DATA SCIENCE AND STATISTICS - 23SC3201

TUTORIAL - 1

NAME OF THE STUDENT	**KUNAL KUMAR DAPPU**
ROLL NUMBER	**2320030167**
NAME OF THE INSTRUCTOR	**SARIKA SABAVATH**
DATE AND SIGNATURE OF THE STUDENT	Digitally signed by KUNAL KUMAR DAPPU Date: 2025.08.04 10:27:24 +05'30'

TOPICS COVERED

- Bayes Theorem

- Probability Addition Rule

- Probability Multiplication Rule

Q1. Bayes Theorem

1.1 Definition

Bayes Theorem is a fundamental result in probability theory that describes how to update the probabilities of hypotheses when given evidence. It provides a way to revise existing predictions or theories (probabilities) given new or additional evidence.

Mathematical Formula:

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

Where:

- P(A|B): Posterior Probability (Probability of event A occurring given that B is true)

- P(B|A): Likelihood (Probability of event B occurring given that A is true)

- P(A): Prior Probability (Probability of event A occurring)

- P(B): Marginal Probability (Total probability of event B occurring)

1.2 Types of Events

- **Independent Events:** The occurrence of one does not affect the probability of the other.
- **Dependent Events:** The probability of one event depends on the outcome of another.
- **Mutually Exclusive Events:** Events that cannot happen at the same time.

Bayes Theorem is especially useful for **dependent events**, where prior information is updated with new evidence.

1.3 Detailed Example

Medical Diagnosis:

Suppose there is a rare disease affecting 0.5% of the population. A test for this disease is 99% accurate (true positive rate) and has a 5% false positive rate.

Events:

- Let D = Person has disease
- Let ¬D = Person does not have disease
- Let T = Test is positive

Probabilities:

- P(D)=0.005
- P(¬D)=0.995P(¬D) = 0.995P(¬D)=0.995
- P(T|D)=0.99P(T|D) = 0.99P(T|D)=0.99
- P(T|¬D)=0.05P(T|¬D) = 0.05P(T|¬D)=0.05

Question:

What is the probability that a person who tests positive **actually has the disease**?

Step-by-step Solution:

1. **Calculate total probability of testing positive :**

$$P(T) = P(T|D) \cdot P(D) + P(T|\neg D) \cdot P(\neg D)$$

$$P(T) = (0.99 \times 0.005) + (0.05 \times 0.995) = 0.00495 + 0.04975 = 0.0547$$

2. **Apply Bayes Theorem:**

$$P(D|T) = \frac{P(T|D) \cdot P(D)}{P(T)}$$

$$P(D|T) = \frac{0.99 \times 0.005}{0.0547} \approx 0.0905$$

So, **even after a positive test, the probability the person has the disease is only ~ 9%**.

1.4 Python Program

```
# Bayes Theorem Example: Medical Diagnosis

P_D = 0.005
P_not_D = 0.995
P_T_given_D = 0.99
P_T_given_not_D = 0.05

P_T = P_T_given_D * P_D + P_T_given_not_D * P_not_D
P_D_given_T = (P_T_given_D * P_D) / P_T

print(f"Probability of having the disease after positive test:
    {P_D_given_T:.2%}")
```

1.5 Real-life Use Cases

- **Medical Testing:** Diagnosing diseases, as above.
- **Spam Detection:** Updating the probability an email is spam based on certain keywords.
- **Machine Learning:** Naive Bayes classifiers for text classification.
- **Forensics:** Updating likelihood of guilt given new evidence.

1.6 Types of Implementations

- Manual calculation for simple cases
- **Naive Bayes algorithms** in machine learning (e.g., sklearn.naive_bayes)
- **Bayesian Networks** for complex dependencies (e.g., pgmpy in Python)
- **Markov Chain Monte Carlo (MCMC)** for Bayesian inference in advanced statistics

2. Probability Addition Rule

2.1 Definition

The Addition Rule determines the probability that at least one of several events occurs.

Mutually Exclusive Events:

$$P(A \cup B) = P(A) + P(B)$$

Non-Mutually Exclusive Events:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

where $P(A \cap B)$ is the probability that both events occur.

2.2 Types of Events

- **Mutually Exclusive Events:** Two events cannot happen at the same time (e.g., drawing a king or a queen from a single card draw).
- **Non-Mutually Exclusive Events:** Two events can occur together (e.g., drawing a red card or a king).

