SNS COLLEGE OF TECHNOLOGY

Coimbatore-35
An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

23AMB201 - MACHINE LEARNING

II YEAR IV SEM

UNIT I – INTRODUCTION

TOPIC 5 – Probability theory - Probability Distributions

Recall Probability & Statistics

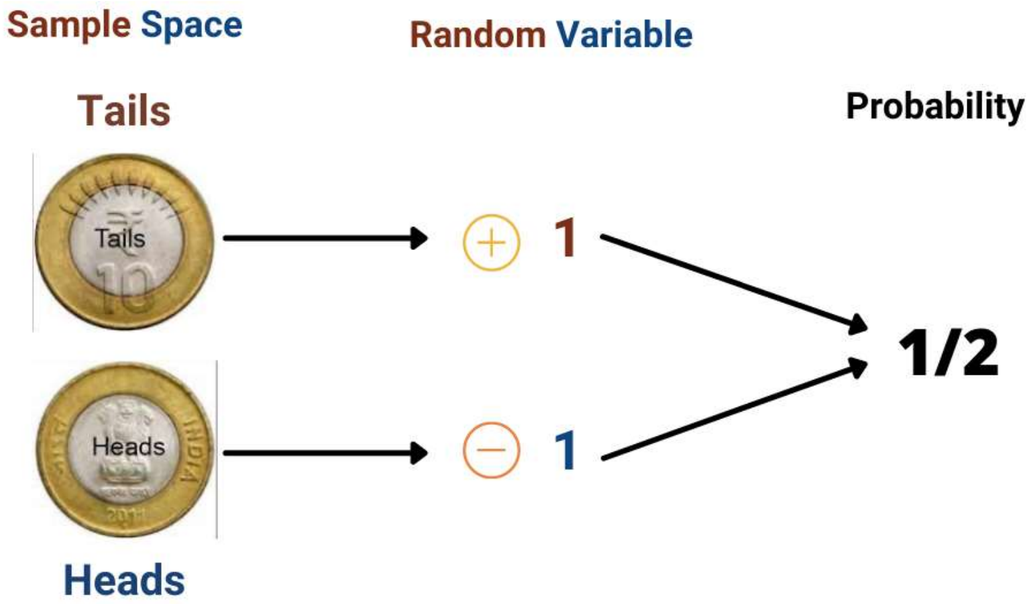
1. What do you mean by Probability?
2. Formula?
3. Example: Coin? How to find the favorable outcomes of 4 coins?
4. Function of Statistics in Machine Learning?
5. Basic methods used to avoid noise?

Probability Theory

Probability simply talks about how **likely is the event to occur,** and its value always lies between 0 and 1 (inclusive of 0 and 1)

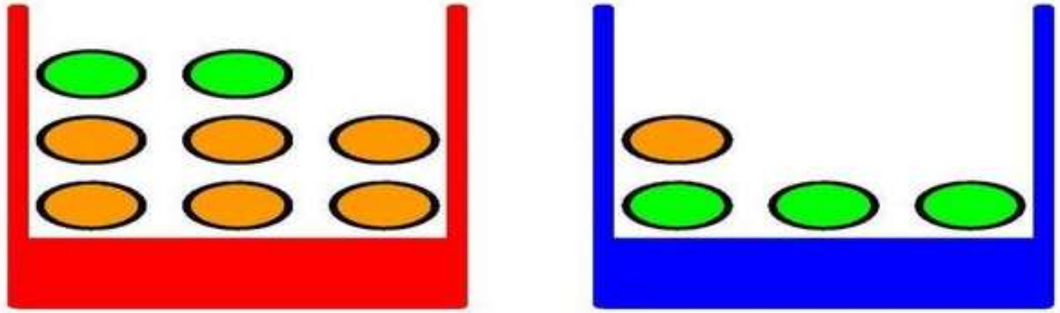
Probability Theory Concepts : Terminologies

1. **Random Experience:** Any event that can be repeated multiple times and its outcome is not hampered by its repetition is called a Random Experiment.
    1. Discrete
    2. Continuous
2. **Sample space:** The set of all possible outcomes for any random experiment is called sample space. {Tails, Heads}
3. **Event:** The outcome of any experiment is called an event. Ex: Independent, Dependent, Mutual and Equally likely events
4. **Probability:** The probability of an event is a number between 0 and 1 that represents the likelihood of the event occurring. A chance of 0 means that the event is impossible, and a probability of 1 means that the event is specific



Probability Theory

- Imagine two boxes one red and another blue.
- Red box has 6 apple and 2 oranges
- Blue box has 1 apple and 3 orange
- Suppose, we pick the red box 40% of the time and blue box 60% of the time.



