambda = 4$$

- iii. (3 points): Calculate the probability that the warehouse will receive fewer than 3 defective products on a given day, i.e.,  $P(X < 3)$ . Discuss what this probability means in terms of real-life decision-making for warehouse management, particularly in planning inventory and customer service responses.

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{e^{-4}4^0}{0!} + \frac{e^{-4}4^1}{1!} + \frac{e^{-4}4^2}{2!} = \mathbf{0.2381}$$

The probability is 0.2381. This probability indicates that the warehouse has a 23.81% chance of receiving fewer than 3 defective products on a given day. This information can be used to plan inventory levels and customer service responses based on the expected number of returns.

## 2. Asymptotic Relationship Between Binomial and Poisson Distributions (18 Points)

A small factory produces electronic components where each component has a probability  $p = 0.01$  of being defective. The factory inspects batches of 100 components each day.

### (a) Proof of concept

- i. (2 points) Let  $X$  be the number of defective components in a batch of  $n = 100$  components, where each component has a defect probability  $p = 0.01$ . Show that as  $n$  becomes large and  $p$  becomes small such that  $\lambda = np$  remains constant, the binomial distribution  $X \sim B(n, p)$  approaches a Poisson distribution with parameter  $\lambda = np$ .  
for a binomial distribution  $X \sim B(n, p)$ , the PMF is given by:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \text{ where } n = 100 \text{ and } p = 0.01$$

$$P(X = x) = \binom{100}{x} 0.01^x (1 - 0.01)^{100-x}$$

$$P(X = x) = \frac{100!}{x!(100-x)!} 0.01^x 0.99^{100-x}$$

$$P(X = x) = \frac{100!}{x!(100-x)!} 0.01^x 0.99^{100} 0.99^{-x}$$

let  $x$  be zero:

$$P(X = 0) = \frac{100!}{0!(100-0)!} 0.01^0 0.99^{100} 0.99^{-0} = \frac{100!}{100!} 0.99^{100} = \mathbf{0.3660}$$

for poisson distribution  $X \sim \text{Poisson}(\lambda)$ , the PMF is given by:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \text{ where } \lambda = np = 100 \times 0.01 = 1$$

let  $x$  be zero:

$$P(X = 0) = \frac{e^{-1} 1^0}{0!} = e^{-1} = \mathbf{0.3679}$$

$$\mathbf{0.3660 \approx 0.3679}$$

for large  $n$  and small  $p$  such that  $\lambda = np$  remains constant, the binomial distribution  $X \sim B(n, p)$  approaches a Poisson distribution with parameter  $\lambda = np$ .

- ii. (2 points) Prove that:

$$\lim_{n \rightarrow \infty, p \rightarrow 0} P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where  $\lambda = np$  is fixed.

for a binomial distribution  $X \sim B(n, p)$ , the PMF is given by:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ where } n = 100 \text{ and } p = 0.01$$

$$P(X = k) = \frac{100!}{k!(100-k)!} 0.01^k 0.99^{100-k}$$

$$P(X = k) = \frac{100!}{k!(100-k)!} 0.01^k 0.99^{100} 0.99^{-k}$$

let  $k$  be zero:

$$P(X = 0) = \frac{100!}{0!(100-0)!} 0.01^0 0.99^{100} = \frac{100!}{100!} 0.99^{100} = 0.3660$$

for poisson distribution  $X \sim \text{Poisson}(\lambda)$ , the PMF is given by:

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \text{ where } \lambda = np = 100 \times 0.01 = 1$$

let  $k$  be zero:

$$P(X = 0) = \frac{e^{-1} 1^0}{0!} = e^{-1} = 0.3679$$

$$\lim_{n \rightarrow \infty, p \rightarrow 0} P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{1^k e^{-1}}{k!} = \frac{e^{-1}}{k!} = 0.3679$$

- iii. (2 points) Provide a detailed explanation of why the Poisson distribution is a good approximation for the binomial distribution in this context.

As the number of trials becomes large and the probability of success becomes small, the binomial distribution approaches the Poisson distribution with a fixed rate parameter. Therefore, the Poisson distribution is a good approximation for the binomial distribution in this context because it is difficult to calculate the binomial distribution for a large number of trials and a small probability of success. The Poisson distribution provides a simpler and more efficient way to calculate probabilities and make decisions based on the expected number of successes.

