Probability and Random Variables with Python Simulations

Applied Stochastic Processes: HW 1

Due Date: Spetember 14, 2024

Objectives

This assignment is designed to help students develop a deep understanding of probability and random variables. Students will enhance their problem-solving skills and gain practical experience through Python-based simulations.

Policy

- You can discuss HW problems with other students, but the work you submit must be written in your own words and not copied from anywhere else. This includes codes.
- However, do write down (at the top of the first page of your HW solutions) the names of all the people with whom you discussed this HW assignment.
- You may decide to write out your solution with pen and paper and use a scanning app to turn in a PDF submission or may choose to type out your solution with LATEX. We strongly encourage the latter.

Expected Topics Covered

1. Matrix Operations and Applications
2. Probability Basics
3. Combinatorics

Q1: Probability Density Functions and Real-Life Applications (10 Points)

Reading: A random variable is a function that maps outcomes from an experiment's sample space to real numbers $\mathbb{R}$, assigning numerical values to quantify uncertainty and variability. Its range consists of all possible values, while the sample space encompasses all potential outcomes.
Random variables are classified as discrete or continuous. Discrete random variables take on a countable number of distinct values, often from finite processes like rolling a die or counting successes. Continuous random variables can assume any value within a specific interval, such as time, height, or temperature. The probability distribution describes the likelihood of outcomes. Discrete random variables use a probability mass function (PMF) to assign probabilities, while continuous random variables utilize a probability density function (PDF), where probabilities for intervals are derived through integration. Understanding random variables is crucial for statistical analysis and inferential statistics, allowing researchers to model real-world phenomena, draw conclusions, and evaluate the reliability of their findings amid uncertainties.

1. Consider the PDF of a continuous random variable Y that models the duration (in hours) of a sudden electrical outage in a city, described by the function:

$$g(y) = \lambda^2 y e^{-\lambda y}, \quad \text{for } y > 0.$$

   (a) **(2 points):** Prove that $g(y)$ is a valid probability density function by verifying the normalization condition:

$$\int_0^{\infty} \lambda^2 y e^{-\lambda y}\, dy = 1.$$

   (b) **(4 points):** Show whether $g(y)$ is concave or convex over its domain by analyzing its second derivative. Discuss the practical implications of this concavity in predicting the likelihood of prolonged outages.

   (c) **(4 points):** Find the mode of the distribution $g(y)$ by solving for the critical points of the function and determining which point corresponds to the maximum or minimum value. Using $\lambda = \begin{bmatrix} 0.5 & 2 \end{bmatrix}$ relate the mode to real-life scenarios where outages are frequent and brief or infrequent but prolonged.

2. Two manufacturing plants, Plant A and Plant B, produce goods with daily outputs X (in tons) and Y (in tons) respectively. The production efficiency between the two plants can be modeled by the joint probability density function:

$$f(x, y) = e^{-(x+y)} \quad \text{for} \quad x > 0 \quad \text{and} \quad y > 0.$$

   (a) **(2 points):** Show that $f(x, y)$ is a valid joint probability density function by verifying the normalization condition:

$$\iint_{x>0, y>0} e^{-(x+y)}\, dx\, dy = 1.$$

   (b) **(4 points):** Find the probability that both plants produce more than 1 ton of goods in a day by calculating:

$$P(X > 1, Y > 1) = \iint_{x>1, y>1} e^{-(x+y)}\, dx\, dy.$$

   Discuss the practical interpretation of this result for plant managers who are aiming to maintain high production levels.

(c) **(4 points):** Calculate the probability that Plant A produces more than 2 tons and Plant B produces less than 2 tons:

$$P(X > 2, Y < 2) = \iint_{x>2, 0<y>2} e^{-(x+y)} \, dx \, dy.$$

Explain how this information might be useful for balancing production targets between the two plants.

Q2: Eigenvalues and Stability Analysis in Machine Learning Context (8 Points)

Reading: Eigenvalues and vectors are fundamental concepts in linear algebra, playing a crucial role in various mathematical, scientific, and engineering applications. An eigenvalue is a scalar that indicates how much a corresponding eigenvector is stretched or compressed during a linear transformation represented by a matrix. More specifically, if A is a square matrix, an eigenvalue λ satisfies the equation $Av = \lambda v$, where v is the eigenvector associated with λ.
The determination of eigenvalues and eigenvectors is essential for solving systems of linear equations, performing matrix diagonalization, and analyzing stability in dynamical systems. They help in understanding the characteristics of linear transformations, such as rotations, reflections, and scaling. In computational applications, eigenvalue problems appear in methods like Principal Component Analysis (PCA), which reduces data dimensionality while retaining essential information.
Finding eigenvalues often involves computing the characteristic polynomial, defined as $det(A - \lambda I) = 0$, where I is the identity matrix. The roots of this polynomial yield the eigenvalues, while substituting these back into the equation enables the calculation of the corresponding eigenvectors. Thus, the study of eigenvalues and eigenvectors is central to grasping the behavior of complex systems and simplifying various mathematical tasks.