2.3 Detailed Example

Example 1: Mutually Exclusive

Question:
What is the probability of rolling a 2 or a 5 on a six-sided die?

Solution:

- P(2) = 1/6
- P(5) = 1/6
- P(2 and 5) = 0 (cannot roll both at once)

$$P(2 \text{ or } 5) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

Example 2 :Non Mutually Exclusive

Question:

What is the probability of drawing a card that is a **heart** or a **king** from a standard deck?

Solution:

- P(heart) = 13/52
- P(king) = 4/52
- P(heart and king) = 1/52 (king of hearts)

$$P(\text{heart or king}) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

2.4 Python Program

```
# Addition Rule: Mutually Exclusive
P_2 = 1/6
P_5 = 1/6
P_2_or_5 = P_2 + P_5    # Mutually exclusive
print(f"Probability of rolling 2 or 5: {P_2_or_5:.2f}")

# Addition Rule: Non-Mutually Exclusive
P_heart = 13/52
P_king = 4/52
P_heart_and_king = 1/52
P_heart_or_king = P_heart + P_king - P_heart_and_king
print(f"Probability of heart or king: {P_heart_or_king:.2f}")
```

2.5 Real-life Use Cases

- **Quality Control:** Probability of a product failing due to defect A or B.
- **Elections:** Probability that a voter supports candidate X or Y.
- **Insurance:** Probability of claims from fire or theft.

2.6 Types of Implementations

- Direct calculation using rules
- Simulation (e.g., Monte Carlo) for complex real-world events
- Use of probability libraries (numpy, scipy.stats)

3. Probability Multiplication Rule

3.1 Definition

The Multiplication Rule calculates the probability that **two or more events** occur together:

Independent Events:

$$P(A \cap B) = P(A) \times P(B)$$

Dependent Events:

$$P(A \cap B) = P(A) \times P(B|A)$$

where $P(B|A)$ is the probability of B given A.

3.2 Types of Events

- **Independent Events:** The occurrence of one does not affect the other (e.g., flipping two coins).
- **Dependent Events:** The outcome of one event affects the probability of the next (e.g., drawing cards without replacement).

3.3 Detailed Examples

Example 1: Independent Events

Question:

What is the probability of flipping heads and rolling a 6 on a die?

Solution:

P(heads) = 0.5

P(6) = 1/6

$$P(\text{heads and } 6) = 0.5 \times \frac{1}{6} = \frac{1}{12} \approx 0.083$$

Example 2: Dependent Events

Question:

What is the probability of drawing 2 aces in a row (without replacement) from a deck?

Solution:

P(first ace) = 4/52

P(second ace | first ace drawn) = 3/51

$$P(2\text{ aces}) = \frac{4}{52} \times \frac{3}{51} = \frac{12}{2652} \approx 0.0045$$

3.4 Python Program

```
# Multiplication Rule: Independent Events
P_heads = 0.5
P_six = 1/6
P_heads_and_six = P_heads * P_six
print(f"Probability of heads and rolling a 6: {P_heads_and_six
    :.3f}")

# Multiplication Rule: Dependent Events
P_first_ace = 4/52
P_second_ace_given_first = 3/51
P_two_aces = P_first_ace * P_second_ace_given_first
print(f"Probability of drawing 2 aces in a row: {P_two_aces:.4f}")
```

3.5 Real-life Use Cases

- **Manufacturing:** Probability two machines fail on same day.
- **Genetics:** Probability of inheriting two independent traits.
- **Project Management:** Probability two risk events occur together.

3.6 Types of Implementations

- Direct calculation for simple cases
- Probability trees for dependent events
- Simulation for complex dependencies
- Bayesian Networks for advanced dependencies

References & Further Reading

1. [Khan Academy: Probability and Statistics]
(https://www.khanacademy.org/math/statistics-probability)

2. [Wikipedia: Bayes' Theorem]

(https://en.wikipedia.org/wiki/Bayes%27_theorem)

3. [Python scikit-learn: Naive Bayes]

(https://scikit-learn.org/stable/modules/naive_bayes.html)

4. [Probability Rules: Addition and Multiplication]

(https://stattrek.com/probability/probability-rules.aspx)

5. [pgmpy: Probabilistic Graphical Models in Python]

(https://pgmpy.org/)

Submitted by Kunal Kumar Dappu – 2320030167

04-10-2025.