### (b) Poisson Approximation for Defect Probability (4 points)

- i. (1 point) Using the Poisson approximation, calculate the probability of finding exactly 3 defective components in a batch of 100.

$$P(X = 3) = \frac{e^{-1} 1^3}{3!} = \mathbf{0.0613}$$

- ii. (1 point) Compute the probability of finding at most 5 defective components using the Poisson distribution

$$P(X \leq 5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$$

$$P(X \leq 5) = \frac{e^{-1}1^0}{0!} + \frac{e^{-1}1^1}{1!} + \frac{e^{-1}1^2}{2!} + \frac{e^{-1}1^3}{3!} + \frac{e^{-1}1^4}{4!} + \frac{e^{-1}1^5}{5!}$$

$$0.3679 + 0.3679 + 0.1839 + 0.0613 + 0.0153 + 0.0031 = \mathbf{0.9994}$$

- iii. (2 points) Explain the implications of these probabilities for quality control in the factory, and discuss how this approximation simplifies the analysis

The probability of finding exactly 3 defective components in a batch of 100 is **0.0613 - 6.13%**, while the probability of finding at most 5 defective components is **0.9994 - 99.4%**. These probabilities provide insights into the expected number of defects in a batch and can be used to set quality control targets.

The Poisson approximation simplifies the analysis as it is more easier to calculate probabilities using the Poisson distribution than the binomial distribution for a large number of trials and a small probability of success

### 3. Poisson Distribution in Real-World Applications (35 points)

- (a) Airtel Rwanda provides communication services to high school students in Rwanda at a subsidized cost. To maximize their market returns, their Data Analyst collects data and reports that students make an average of 5 calls per minute with a 7:3 ratio for intra-network to inter-network calls. Following her recommendation to reduce call tariffs, her subsequent monthly report reveals that the average number of calls increases by 2 calls per minute. Given the following tariff information:

**Before reduction:**

- Intra-network calls: 70 RWF per minute
- Inter-network calls: 90 RWF per minute

**After reduction:**

- Intra-network calls: 60 RWF per minute
- Inter-network calls: 80 RWF per minute

Assume the number of calls follows a Poisson distribution and that a cap of 10 calls are made per minute.

- i. (1 point) What is the probability that exactly 7 calls are made in a given minute before the tariff reduction? Use the Poisson distribution with parameter  $\lambda = 5$ .

$$P(X = 7) = \frac{e^{-5}5^7}{7!} = \mathbf{0.104444863}$$

- ii. What is the probability that fewer than 3 calls are made in a given minute after the tariff reduction? Use the Poisson distribution with parameter  $\lambda = 7$ .

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{e^{-7}7^0}{0!} + \frac{e^{-7}7^1}{1!} + \frac{e^{-7}7^2}{2!}$$

$$0.000911882 + 0.006383174 + 0.022341109 = \mathbf{0.029636165}$$

- iii. (1 point) Before the tariff reduction, what is the expected number of intra-network and internetwork calls per minute based on the 7:3 ratio?

$$\text{Intra-network calls} = 5 \times \frac{7}{10} = \mathbf{3.5}$$

$$\text{Inter-network calls} = 5 \times \frac{3}{10} = \mathbf{1.5}$$

- iv. (2 points) Calculate the expected revenue per minute before and after the tariff reduction based on the expected number of calls and the given tariffs.

**Before reduction:**

$$\text{Expected revenue per minute} = 3.5 \times 70 + 1.5 \times 90 = \mathbf{380}$$

**After reduction:**

$$\text{Expected intra-network calls} = 7 \times \frac{7}{10} = \mathbf{4.9}$$

$$\text{Expected inter-network calls} = 7 \times \frac{3}{10} = \mathbf{2.1}$$

$$\text{Expected revenue per minute} = 4.9 \times 60 + 2.1 \times 80 = \mathbf{462}$$

- v. (8 points) Simulate the probability that the revenue in a given minute exceeds 500 Rwf before and after the tariff reduction for 100,000 repetitions. (Hint: For each case, calculate the Poisson parameter for total calls  $\lambda = 5$  before and  $\lambda = 7$  after the reduction, simulate the number of calls per minute using the Poisson distribution, and compute the probability that the revenue exceeds 500 Rwf.)

