Probability and Statistics course

Introduction

Gabor Vigh

TTK Department of Probability Theory and Statistics
ELTE

February 23, 2026

Week 3

Materials

[Baron, 2014] [Ross, 2012]

Baron, M. (2014).
Probability and statistics for computer scientists.
CRC Press.

Ross, S. (2012).
First Course in Probability.
Pearson.

Random Variables

Frequently, when an experiment is performed, we are interested mainly in some function of the outcome as opposed to the actual outcome itself. For instance, in tossing dice, we are often interested in the sum of the two dice and are not really concerned about the separate values of each die. That is, we may be interested in knowing that the sum is 7 and may not be concerned over whether the actual outcome was $(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)$. These quantities of interest, or more formally these real-valued functions defined in the sample space, are known as random variables.

Example

Examples

Independent trials consisting of the flipping of a coin having probability p of coming up heads are continually performed until either a head occurs or a total of n flips is made. If we let X denote the number of times the coin is flipped, then X is a random variable taking on one of the values $1, 2, 3, \ldots, n$ with respective probabilities

Random Variables

Definition

A random variable is a function $X : \Omega \to \mathbb{R}$ such that for every $t \in \mathbb{R}$

$$\{\omega \in \Omega : X(\omega) \leq t\} = \{X \leq t\}$$

is an event, that is, it is in $\mathcal{A}$. A random variable is discrete, if its range is finite or countably infinite.

(Cumulative) Distribution function

Definition ((Cumulative) Distribution Function of the Random Variable X)

$F_X(x) = P(X < x)$. Properties of the Distribution Function:

- $0 \leq F_X(x) \leq 1$
- monotonically increasing
- left-continuous
- $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.

Theorem

For any random variable X, we have $P(a \leq X < b) = F(b) - F(a)$;
$P(a < X \leq b) = F(b^+) - F(a^+)$.

Discrete distributions

Definition (Discrete Random Variable)

Its range of values is at most countably infinite, that is, it consists of elements $\{x_1, \ldots, x_n, \ldots\}$. Its distribution is:

$$p_i := P(X = x_i) = P(\omega : X(\omega) = x_i)$$

Definition (Expected Value of a Discrete Random Variable)

Notation: $E[X]$. Let X be a discrete random variable that takes on values $x_1, x_2, \ldots$ with probabilities $p_1, p_2, \ldots$. Then

$$E[X] = \sum_{k=1}^{\infty} x_k p_k, \quad \text{if the infinite sum is absolutely convergent.}$$

Variance and Standard deviation

Definition (Variance of X)

$D^2X = E[(X - E[X])^2] = E[X^2] - E[X]^2$

Definition (Standard deviation of X)

$DX = \sqrt{D^2X}$

Bernoulli and Binomial Distribution

Definition (Bernoulli Distribution)

A Bernoulli trial models a single experiment with exactly two outcomes: Success (1) or Failure (0).

$$P(X = k) = p^k \dot{(}1 - p)^k, wherek \in \{0, 1\}$$

A single packet transmission is modeled as a Bernoulli trial where the outcome is binary.

Definition (Binomial Distribution)

The Binomial distribution models the number of successes in n independent Bernoulli trials, where each trial has a success probability p.

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, wherek \in \{0, 1, 2, \ldots, n\}$$

In high availability (HA) systems, we calculate the probability that a specific number of components remain functional.

- **Cluster Size (n):** 10 servers
- **Probability of uptime (p):** 0.99
- **Requirement:** At least 8 servers online ($k \geq 8$)

Example

Examples

Five fair coins are flipped. If the outcomes are assumed independent, find the probability mass function of the number of heads obtained.

Example

Examples

It is known that screws produced by a certain company will be defective with probability .01, independently of each other. The company sells the screws in packages of 10 and offers a money-back guarantee if there is more than 1 screw is defective. What proportion of packages sold must the company replace?