You are analyzing a system modeled by a 3x3 state transition matrix, which represents the interaction of three interconnected features in a neural network, given by:

$$B = \begin{pmatrix} 4 & 2 & 1 \\ 1 & 3 & 2 \\ 0 & 1 & 2 \end{pmatrix}.$$

(a) **(5 points):** Determine if the system described by matrix B is stable by finding its eigenvalues using the characteristic equation $\det(B - \lambda I) = 0$. Discuss whether the eigenvalues meet the stability criterion for a system where stability requires all eigenvalues to have negative real parts.

(b) **(3 points):** Explore how modifications to the off-diagonal elements of the matrix (e.g., increasing the interaction terms) impact the overall stability of the system. Discuss the potential implications of these changes in the context of training machine learning models, such as avoiding exploding gradients.

Q3: Markov Chains in Employee Performance Evaluation (10 Points)

Reading: Markov chains describe a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. This characteristic is known as the Markov property, which simplifies the analysis of complex stochastic processes. In a Markov chain, the system transitions between various states, and these transitions are governed by a set of probabilities that can

be represented in a transition matrix. Each entry in this matrix indicates the likelihood of moving from one state to another, thus providing a complete statistical description of the process.
Markov chains can be categorized into two main types: discrete-time and continuous-time. Discrete-time Markov chains involve state transitions at fixed time intervals, while continuous-time Markov chains allow for transitions at any moment in time. Both types share the fundamental principle that the future state depends solely on the current state, not on the sequence of events that preceded it.
Applications of Markov chains span various fields, including economics, genetics, game theory, and artificial intelligence. For instance, in natural language processing, Markov chains can be utilized to model and predict sequences of words in a given text. By training on a dataset, the chain can learn the transition probabilities between different words or phrases, enabling it to generate coherent sentences based on its training.
Moreover, Markov chains can also be used for decision-making processes, where the objective is to maximize an expected reward over time. In these scenarios, reinforcement learning often incorporates Markov decision processes, an extension of Markov chains that take into account both the states of the system and the actions taken by an agent.
Understanding the principles of Markov chains also opens up avenues for studying more complex models, such as Hidden Markov Models (HMMs), where the states are not directly observable. HMMs are useful in various applications, including speech recognition, bioinformatics, and financial modeling, allowing researchers and practitioners to infer hidden processes based on observable data.

Consider a Markov chain that models the daily performance of an employee categorized into three states: **High Performance**, **Moderate Performance** and **Low Performance**. The transition matrix is:

$$Q = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.5 & 0.3 \end{pmatrix}.$$

where the states represent the probabilities of transitioning from one performance level to another on the next day.

(a) **(6 points):** Calculate the state distribution after one transition given an initial state distribution of $\pi_0 = (0.5, 0.3, 0.2)$. Show your calculation of $\pi_1 = \pi_0 Q$.

(b) **(4 points):** Determine the steady-state distribution π such that $\pi Q = \pi$. Analyze whether the system is likely to converge to a state where most employees perform at a "Moderate Performance" level. Discuss how this steady-state distribution can influence decisions about employee training and incentives in a corporate setting.

Q4: Matrix/Vector Manipulation in Optimization (4 Points)

Given a matrix M that represents a quadratic form $x^T Mx$, discuss under what conditions the function is concave or convex. Connect this to how concavity affects the optimization landscape in machine learning models, such as in support vector machines or neural network loss surfaces.

Q5: Basic Probability (20 Points)

Reading: Sampling refers to the process of selecting a subset (a sample) from a larger set (the population) to make inferences about the population. The goal of sampling is to gather information about the entire population without having to survey or measure every individual in that population.
A simulation is a technique for modeling a real-world system or process through the use of a computer program. It involves creating a virtual model of a system, running experiments on this model, and analyzing the outcomes to gain insights into the behavior of the system under various conditions.