#### 4. Expectation, Variance, and Moments of Discrete Random Variables (13 points)

- (a) A programming problem set has three tasks to solve for an interview of interns. For each task solved, a dollar is awarded to the prospective intern. Let  $X$  represent the amount of money earned by an intern, which depends on how many tasks they solve. The distribution of solving the problems is such that it is twice as likely to solve one task as it is to solve no task. It is four times and three times as likely to solve two tasks and three tasks, respectively, as it is to solve no task. If a candidate does not solve any task in 2 out of 5 attempts, answer the following:  
the probability distribution is:

$$P(X = 0) = P$$

$$P(X = 1) = 2p$$

$$P(X = 2) = 4p$$

$$P(X = 3) = 3p$$

- i. (2 points) What is the average amount of money a candidate can be rewarded with? (Hint: Use the probabilities for solving 0, 1, 2, and 3 tasks to compute the expectation)

$$E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) + 2 \times P(X = 2) + 3 \times P(X = 3)$$

But we know the probabilities are in terms of  $p$ , so we need to find  $p$  first:

$$P(X = 0) = \frac{2}{5} = p = 0.4$$

$$P(X = 1) = 2 \times 0.4 = 0.8$$

$$P(X = 2) = 4 \times 0.4 = 1.6$$

$$P(X = 3) = 3 \times 0.4 = 1.2$$

we need to normalize the probabilities so that they sum to 1 we divide by the sum of the probabilities:

$$P(X = 0) = \frac{2}{5} = 0.4/4 = 0.1$$

$$P(X = 1) = 0.8/4 = 0.2$$

$$P(X = 2) = 1.6/4 = 0.4$$

$$P(X = 3) = 1.2/4 = 0.3$$

$$E(X) = 0 \times 0.1 + 1 \times 0.2 + 2 \times 0.4 + 3 \times 0.3 = \mathbf{2}$$

- ii. (3 points) How would the amount of money earned vary with respect to the number of tasks solved? (Hint: Calculate the variance of  $X$  to assess the variability in rewards)

$$Var(X) = E(X^2) - E(X)^2$$

$$E(X^2) = 0^2 \times 0.1 + 1^2 \times 0.2 + 2^2 \times 0.4 + 3^2 \times 0.3 = 0 + 0.2 + 1.6 + 2.7 = \mathbf{4.5}$$

$$Var(X) = 4.5 - 2^2 = 4.5 - 4 = \mathbf{0.5}$$

- (b) In a series of independent trials with probability 0.3 of success, we want to calculate the probability that the first success occurs on the 4th trial. Let  $X_1$  be a random variable representing the number of trials until the first success.

- i. What is the probability that the first success occurs on the 4th trial? (Hint: Use the PMF of the geometric distribution.)

$$P(X_1 = k) = (1 - p)^{k-1}p$$

$$P(X_1 = 4) = (1 - 0.3)^3 \times 0.3 = \mathbf{0.1029}$$

- ii. During a fishing competition, each angler has a 10% chance of catching a fish with each cast. Assuming independence in each cast, use  $X_r$  to represent the number of casts needed to secure  $r$  successful catches.

- A. (1 point) Calculate  $P(X_1 = x_1)$ , the probability of securing the first catch on the  $x_1$ -th cast.

$$P(X_1 = x_1) = (1 - 0.1)^{x_1-1} \times 0.1$$

$$P(X_1 = x_1) = 0.9^{x_1-1} \times 0.1$$

- B. (2 points) Determine the expected number of casts to catch the first fish,  $E[X_1]$ .

$$E[X_1] = \frac{1}{p} = \frac{1}{0.1} = \mathbf{10}$$

- C. (2 points) Find  $P(X_4 = x_4)$ , the probability that it takes exactly  $X_4$  casts to catch four fish.

$$P(X_4 = x_4) = \binom{x_4 - 1}{3} \times P^4 \times (1 - P)^{x_4 - 4}$$

This is a negative binomial distribution.