Poisson distribution

Definition (Poisson Distribution)

The probability of observing exactly k events in an interval is given by the Probability Mass Function:

$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$

Used to model the number of independent events occurring within a fixed time interval, given an average rate λ. ite Reliability Engineering (SRE) use the Poisson distribution to model discrete events (incoming requests) over a fixed time interval.

The Scenario: API Load Analysis

- **Average Arrival Rate (λ):** 500 requests/minute.
- **The Question:** What is the probability of a spike to exactly 800 requests ($k = 800$) in a single minute?

Example

Examples

Suppose that the number of typographical errors on a single page of this book has a Poisson distribution with parameter $\lambda = \frac{1}{2}$. Calculate the probability that there is at least one error on this page.

Example

Examples

Suppose that the probability that an item produced by a certain machine will be defective is 0.1. Find the probability that a sample of 10 items will contain at most 1 defective item.

Poisson Expected value

The expected value $E[X]$ of a discrete random variable X is defined as the sum of each possible value multiplied by its probability:

$$E[X] = \sum_{k=0}^{\infty} k \cdot P(X = k) = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

To simplify the summation, we can change the index. Let $j = k - 1$. When $k = 1$, $j = 0$. The sum becomes:

$$E[X] = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$

The series $\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$ is the Maclaurin series expansion for the exponential function e^{λ}:

$$E[X] = \lambda e^{-\lambda}(e^{\lambda}) = \lambda e^0 = \lambda \cdot 1 = \lambda$$

Important Discrete Distributions

Name (parameters)	Values (k)	$P(X=k)$	EX	D^2X
Bernoulli (p) ($=$ Binomial ($1,\ p$))	0, 1	$p^k(1-p)^{1-k}$	p	$p(1-p)$
Binomial (n, p)	$0, 1, \ldots, n$	$\binom{n}{k} p^k(1-p)^{n-k}$	np	$np(1-p)$
Poisson (λ)	$0, 1, \ldots$	$\dfrac{\lambda^k}{k!} e^{-\lambda}$	λ	λ
Geometric/Pascal (p) ($=$ Negative binomial ($1,\ p$))	$1, 2, \ldots$	$p(1-p)^{k-1}$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$
Negative binominal (n, p)	$n, n+1, \ldots$	$\binom{k-1}{n-1} p^n(1-p)^{k-n}$	$\dfrac{n}{p}$	$\dfrac{n(1-p)}{p^2}$
Hyper-geometric (N, M, n)	$0, 1, \ldots, n$	$\dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$	$n\dfrac{M}{N}$	$n\dfrac{M}{N}\left(1-\dfrac{M}{N}\right)\left(1-\dfrac{n-1}{N-1}\right)$

Further Examples

Examples

The following gambling game, known as the wheel of fortune (or chuck-a-luck), is quite popular at many carnivals and gambling casinos: A player bets on one of the numbers 1 through 6. Three dice are then rolled, and if the number bet by the player appears i times, $i = 1, 2, 3$, then the player wins i units; if the number bet by the player does not appear on any of the dice, then the player loses 1 unit. Is this game fair to the player?

Solution: The Wheel of Fortune (Chuck-a-Luck)

Problem Analysis: Expected Value of a Bet
Let X be the random variable representing the player's gain/loss. The player bets on a number (e.g., "4"). Three fair dice are rolled independently. The number of times the chosen number appears, Y, follows a Binomial distribution: $Y \sim B(n = 3, p = 1/6)$.
1. Probability Mass Function of Y (Matches)
The probability that the chosen number appears k times is $P(Y = k) = \binom{3}{k}(\frac{1}{6})^k(\frac{5}{6})^{3-k}$:

- $P(Y = 0) = \binom{3}{0}(5/6)^3 = 125/216 \approx 0.5787$ (Loss of 1 unit)
- $P(Y = 1) = \binom{3}{1}(1/6)^1(5/6)^2 = 75/216 \approx 0.3472$ (Win 1 unit)
- $P(Y = 2) = \binom{3}{2}(1/6)^2(5/6)^1 = 15/216 \approx 0.0694$ (Win 2 units)
- $P(Y = 3) = \binom{3}{3}(1/6)^3 = 1/216 \approx 0.0046$ (Win 3 units)