In this question, we want to conduct an experiment of sampling from game cards to convey the concept of probability. The probability of an event is a measure of the likelihood that the event will occur, expressed as a number between 0 and 1. Thus, it is the number of times the event is observed relative to the number of experiments conducted. A probability of 0 indicates that the event is impossible, while a probability of 1 signifies that the event is certain to happen.

Define the sample space for drawing a card from a standard deck of 52 cards.

(a) **(4 points)** Calculate the probability of drawing a heart.

(b) **(4 points)** Calculate the probability of drawing a face card (Jack, Queen, King).

(c) **(8 points)** Using the standard deck of cards as the population, simulate the process of drawing 10 cards (without replacement) and calculate the above probabilities based on your sample. Compare your answers with the computed probabilities.

(d) **(4 points)** Given the sample size range $[1, N]$, repeat this process and plot the probabilities of the computed events above as well as the calculated in (b) and (c). What are your observations?

Steps to Carry Out the Simulation

1. **Define the Population:**

   The standard deck of cards has 52 cards: 13 ranks (Ace to King) in each of 4 suits (hearts, diamonds, clubs, spades).

2. **Simulate Drawing Cards:**

   - Draw a sample of 10 cards from the deck without replacement.
   - This means once a card is drawn, it is not put back into the deck for subsequent draws.

3. **Calculate Probabilities:**

   - Based on the drawn sample, calculate the probability of certain events (e.g., drawing a specific rank or suit).
   - Compare these sample-based probabilities with the theoretical probabilities computed from the entire deck.

4. **Repeat the Process:**

   Perform the drawing and probability calculations for sample sizes ranging from 1 to 52.

5. **Plot and Analyze:**

   - Plot the probabilities obtained from the simulations and compare them with the theoretical probabilities.
   - Observe and report how the simulated probabilities behave.

Q6: Conditional Probability (5 points)

Reading: Conditional probability is a fundamental concept in probability theory that quantifies the likelihood of an event occurring given that another event has already taken place. Formally, the conditional probability of an event A given that event B has occurred is denoted by $P(A|B)$. This notation signifies that we are interested in event A within the context of event B, which can significantly alter our perception of the likelihood of A.
To compute conditional probability, we use the formula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

provided that $P(B) > 0$. Here, $P(A \cap B)$ represents the probability that both events A and B occur simultaneously, while $P(B)$ is the probability of event B alone. This relationship underscores the importance of the intersection in determining how these events influence each other.
Conditional probability is particularly vital in various fields such as statistics, finance, and machine learning. In statistical analysis, understanding how variables interact conditionally allows researchers to bridge causal inferences and make informed predictions. For instance, in medical research, one may wish to determine the probability of a patient developing a condition based on their exposure to a risk factor, effectively leveraging the concept of conditional probability for better healthcare outcomes.
Moreover, the idea extends to concepts like Bayes' theorem, which interrelates conditional probabilities and provides a powerful framework for updating probabilities as new evidence becomes available. This iterative updating mechanism is central to many algorithms in artificial intelligence, where models continuously refine their predictions based on incoming data.

An insurance company examines its pool of auto insurance customers and gathers the following information:

- All customers insure at least one car.
- 70% of the customers insure more than one car.
- 20% of the customers insure a sports car.
- Of those customers who insure more than one car, 15% insure a sports car.

Calculate the probability that a randomly selected customer insures one car and that car is not a sports car.

Q7: Bayes' Theorem (8 points)

Reading: Bayes' theorem is a fundamental principle in probability theory that describes how to update the probability of a hypothesis based on new evidence. Formulated by the Reverend Thomas Bayes in the 18th century, it provides a way to incorporate prior knowledge with observed data. Mathematically, Bayes' theorem is expressed as:

$$P(H|E) = \frac{(P(E|H) * P(H)}{P(E)}$$

where:

- $P(H|E)$ is the posterior probability, the probability of the hypothesis H given the evidence E.
- $P(E|H)$ is the likelihood, the probability of observing the evidence E if the hypothesis H is true.
- $P(H)$ is the prior probability, the initial belief in the hypothesis before considering the evidence.
- $P(E)$ is the marginal likelihood, the total probability of the evidence under all possible hypotheses.

This theorem is widely used in various fields such as statistics, economics, medicine, and machine learning. It allows for a structured approach to incorporate incoming data, making it particularly useful in predictive modeling and decision-making processes. By providing a clear framework for updating beliefs, Bayes' theorem helps us navigate uncertainty and make informed choices based on the best available information.