$$P(X_4 = x_4) = \binom{x_4 - 1}{3} \times 0.1^4 \times 0.9^{x_4 - 4}$$

$$P(X_4 = x_4) = \frac{(x_4 - 1)!}{3!(x_4 - 4)!} \times 0.1^4 \times 0.9^{x_4 - 4}$$

$$P(X_4 = x_4) = \frac{(x_4 - 1)(x_4 - 2)(x_4 - 3)}{3!} \times 0.1^4 \times 0.9^{x_4 - 4}$$

$$P(X_4 = x_4) = \frac{(x_4 - 1)(x_4 - 2)(x_4 - 3)}{6} \times 0.1^4 \times 0.9^{x_4 - 4}$$

$$P(X_4 = x_4) = \frac{(x_4 - 1)(x_4 - 2)(x_4 - 3)}{6} \times 0.0001 \times 0.9^{x_4 - 4}$$

$$P(X_4 = x_4) = \frac{(x_4 - 1)(x_4 - 2)(x_4 - 3)}{60000} \times 0.9^{x_4 - 4}$$

D. (2 points) Find  $E[X_6]$ , the expected number of casts to catch 6 fish. (Hint: Use the fact that the sum of independent geometric random variables follows a negative binomial distribution.)

$$E[X_6] = \frac{r}{p} = \frac{6}{0.1} = \mathbf{60}$$

## 5. Expectation, Variance, and Moments of Continuous Random Variables (36 points)

### (a) Exponential Distribution and Reliability (12 points)

A machine part fails according to an exponential distribution with a mean time between failures of 10 hours. Let  $X$  represent the time (in hours) until the next failure

- i. (2 points): Write the PDF for the exponential distribution and calculate the expectation  $E[X]$ .

$$f_X(x) = \lambda e^{-\lambda x} \text{ for } x \geq 0$$

our rate parameter is:

$$\lambda = \frac{1}{\mu} = \frac{1}{10} = 0.1$$

$$f_X(x) = \frac{1}{10} e^{-\frac{x}{10}} \text{ for } x \geq 0$$

$$E[X] = \frac{1}{\lambda} = \frac{1}{0.1} = \mathbf{10}$$

- ii. (3 points): Derive the variance of  $X$ . Interpret what this variance means in terms of the variability of failure times.

$$\text{Var}(X) = \frac{1}{\lambda^2} = \frac{1}{0.1^2} = \mathbf{100}$$

The variance of  $X$  is 100. This means that the failure times are highly variable, with a large spread around the mean time between failures.

- iii. (7 points): Suppose the machine part is replaced after each failure, and you track 1000 failure times. Simulate the total downtime over these 1000 failures, assuming each failure follows the exponential distribution with a mean of 10 hours. Plot the histogram of failure times and discuss the practical implications for managing the machine's operational efficiency.

### (b) Normal Distribution and Product Quality (12 points)

In a factory, the weight of a certain product follows a normal distribution with a mean of 500 grams and a standard deviation of 10 grams.

- i. (2 points) Write the PDF of the normal distribution for this product weight. Identify the expectation  $E[X]$  and variance  $\text{Var}(X)$ .

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}10} e^{-\frac{(x-500)^2}{2 \times 10^2}}$$

$$E[X] = \mu = \mathbf{500}$$

$$\text{Var}(X) = \sigma^2 = 10^2 = \mathbf{100}$$

- ii. (3 points) Calculate the probability that a randomly selected product weighs between 490 and 510 grams. How would this result influence quality control decisions?

$$P(490 < X < 510) = P(X < 510) - P(X < 490)$$

in standard normal form:

$$P(X < 510) = P\left(\frac{510 - 500}{10}\right) = P(Z < 1)$$

$$P(X < 490) = P\left(\frac{490 - 500}{10}\right) = P(Z < -1)$$

$$P(490 < X < 510) = P(-1 < Z < 1) = \mathbf{0.6826}$$

The probability is 0.6826. This result indicates that there is a 68.26% chance that a randomly selected product weighs between 490 and 510 grams. This information can be used to set quality control limits and identify products that fall outside the acceptable weight range.