- The **identity of the box** that will be chosen is **random variable** which we here denote by $\boldsymbol{B}$ and can take two possible values $\boldsymbol{r}(red\ box)$ and $\boldsymbol{b}(blue\ box)$.
- The **identity of the fruit** is also a **random variable** which we here denote by $\boldsymbol{F}$ and can take two values $\boldsymbol{a}(for\ apple)$ and $\boldsymbol{o}(for\ orange)$

$$\boldsymbol{p(B=r)=\frac{4}{10}} \qquad \boldsymbol{p(B=b)=\frac{6}{10}}$$

Probability Theory Formulas

- **Theoretical Probability Formula:** (Number of Favourable Outcomes) / (Number of Total Outcomes)
- **Empirical Probability Formula:** (Number of times event A happened) / (Total number of trials)
- **Addition Rule of Probability:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Complementary Rule of Probability:** $P(A') = 1 - P(A)$
- **Independent Events:** $P(A \cap B) = P(A) \cdot P(B)$
- **Conditional Probability:** $P(A \mid B) = P(A \cap B) / P(B)$
- **Bayes' Theorem:** $P(A \mid B) = P(B \mid A) \cdot P(A) / P(B)$

Example

Example 2: A fair coin is tossed three times. What is the probability of getting exactly two heads?

Total possible outcomes when tossing a coin three times = 23 = 8.

Possible outcomes: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT.

Outcomes with exactly two heads: HHT, HTH, THH (3 outcomes).

Probability of getting exactly two heads:

P(exactly 2 heads)=Number of favorable outcomes/ Total outcomes.

P(exactly 2 heads)=3/ 8.

Bayes Theorem

Used to determine the conditional probability of event

A when event B has already occurred.

$$P(A|B) = P(B|A)P(A) / P(B)$$

P(A) and P(B) are the probabilities of events A and B

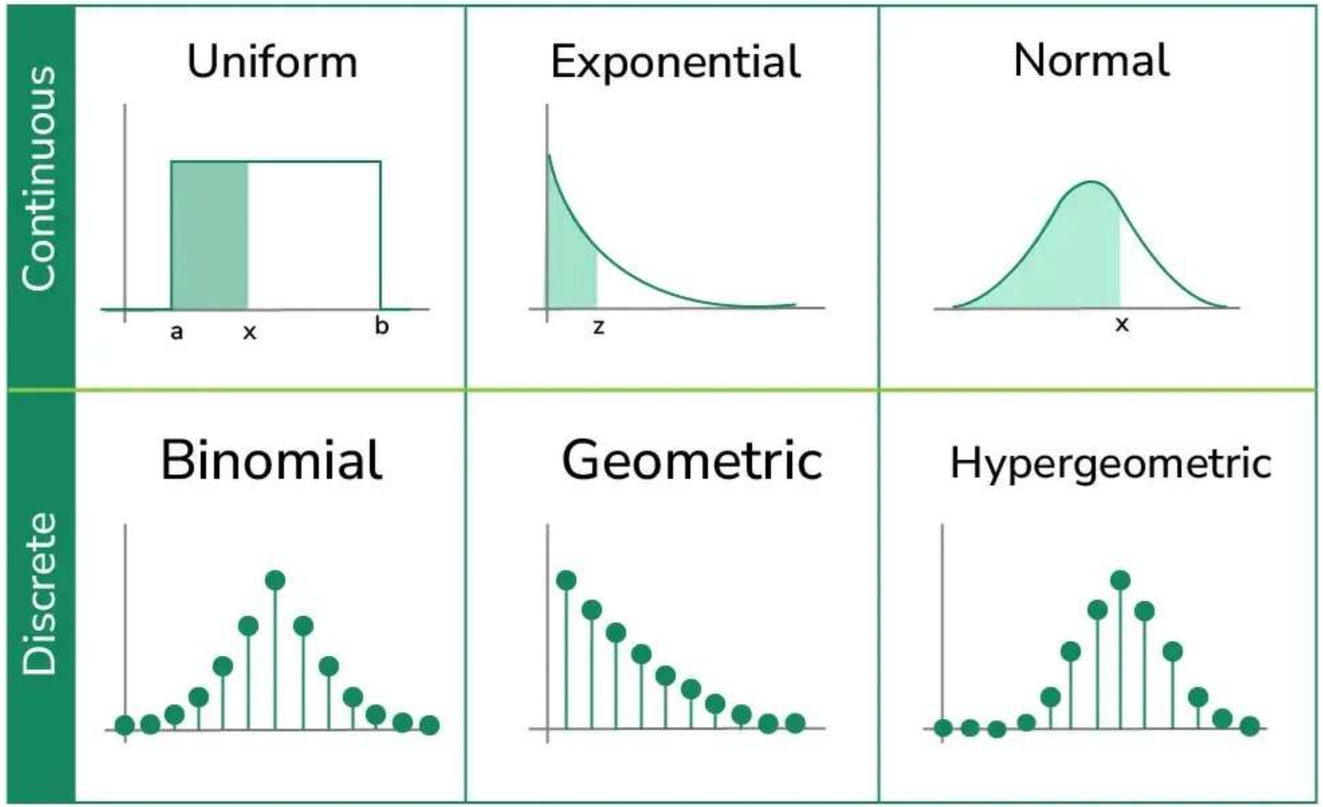
P(A|B) is the probability of event A when event B happens