2. Calculating the Expected Value $E[X]$
$E[X] = \sum x \cdot P(X = x)$

$$E[X] = (-1) \cdot \frac{125}{216} + (1) \cdot \frac{75}{216} + (2) \cdot \frac{15}{216} + (3) \cdot \frac{1}{216}$$

$$E[X] = \frac{-125 + 75 + 30 + 3}{216} = \frac{-17}{216} \approx -0.0787$$

Conclusion: The game is **not fair**. On average, the player loses about 7.87 cents for every 1 unit bet. This "house edge" is why the game is popular in casinos.

Further Examples

Examples

A batch of goods contains 1% defective items. How many items must we randomly select and examine so that there is at least one defective item with a probability of at least 0.95, if we replace each selected item after inspection?

Solution: Determining Minimum Sample Size

Problem Analysis: Geometric Logic and the Complement
Let X denote the number of defective items found in n independent trials. The probability of an item being defective is $p = 0.01$, and the probability of it being functional is $q = 1 - p = 0.99$.

1. Defining the Requirement
We want to find the smallest n such that the probability of finding *at least one* defective item is at least 0.95:

$$P(X \geq 1) \geq 0.95$$

2. Using the Complement
The event $X \geq 1$ is the complement of finding zero defective items ($X = 0$).

$$1 - P(X = 0) \geq 0.95 \implies P(X = 0) \leq 0.05$$

Since the trials are independent and we replace the items:

$$P(X = 0) = (0.99)^n$$

3. Solving for n using Logarithms

$$(0.99)^n \leq 0.05$$

Taking the natural logarithm (ln) on both sides:

$$n \cdot \ln(0.99) \leq \ln(0.05)$$

Note that ln(0.99) is negative, so the inequality sign flips when we divide:

$$n \geq \frac{\ln(0.05)}{\ln(0.99)} \approx \frac{-2.9957}{-0.01005} \approx 298.07$$

Conclusion: We must select and examine at least **299 items** to be at least 95% sure that we find at least one defective item.

Further Examples

Examples

Roll a die as many times as the number of heads we get with two fair coins. Let X be the sum of the resulting numbers. Determine the distribution of X.

Solution: Sum of Dice with Random Number of Rolls

Problem Analysis: Compound Distribution
Let N be the number of heads obtained when flipping two fair coins. $N \sim B(n = 2, p = 0.5)$. The possible values for N are $\{0, 1, 2\}$ with probabilities:

$$P(N = 0) = 1/4, \quad P(N = 1) = 2/4 = 1/2, \quad P(N = 2) = 1/4$$

X is the sum of N dice rolls. If $N = 0$, then $X = 0$.

1. Conditional Distributions of X

- **Case** $N = 0$**:** $P(X = 0|N = 0) = 1$.
- **Case** $N = 1$**:** X is the result of one die. $P(X = k|N = 1) = 1/6$ for $k \in \{1, \ldots, 6\}$.
- **Case** $N = 2$**:** X is the sum of two dice. $P(X = k|N = 2)$ follows the triangular distribution:

$$P(X = k|N = 2) = \frac{6 - |7 - k|}{36} \text{ for } k \in \{2, \ldots, 12\}.$$

2. Applying the Law of Total Probability
$P(X = k) = \sum_{n=0}^{2} P(X = k|N = n)P(N = n)$.

- $P(X = 0) = 1 \cdot (1/4) = 1/4 = 9/36$
- For $k \in \{1, \ldots, 6\}$, $P(X = k)$ receives contributions from $N = 1$ and $N = 2$.
- For $k > 6$, $P(X = k)$ receives contributions only from $N = 2$.