- iii. (4 points) If the factory wants to ensure that 95% of the products weigh between 495 and 505 grams, what should the standard deviation be? What does this say about how much the manufacturing process needs to be improved?

$$P(495 < X < 505) = 0.95$$

$$P\left(\frac{495 - 500}{\sigma}\right) < P(Z) < P\left(\frac{505 - 500}{\sigma}\right) = 0.95$$

$$P(-0.5 < Z < 0.5) = 0.95$$

$$P(Z < 0.5) - P(Z < -0.5) = 0.95$$

$$0.6915 - 0.3085 = 0.95$$

$$\sigma = \frac{5}{0.6915} = \mathbf{7.23}$$

The standard deviation should be 7.23 grams to ensure that 95% of the products weigh between 495 and 505 grams. This indicates that the manufacturing process needs to be improved to reduce the variability in product weights.

- iv. (3 points) Simulate 10,000 product weights using the normal distribution with the given parameters. Plot the distribution and calculate the percentage of products that fall within the range of 495 to 505 grams. Compare this with the theoretical value and discuss any discrepancies.



6. **Lognormal Distribution and Stock Returns (12 points)**

The daily returns of a stock are modeled using a lognormal distribution with parameters  $\mu = 0.001$  and  $\sigma = 0.02$ .

- (a) (2 points) Write the PDF for the lognormal distribution and calculate the expectation and variance of the stock returns.

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$

$$f_X(x) = \frac{1}{x \times 0.02 \times \sqrt{2\pi}} e^{-\frac{(\ln(x)-0.001)^2}{2 \times 0.02^2}}$$

$$E[X] = e^{\mu + \frac{\sigma^2}{2}} = e^{0.001 + \frac{0.02^2}{2}} = \mathbf{1.001}$$

$$Var(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1) = e^{2 \times 0.001 + 0.02^2}(e^{0.02^2} - 1) = \mathbf{0.0004}$$

- (b) (3 points) Calculate the probability that the stock returns more than 5% in a day. Discuss the implications of this for investors considering the stock as a high-risk, high-reward asset.

$$P(X > 1.05) = 1 - P(X < 1.05)$$

$$Z = \frac{\ln(1.05) - 0.001}{0.02} = 2.236$$

$$P(X > 1.05) = 1 - P(Z < 2.236) = 1 - 0.9877 = \mathbf{0.0123}$$

- (c) (4 points) Simulate 1,000 days of stock returns using the lognormal distribution. Plot the histogram of daily returns and calculate the proportion of days where the return is positive. How does this relate to typical stock market behavior?
- (d) (3 points) Calculate the 95th percentile of the daily returns and explain its significance in a financial risk management context

$$P(X < x) = 0.95$$

$$Z = \frac{\ln(x) - 0.001}{0.02} = 1.645$$

$$x = e^{0.001 + 1.645 \times 0.02} = \mathbf{1.034}$$

## 7. Joint Probability Distributions and Covariance (32 points)

- (a) In a certain suburb, each household reported the number of cars and the number of television sets that they own. Let  $X$  represent the number of cars, and  $Y$  represent the number of television sets owned by a randomly selected household. The table below gives the joint probability distribution for  $X$  and  $Y$  :

	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$
$X = 1$	0.1	0	0.1	0
$X = 2$	0.3	0	0.1	0.2
$X = 3$	0	0.2	0	0

- i. (2 points) Find the probability that a randomly selected household owns at most two cars and at most two television sets. That is, compute:

$$P(X \leq 2 \text{ and } Y \leq 2)$$

Provide an explanation of how the probabilities from the table contribute to this total probability.

$$P(X \leq 2 \text{ and } Y \leq 2) = P(X = 1, Y = 1) + P(X = 1, Y = 2) + P(X = 2, Y = 1) + P(X = 2, Y = 2)$$

$$P(X \leq 2 \text{ and } Y \leq 2) = 0.1 + 0 + 0.3 + 0 = \mathbf{0.4}$$

The probabilities from the table contribute to this total probability by providing the likelihood of each combination of cars and television sets owned by a household.