Sensitivity measures the probability that the test correctly identifies a positive case given that the condition is present. This means that sensitivity indicates how well the test can accurately detect individuals with the condition.
Specificity, on the other hand, measures the probability that the test correctly identifies a negative case given that the condition is not present. In other words, specificity indicates how well the test can accurately rule out individuals who do not have the condition.

(a) **(4 points)** A medical test for a new strain of a viral respiratory disease has a 98% sensitivity and a 97% specificity. If 0.5% of the population has the disease, calculate the probability that a person has the disease given they tested positive.

(b) **(4 points)** A financial credit scoring model is 95 percent effective in identifying individuals who are likely to default on their loans when they actually will default. However, the model also yields a "false positive" result for 1 percent of individuals who are creditworthy. (That is, if a creditworthy individual is tested, then, with probability 0.01, the model will incorrectly classify them as likely to default.) If 0.5 percent of the population actually defaults on their loans, what is the probability that an individual will default given that the model predicts they are likely to default?

Q8: Independence of Events, Inclusion-Exclusion Principle, Mutual Exclusivity (15 points)

Reading: The notion of independence posits that events occur regardless of one another. Each occurrence stands on its own, unaffected by the preceding or forthcoming actions. This principle is fundamental in probability theory, where the likelihood of multiple events transpiring simultaneously can be calculated by multiplying their individual probabilities ($i.e. P(A \cap B) = P(A) \cdot P(B)$).
When we examine the world around us, we find that this concept applies across a multitude of scenarios – from the flipping of a coin to the roll of a die, each action produces an outcome that doesn't rely on the past or future. The roads of causality may twist and turn, yet the essence of independence reveals that there exists a freedom in every choice and chance that unfolds.
Mutual exclusiveness on the other hand, means that two events cannot occur simultaneously. In other words, if one event happens, the other cannot occur at the same time. For example, one cannot test positive and negative for a disease at the same time. In contrast, non-mutually exclusive events allow for the possibility that more than one event can occur simultaneously. For example, a person can be both a student and an employee at the same time.

Mutual Exclusivity Does Not Imply Independence

If A and B are mutually exclusive, then

$$P(A \cap B) = 0.$$

For independent events,

$$P(A \cap B) = P(A) \cdot P(B).$$

For mutually exclusive events (except when one of them has zero probability), $P(A \cap B) = 0$, and it can only equal $P(A) \cdot P(B)$ if either $P(A) = 0$ or $P(B) = 0$.
Thus, if both $P(A)$ and $P(B)$ are greater than zero, mutually exclusive events cannot be independent, because:

$$P(A) \cdot P(B) > 0 \neq 0.$$

Independence Does Not Imply Mutual Exclusivity

Independent events A and B can occur together. For example, if you flip a coin and roll a die, the event **getting heads** (coin) and "rolling a 4" (die) are independent.
Since they are independent,

$$P(A \cap B) = P(A) \cdot P(B),$$

which is generally not zero if $P(A)$ and $P(B)$ are both non-zero. Therefore, independent events are not mutually exclusive unless one of the events has zero probability.

Inclusion-Exclusion Principle

This principle helps to count the number of elements in the union of overlapping sets by correcting for over-counting:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

For three sets A, B, and C:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

1. Show that if A and B are events with positive probabilities and are independent, then so are:
   (a) **(2 points)** A^c and B
   (b) **(2 points)** A and B^c
   (c) **(2 points)** A^c and B^c
2. **(3 points)** In the ECE class of 2026 consisting of 25 students, 15 take Robotics, 10 take Introduction to Systems Software Engineering, and 5 take both. Calculate the number of students who take either Robotics or Intro. to Systems Software Engineering or both.
3. You are at Nyandungu amusement park with your friends and you want to play a game where you roll a fair six-sided die. There are prizes for rolling specific numbers:
   - If you roll a 1 or a 2, you win a small prize (Event A).
   - If you roll a 3, 4, or 5, you win a medium prize (Event B).
   - If you roll a 6, you win a large prize (Event C).

   (a) **(2 points)** Are Events A and B mutually exclusive? Why or why not?
   (b) **(2 points)** Are Events A and C mutually exclusive? Why or why not?
   (c) **(2 points)** What is the probability of winning either a small or a medium prize?

Q9: Combinatorics (12 points)

Reading: Combinatorics is a branch of mathematics focused on counting, arranging, and analyzing the structure of sets. It plays a critical role in various fields, including computer science, probability, statistics, and optimization. The fundamental concepts in combinatorics include permutations, combinations, the binomial theorem, and the principles of counting.