3. Final Distribution (Example Values)
$P(X = 1) = P(X = 1|N = 1)P(N = 1) + P(X = 1|N = 2)P(N = 2) = (1/6)(1/2) + 0 = 1/12 \approx 0.0833$.
$P(X = 7) = P(X = 7|N = 1)P(N = 1) + P(X = 7|N = 2)P(N = 2) = 0 + (6/36)(1/4) = 1/24 \approx 0.0417$.
Conclusion: X takes values in $\{0, 1, \ldots, 12\}$. The total probability $\sum P(X = k) = 1$.

Further Examples

Examples

Flipping a coin (let p be the probability of heads), let X denote the length of the first sequence of identical outcomes. (For example, if the sequence is HHT..., then $X = 2$.) Determine the distribution of X.

Solution: Length of the First Sequence

Problem Analysis: Geometric Logic with Complementary Outcomes
Let X be the number of consecutive identical outcomes at the start of the sequence. The sequence can start with either a Head (H) or a Tail (T). Let p be the probability of Heads and $q = 1 - p$ be the probability of Tails.

1. Partitioning by the First Outcome
We use the Law of Total Probability by conditioning on the first flip:

$$P(X = k) = P(X = k \mid \text{1st is H})P(\text{1st is H}) + P(X = k \mid \text{1st is T})P(\text{1st is T})$$

2. Calculating Conditional Probabilities

- **Case 1st is H:** For $X = k$, we must have exactly k Heads followed by a Tail. Since the first flip is already known to be H, we need $k - 1$ more Heads and then 1 Tail:
$$P(X = k \mid \text{1st is H}) = p^{k-1}q$$
- **Case 1st is T:** Similarly, we need $k - 1$ more Tails and then 1 Head:
$$P(X = k \mid \text{1st is T}) = q^{k-1}p$$

3. The Probability Mass Function (PMF)
Substituting these into the total probability formula:

$$P(X = k) = (p^{k-1}q)p + (q^{k-1}p)q = p^k q + q^k p, \quad k = 1, 2, 3, \ldots$$

4. Verification for Fair Coin ($p = q = 1/2$)
If the coin is fair, the distribution simplifies to:

$$P(X = k) = (1/2)^k(1/2) + (1/2)^k(1/2) = (1/2)^k, \quad k = 1, 2, 3, \ldots$$

This is a Geometric distribution with parameter $1/2$.
Conclusion: The distribution of X is $P(X = k) = p^k q + q^k p$.

Further Examples

Examples

Suppose that 3 probability practice sessions have 15, 20, and 25 students, respectively. What is the expected size of a randomly selected student's group?

Solution: Expected Size of a Student's Group

Problem Analysis: Size-Biased Sampling
Let X be the size of the group of a randomly selected student. The probability that a student is in a specific group is proportional to the number of students in that group.

1. Total Population
The total number of students across all three sessions is:

$$N = 15 + 20 + 25 = 60$$

2. Probability Mass Function (PMF)
A student is selected with equal probability $1/60$. The probability that the selected student belongs to a group of size x is:

- $P(X = 15) = 15/60 = 1/4 = 0.25$
- $P(X = 20) = 20/60 = 1/3 \approx 0.3333$
- $P(X = 25) = 25/60 = 5/12 \approx 0.4167$

3. Calculating the Expected Value $E[X]$
The expected group size from the student's perspective is the weighted average:

$$\begin{aligned} E[X] &= \sum x \cdot P(X = x) \\ &= 15\left(\frac{15}{60}\right) + 20\left(\frac{20}{60}\right) + 25\left(\frac{25}{60}\right) \\ &= \frac{225 + 400 + 625}{60} = \frac{1250}{60} = \frac{125}{6} \approx 20.833 \end{aligned}$$

Conclusion: While the average group size is 20, the expected size of a *randomly selected student's* group is **20.833**. This is because students are more likely to be found in larger groups.

Further Examples

Examples

In a beech forest, the number of beech saplings per square meter follows a Poisson distribution with a parameter $\lambda = 2.5$ saplings / m^2. What is the probability that in a 1 m^2 sample we find:

a at most one, or

b more than three saplings?

c State the expected value and the standard deviation of the number of saplings.