- ii. (3 points) Using the joint probability distribution, determine the marginal distribution for the number of cars  $X$  owned by a household. In other words, compute the probabilities  $P(X = x)$  for each possible value of  $X$ . Specifically:

$$P(X = 1), P(X = 2), P(X = 3)$$

Summarize these values in a clear format and provide an interpretation of what the marginal distribution of cars tells us about car ownership in the suburb.

$$P(X = 1) = 0.1 + 0 + 0.1 = \mathbf{0.2}$$

$$P(X = 2) = 0.3 + 0 + 0.1 + 0.2 = \mathbf{0.6}$$

$$P(X = 3) = 0 + 0.2 + 0 + 0 = \mathbf{0.2}$$

**Marginal Distribution of cars:**

$X$	$P(X)$
1	0.2
2	0.6
3	0.2

The marginal distribution of cars tells us about car ownership in the suburb by providing the probability of households owning a certain number of cars, regardless of the number of television sets they own.

- iii. (3 points) Similarly, determine the marginal distribution for the number of television sets  $Y$  owned by a household. Compute the probabilities  $P(Y = y)$  for each possible value of  $Y$ . Specifically:

$$P(Y = 1), P(Y = 2), P(Y = 3), P(Y = 4)$$

Present the results in a clear format and discuss what the marginal distribution reveals about television set ownership in the suburb.

$$P(Y = 1) = 0.1 + 0.3 + 0 = \mathbf{0.4}$$

$$P(Y = 2) = 0 + 0 + 0.2 = \mathbf{0.2}$$

$$P(Y = 3) = 0.1 + 0.1 + 0 = \mathbf{0.2}$$

$$P(Y = 4) = 0 + 0.2 + 0 = \mathbf{0.2}$$

The marginal distribution of television sets tells us about television ownership in the suburb by providing the probability of households owning a certain number of televisions, regardless of the number of cars they own.

**Marginal Distribution of televisions:**

$Y$	$P(Y)$
1	0.4
2	0.2
3	0.2
4	0.2

- iv. (3 points) Are the number of cars  $X$  and the number of television sets  $Y$  owned by a household independent? Recall that  $X$  and  $Y$  are independent if and only if:

$$P(X = x, Y = y) = P(X = x)P(Y = y) \text{ for all } x, y$$

For this part, verify whether this condition holds for any pair of values of  $X$  and  $Y$ . Show your calculations and state whether  $X$  and  $Y$  are independent or not.

take  $X = 1, Y = 1$ :

$$P(X = 1, Y = 1) = 0.1$$

$$P(X = 1)P(Y = 1) = 0.2 \times 0.4 = 0.08 \text{ (marginal probabilities)}$$

$$P(X = 1, Y = 1) \neq P(X = 1)P(Y = 1)$$

The number of cars  $X$  and the number of television sets  $Y$  owned by a household are not independent.

- v. (2 points) Find the conditional probability that a randomly selected household owns exactly two television sets given that they own exactly two cars. That is, calculate:

$$P(Y = 2|X = 2)$$

Explain the result and provide insight into how owning two cars may affect the likelihood of owning two television sets.

$$P(Y = 2|X = 2) = \frac{P(X = 2, Y = 2)}{P(X = 2)}$$

$$P(X = 2, Y = 2) = 0$$

$$P(X = 2) = 0.6$$

$$P(Y = 2|X = 2) = \frac{0}{0.6} = 0$$

The result indicates that the conditional probability of owning exactly two television sets given that a household owns exactly two cars is 0. This suggests that owning two cars does not affect the likelihood of owning two television sets, as there are no households in the sample that own two cars and two television sets.

- vi. (3 points) Using the marginal distributions, compute the expected number of cars  $E[X]$  and the expected number of television sets  $E[Y]$  a randomly selected household owns. Use the formulas for the expected value of a discrete random variable:

$$E[X] = \sum_x xP(X = x)$$

$$E[X] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2$$

$$E[X] = 2$$

$$E[Y] = \sum_y yP(Y = y)$$

$$E[Y] = 1 \times 0.4 + 2 \times 0.2 + 3 \times 0.2 + 4 \times 0.2 = 2.2$$

$$E[Y] = 2.2$$

$$E[X] = 2.0, E[Y] = 2.2$$

The expected number of cars owned by a household is 2, while the expected number of television sets owned is 2.2, indicating that households in the suburb are more likely to own more television sets than cars.