P(B|A) is the probability of event B when A happens

1. **Hypotheses**
2. **Priori Probability**
3. **Posterior Probability**

Probability Distribution

Probabilities of different outcomes are assigned to the possible values of a random variable.



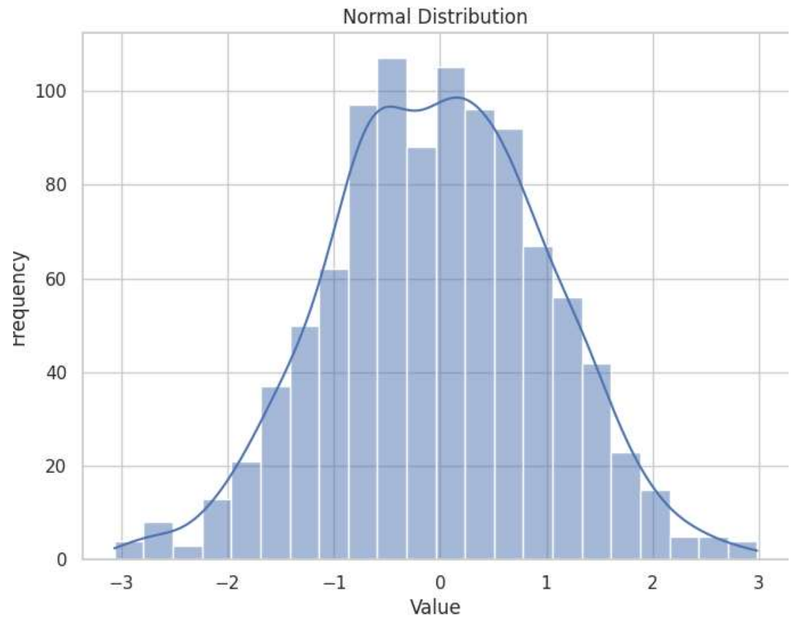
Probability Distribution: Example

```
[1]  import numpy as np

     mean = 0
     std_dev = 1
     num_samples = 1000

     samples = np.random.normal(mean, std_dev, num_samples)
```

0s




```
import matplotlib.pyplot as plt
import seaborn as sns

sns.set(style="whitegrid")

plt.figure(figsize=(8, 6))
sns.histplot(samples, kde=True)
plt.title("Normal Distribution")
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.show()
```

8s

References

1. Aurélien Géron "Hands-On Machine Learning with Scikit-Learn and TensorFlow" Publisher(s): O'Reilly Media, Inc 2017.
2. https://www.geeksforgeeks.org/probability-theory/
3. https://www.analyticsvidhya.com/blog/2021/04/probability-theory-basics-in-machine-learning/
4. https://www.enjoymathematics.com/blog/probability-theory-for-machine-learning