Solution: Beech Saplings Distribution

Problem Analysis: Poisson Distribution

Let X be the number of beech saplings in 1 m^2. $X \sim \text{Pois}(\lambda = 2.5)$. The PMF is given by: $P(X = k) = \frac{e^{-2.5} \cdot 2.5^k}{k!}, \quad k = 0, 1, 2, \ldots$

a) Probability of at most one ($X \leq 1$)

$P(X \leq 1) = P(X = 0) + P(X = 1)$

- $P(X = 0) = e^{-2.5} \approx 0.0821$
- $P(X = 1) = e^{-2.5} \cdot 2.5 \approx 0.2052$

$P(X \leq 1) = 0.0821 + 0.2052 = \mathbf{0.2873}$

b) Probability of more than three ($X > 3$)

$P(X > 3) = 1 - P(X \leq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$

- $P(X = 2) = \frac{e^{-2.5} \cdot 2.5^2}{2} = \frac{0.0821 \cdot 6.25}{2} \approx 0.2565$
- $P(X = 3) = \frac{e^{-2.5} \cdot 2.5^3}{6} = \frac{0.0821 \cdot 15.625}{6} \approx 0.2138$

$P(X > 3) = 1 - (0.0821 + 0.2052 + 0.2565 + 0.2138) = 1 - 0.7576 = \mathbf{0.2424}$

c) Expected Value and Standard Deviation

For a Poisson distribution $X \sim \text{Pois}(\lambda)$:

- **Expected Value:** $E[X] = \lambda = \mathbf{2.5}$
- **Standard Deviation:** $\sigma = \sqrt{\text{Var}(X)} = \sqrt{\lambda} = \sqrt{2.5} \approx \mathbf{1.5811}$

Further Examples

Examples

We organize a party, and we know in advance that the number of participants is eight with probability $1/4$, nine with probability $1/3$ and ten otherwise. Calculate the expected number and the variance of the number of participants.

Solution: Party Participants Distribution

Problem Analysis: Expected Value and Variance
Let X be the random variable representing the number of participants. The possible values are $\{8, 9, 10\}$.
1. Determining the Probability Mass Function (PMF)
Based on the problem statement:

- $P(X = 8) = 1/4 = 0.25$
- $P(X = 9) = 1/3 \approx 0.3333$
- $P(X = 10) = 1 - (1/4 + 1/3) = 1 - 7/12 = 5/12 \approx 0.4167$

2. Calculating the Expected Number $E[X]$

$$E[X] = \sum x \cdot P(X = x)$$

$$E[X] = 8\left(\frac{3}{12}\right) + 9\left(\frac{4}{12}\right) + 10\left(\frac{5}{12}\right) = \frac{24 + 36 + 50}{12} = \frac{110}{12} = \frac{55}{6} \approx \mathbf{9.1667}$$

3. Calculating the Variance Var(X)
We use the formula $\text{Var}(X) = E[X^2] - (E[X])^2$. First, calculate $E[X^2]$:

$$E[X^2] = 8^2\left(\frac{3}{12}\right) + 9^2\left(\frac{4}{12}\right) + 10^2\left(\frac{5}{12}\right) = \frac{192 + 324 + 500}{12} = \frac{1016}{12} = \frac{254}{3} \approx 84.6667$$

Now, calculate Variance:

$$\text{Var}(X) = \frac{254}{3} - \left(\frac{55}{6}\right)^2 = \frac{254}{3} - \frac{3025}{36} = \frac{3048 - 3025}{36} = \frac{23}{36} \approx \mathbf{0.6389}$$

Conclusion: The expected number of participants is **9.1667**, with a variance of **0.6389**.