- (b) The daily returns of two correlated stocks, X and Y, follow a joint lognormal distribution with the following parameters:

$$\mu_X = 0.001, \mu_Y = 0.002, \sigma_X = 0.02, \sigma_Y = 0.03, \rho_{X,Y} = 0.8$$

- i. (2 points) Write the joint PDF for the lognormal distribution of X and Y.

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \exp\left(-\frac{1}{2(1-\rho_{X,Y}^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho_{X,Y}\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right)$$

$$f_{X,Y}(x,y) = \frac{1}{2\pi \times 0.02 \times 0.03 \times \sqrt{1-0.8^2}} \exp\left(-\frac{1}{2(1-0.8^2)} \left[\frac{(x-0.001)^2}{0.02^2} - 2 \times 0.8 \times \frac{(x-0.001)(y-0.002)}{0.02 \times 0.03} + \frac{(y-0.002)^2}{0.03^2}\right]\right)$$

- ii. (3 points) Calculate the marginal distributions of X and Y.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi \times 0.02 \times 0.03 \times \sqrt{1-0.8^2}} \exp\left(-\frac{1}{2(1-0.8^2)} \left[\frac{(x-0.001)^2}{0.02^2} - 2 \times 0.8 \times \frac{(x-0.001)(y-0.002)}{0.02 \times 0.03} + \frac{(y-0.002)^2}{0.03^2}\right]\right) dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

$$f_Y(y) = \int_{-\infty}^{\infty} \frac{1}{2\pi \times 0.02 \times 0.03 \times \sqrt{1-0.8^2}} \exp\left(-\frac{1}{2(1-0.8^2)} \left[\frac{(x-0.001)^2}{0.02^2} - 2 \times 0.8 \times \frac{(x-0.001)(y-0.002)}{0.02 \times 0.03} + \frac{(y-0.002)^2}{0.03^2}\right]\right) dx$$

- iii. (4 points) Simulate 1,000 days of returns for both stocks using the joint lognormal distribution. Plot the scatter plot and calculate the empirical correlation.
- iv. (2 points) Using the simulated data, calculate the percentage of days where both stocks have positive returns. Compare this to the theoretical correlation.
- (c) (5 points) Let X be the size of a surgical claim and let Y denote the size of the associate hospital claim. A risk analyst uses a model in which  $E(X) = 5$ ,  $E(X^2) = 27.4$ ,  $E(Y) = 7$ ,  $E(Y^2) = 51.4$  and  $\text{Var}(X+Y) = 8$ . Let  $C1 = X+Y$  denote the size of the combined claims before the application of a 20% surcharge on the hospital claim and  $C2$  denote the size of the combined claim after the application of the surcharge. Calculate  $\text{Cov}(C1, C2)$ .

$$\text{Cov}(C1, C2) = E[C1C2] - E[C1]E[C2]$$

$$\text{Cov}(C1, C2) = E[(X+Y)(X+1.2Y)] - E[X+Y]E[X+1.2Y]$$

$$\text{Cov}(C1, C2) = E[X^2 + 1.2XY + XY + 1.44Y^2] - (E[X] + E[Y])(E[X] + 1.2E[Y])$$

$$Cov(C1, C2) = E[X^2] + 2.2E[XY] + 1.44E[Y^2] - (E[X] + E[Y])(E[X] + 1.2E[Y])$$

$$Cov(C1, C2) = 27.4 + 2.2 \times Cov(X, Y) + 1.44 \times 51.4 - (5 + 7)(5 + 1.2 \times 7)$$

$$Cov(C1, C2) = 27.4 + 2.2 \times Cov(X, Y) + 1.44 \times 51.4 - 12 \times 12.4$$

$$Cov(C1, C2) = 27.4 + 2.2 \times Cov(X, Y) + 74.016 - 148.8$$

$$Cov(C1, C2) = 27.4 + 2.2 \times Cov(X, Y) - 74.784$$

$$Cov(C1, C2) = 2.2 \times Cov(X, Y) - 47.384$$

$$8 = Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

$$8 = 27.4 + 51.4 + 2Cov(X, Y)$$

$$8 = 78.8 + 2Cov(X, Y)$$

$$2Cov(X, Y) = -70.8$$

$$Cov(X, Y) = \mathbf{-35.4}$$