Basic Principles of Counting

- **Addition Principle:** If there are m ways to do one thing and n ways to do another, and these two actions cannot occur simultaneously, then there are $m + n$ ways to do either.
- **Multiplication Principle:** If there are m ways to do one thing and n ways to do another, and these actions occur in sequence, then there are $m \times n$ ways to do both.

Permutations

Permutations refer to the arrangements of objects where the order matters.

- **Formula:** The number of permutations of n distinct objects is $n!$ (factorial).

$$n! = n \times (n-1) \times \ldots \times 2 \times 1$$

- **Permutations of a Subset:** If selecting r objects from n distinct objects:

$$P(n, r) = \frac{n!}{(n-r)!}$$

Combinations

Combinations refer to the selection of objects where the order does not matter.

- **Formula:** The number of ways to choose r objects from n without regard to order is given by:

$$C(n, r) = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Binomial Theorem

The binomial theorem provides a way to expand expressions of the form $(x+y)^n$:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

This theorem is essential in probability and algebra, linking combinatorial coefficients with algebraic expansions.

Pigeonhole Principle

The pigeonhole principle states that if $n+1$ or more objects are placed into n containers, at least one container must hold more than one object. It is used in proving the existence of certain properties within sets.

Applications of Combinatorics

- **Probability:** Combinatorial methods help calculate probabilities, especially in scenarios involving random selections, arrangements, or distributions.
- **Graph Theory:** Combinatorics underpins the study of graphs, which are essential in network analysis, computer science, and optimization problems.
- **Design Theory and Coding:** Combinatorial designs are used in experiments, cryptography, and error-correcting codes.

1. **(2 points)** Calculate the number of ways to arrange the letters in the word "ALGORITHM" such that the vowels come together.
2. **(4 points)** Every fall semester, elections are held at CMU-Africa to choose members of the student guild and other club representatives. After, the elections were held, the elected members conducted a survey on students problems and it was revealed that tuition funding and housing ranked first. The guild decided on decentralized task forces to address specific issues from a group of 5 females and 7 males, how many different committees consisting of 2 females and 3 males can be formed? What if 2 of the males are feuding and refuse to serve on the committee together?

3. A recent census summary reveals that there are about 1.75 million people in Kigali, of which approximately 55% are females. A survey conducted by a major exchange student in cosmetics reveals an alarming balding rate among men. The survey reports that the average hair count for people in Kigali is $150,000$ hairs. Men aged 35 years and above have about 35% of the average hair count. This age group constitutes 60% of the male population.

   Assume the hair counts for men aged 35 years and above follow a normal distribution with a mean of $52,500$ hairs and a standard deviation of $10,000$ hairs.

   (a) **(2 points)** Calculate the expected number of men aged 35 years and above in Kigali.

   (b) **(2 points)** Determine the range within which approximately 68% of the hair counts for men aged 35 years and above lie, using the given normal distribution.

   (c) **(2 points)** Considering the calculated range, use the pigeonhole principle to show that at least 2 men will have the same hair count within this range. Assume hair counts are discrete values.

Law of Total Probability (8 marks)

Reading: The law of total probability states that if events $B_1, B_2, \ldots, B_n$ form a partition of the sample space, then the probability of any event A can be expressed as the sum of the probabilities of the event occurring given each partition multiplied by the probability of each partition. Mathematically, this can be formulated as:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n)$$

This expression is further simplified using the definition of conditional probability, yielding:

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_n)P(B_n)$$

This principle is particularly useful in scenarios where directly calculating $P(A)$ is complex but evaluating the conditional probabilities $P(A|B_i)$ and the probabilities of the conditioning events $P(B_i)$ is more feasible. Thus, the law of total probability serves as a fundamental tool in probability theory, enabling the analysis of complex events through their relationships with simpler, mutually exclusive events.

1. **(4 marks)** A factory produces three types of gadgets. Type A constitutes 50%, Type B constitutes 30%, and Type C constitutes 20% of the total production. The defect rates for these gadgets are 1%, 2%, and 3% respectively. Calculate the probability that a randomly selected gadget is defective.

2. **(4 marks)** Samsung has three factories—Factory Korea, Factory Japan, and Factory USA—that produce the same type of electronic component. The factory in Korea produces 50% of the components, 30% are produced in Japan, and the factory in the USA is responsible for 20% of production. The probability of a component being defective is 2% for Factory Korea, 4% for Factory Japan, and 5% for Factory USA.

   Suppose a randomly selected component from the company's entire production is found to be defective. What is the probability that it was produced by Factory Japan?