Further Examples

Examples

Peter is late from the university with probability 0.1 every day, independently of the other days. We know that there are 21 days in March where he goes to the university.
a) What is the probability that he is never late in March? What is the probability of exactly 1 occasion when he is late? Of 2? In general, what is the probability that he is late in March exactly k times?
b) What is the distribution of the number of occasions when he is late?
c) Randomize 100 samples from the distribution of the number of late arrivals of Peter in March. Make a histogram, and calculate the average in python.
d) What is the expectation of number of occasions when he is late?
e) What is the variance of the number of occasions when he is late?

Solution: Peter's Late Arrivals in March

a) Calculation of Probabilities
Let X be the number of days Peter is late. X follows a Binomial distribution with $n = 21$ and $p = 0.1$. The probability of being late exactly k times is:

$$P(X = k) = \binom{21}{k}(0.1)^k(0.9)^{21-k}$$

- **Never late ($k = 0$):** $P(X = 0) = (0.9)^{21} \approx 0.1094$
- **Exactly 1 time ($k = 1$):** $P(X = 1) = \binom{21}{1}(0.1)^1(0.9)^{20} \approx 0.2553$
- **Exactly 2 times ($k = 2$):** $P(X = 2) = \binom{21}{2}(0.1)^2(0.9)^{19} \approx 0.2837$

b) Distribution
The number of occasions X follows a **Binomial Distribution**: $X \sim B(21, 0.1)$. **c) Python Simulation (Example Code)**

```
import numpy as np
import matplotlib.pyplot as plt

samples = np.random.binomial(n=21, p=0.1, size=100)
print(f"Average of samples: {np.mean(samples)}")

plt.hist(samples, bins=range(22), align='left', rwidth=0.8)
plt.title("Histogram of 100 Samples (n=21, p=0.1)")
plt.show()
```

d) Expectation
For a Binomial distribution, $E[X] = n \cdot p$:

$$E[X] = 21 \cdot 0.1 = \mathbf{2.1}$$

e) Variance
For a Binomial distribution, $\mathrm{Var}(X) = n \cdot p \cdot (1 - p)$:

$$\mathrm{Var}(X) = 21 \cdot 0.1 \cdot 0.9 = \mathbf{1.89}$$

Further Examples

Examples

Suppose that we have 10 servers in a system. On each day, each of them breaks down with probability 0.01, independently of each other. Let Z be the number of servers (among these 10) which break down tomorrow (that is, on a given day). Calculate the probability $\mathbb{P}(Z = 2)$, and the expectation and variance of Z.

Solution: Server System Reliability

Problem Analysis: Binomial Distribution
Let Z be the number of servers that break down. Since failures are independent and occur with a constant probability, Z follows a Binomial distribution: $Z \sim B(n = 10, p = 0.01)$.

1. Calculating the Probability $P(Z = 2)$
Using the binomial probability mass function $P(Z = k) = \binom{n}{k} p^k (1-p)^{n-k}$:

$$\begin{aligned} P(Z = 2) &= \binom{10}{2} (0.01)^2 (0.99)^{10-2} \\ &= 45 \cdot (0.0001) \cdot (0.99)^8 \\ &\approx 45 \cdot 0.0001 \cdot 0.9227 \\ &\approx \mathbf{0.00415} \end{aligned}$$

2. Expectation of Z
For a Binomial distribution, the expected value is $E[Z] = n \cdot p$:

$$E[Z] = 10 \cdot 0.01 = \mathbf{0.1}$$

On average, 0.1 servers break down per day (or one server every 10 days).

3. Variance of Z
The variance is given by $\text{Var}(Z) = n \cdot p \cdot (1-p)$:

$$\text{Var}(Z) = 10 \cdot 0.01 \cdot 0.99 = \mathbf{0.099}$$

Conclusion: The probability of exactly two servers failing is quite low (0.415%), reflecting the high reliability of the individual components.

Examples

Suppose that, according to our data from the last decades, the average number of earthquakes per year is 3.42 in a given city. Suppose that the number of earthquakes in a given year has Poisson distribution, and that its expectation is equal to the average that we observed.

a) Let us randomize a sample of size 200 from the distribution of the number of earthquakes in a given year, and make a histogram. What is the mean of the data? What is the proportion of 3 in the sample? What is the proportion of numbers which are at least 4?

b) What is the probability that there are exactly 3 earthquakes in a year?

c) What is the probability that there are at least 4 earthquakes in a year?

Solution: Earthquake Frequency Analysis

Problem Analysis: Poisson Distribution

The number of earthquakes per year X follows a Poisson distribution with $\lambda = 3.42$. The PMF is $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$.

a) Simulation Results (Sample size $n = 200$)

Running a Python simulation (using `np.random.poisson(3.42, 200)`):

- **Mean of the data:** The sample mean is approximately 3.34 (close to the theoretical $\lambda = 3.42$).
- **Proportion of 3:** The frequency of years with exactly 3 earthquakes in the sample is 0.22.
- **Proportion at least 4:** The frequency of years with 4 or more earthquakes is 0.425.

b) Probability of Exactly 3 Earthquakes

Using the theoretical Poisson formula:

$$P(X = 3) = \frac{e^{-3.42} \cdot 3.42^3}{3!} = \frac{0.0327 \cdot 40.0017}{6} \approx \mathbf{0.2181}$$

c) Probability of At Least 4 Earthquakes

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$$

Calculating individual probabilities:

- $P(X = 0) = e^{-3.42} \approx 0.0327$
- $P(X = 1) = e^{-3.42} \cdot 3.42 \approx 0.1119$
- $P(X = 2) = \frac{e^{-3.42} \cdot 3.42^2}{2} \approx 0.1913$
- $P(X = 3) \approx 0.2181$

$$P(X \geq 4) = 1 - (0.0327 + 0.1119 + 0.1913 + 0.2181) = 1 - 0.5540 = \mathbf{0.4460}$$

Conclusion: The simulation proportions (0.22 and 0.425) are quite close to the theoretical probabilities (0.2181 and 0.4460).

Further Examples

Examples

Suppose that the number of downloads of a webpage within an hour has Poisson distribution, and the probability that there are 0 downloads is $1/e^2$. Suppose furthermore that the number of downloads independent for disjoint time intervals. a) What is the variance of the number of downloads within an hour? b) Given that the number of downloads within an hour is at most 1, what is the probability that there are 0 downloads within this hour?

Solution: Webpage Downloads Distribution

Problem Analysis: Identifying the Poisson Parameter
Let X be the number of downloads in an hour. X follows a Poisson distribution: $X \sim \text{Pois}(\lambda)$. The probability of zero downloads is given as $P(X=0) = 1/e^2 = e^{-2}$. Using the Poisson formula $P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$:

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

By comparison, $e^{-\lambda} = e^{-2}$, which implies $\lambda = 2$.
a) Variance of the Number of Downloads
A key property of the Poisson distribution is that the variance is equal to the expected value (λ):

$$\text{Var}(X) = \lambda = \mathbf{2}$$

b) Conditional Probability Calculation
We seek the probability $P(X=0 \mid X \leq 1)$. Using the definition of conditional probability:

$$P(X=0 \mid X \leq 1) = \frac{P(X=0 \cap X \leq 1)}{P(X \leq 1)}$$

Because $X=0$ is a subset of the condition $X \leq 1$, the intersection is simply $P(X=0)$. The denominator is the sum $P(X \leq 1) = P(X=0) + P(X=1)$.
Calculating the terms with $\lambda = 2$:

- $P(X=0) = e^{-2}$
- $P(X=1) = \frac{e^{-2} \cdot 2^1}{1!} = 2e^{-2}$

Substituting these into the fraction:

$$P(X=0 \mid X \leq 1) = \frac{e^{-2}}{e^{-2} + 2e^{-2}} = \frac{e^{-2}}{3e^{-2}} = \mathbf{\frac{1}{3}}$$

Conclusion: Given that at most one download occurred, there is a **1/3** probability that there were no downloads at all.