Basics of Probability theory: All you need to get started with Statistics and/or Machine Learning.

Diptaraj Sen · Follow

8 min read · Jun 2

Probability theory is the branch of mathematics that deals with the analysis of random phenomena. It is concerned with the likelihood of events occurring. It also gives us interesting mathematical ways in which probabilities can be combined to make predictions which is used heavily in the domain of Statistics, Machine Learning and other related fields where we are interested in making a prediction.

In this article I will cover the foundational concepts of probability theory that will help you as a pre requisite for Statistics and Machine Learning. I will try to cover the topics in a very precise way so as you can quickly jump into statistics and/or machine learning after finishing this.

Photo by Jonathan Greenaway on Unsplash

Topics that will be covered here:

- *Basic terminology:* ***sample space, event, outcome, probability, etc.***
- *Different types of probability:* ***classical, empirical, and subjective.***
- *Concept of* ***probability axioms*** *and laws,* ***including the addition and multiplication rules.***
- *Concept of* ***conditional probability***
- ***Bayes' theorem***

Basic Terminology:

1. **Experiment**: An experiment is a process or activity that produces a set of outcomes. It can be a physical or conceptual process, such as flipping a coin, rolling a dice, or selecting a card from a deck.
2. **Sample Space**: The sample space, denoted by S, is the set of all possible outcomes of a random experiment or event. It is the collection of all the distinct outcomes that can occur as a result of an experiment.

3. **Event**: An event is a subset of the sample space. It represents a specific outcome or a collection of outcomes of interest. Events can be simple (single outcome) or compound (multiple outcomes).

4. **Outcome**: An outcome is a particular result of a random experiment or event. It is a single element of the sample space. For example, when rolling a fair six-sided dice, the outcomes can be 1, 2, 3, 4, 5, or 6.

5. **Probability:** Probability is a measure of the likelihood of an event occurring. It is a numerical value between 0 and 1 (inclusive). A probability of 0 means the event is impossible, while a probability of 1 means the event is certain to occur. The probability of an event A is denoted by **P(A)**.

6. **Equally Likely Outcomes**: When all the outcomes in a sample space have the same chance of occurring, they are called equally likely outcomes. For example, when flipping a fair coin, the outcomes (heads or tails) are equally likely.

7. **Complementary Event**: The complementary event of an event A, denoted by **A'**, is the event that consists of all outcomes in the sample space S that are not in A. In other words, it represents all the outcomes that do not satisfy the conditions of event A.

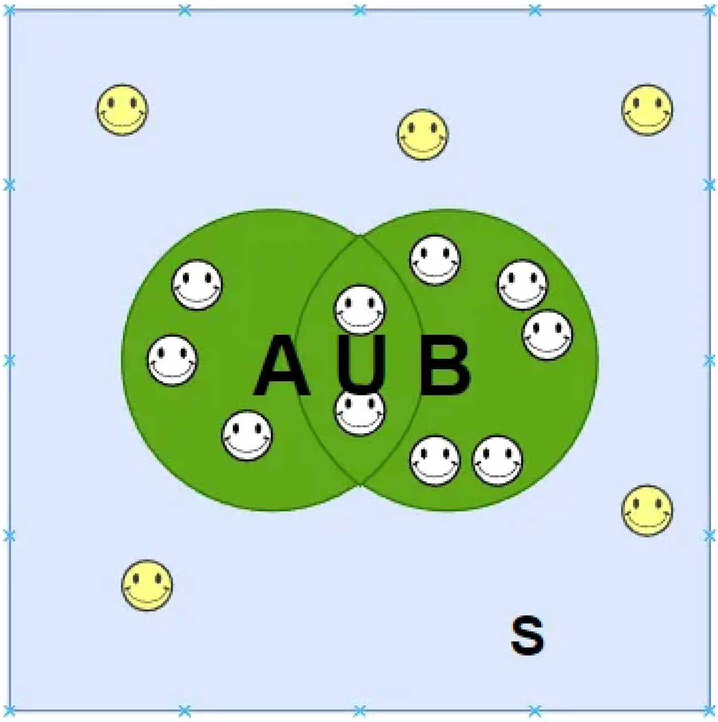
8. **Union of Events**: The union of two events A and B, denoted by **A ∪ B**, is the event that consists of all outcomes that are in either A or B or both.

9. **Intersection of Events**: The intersection of two events A and B, denoted by **A ∩ B**, is the event that consists of all outcomes that are common to both A and B.

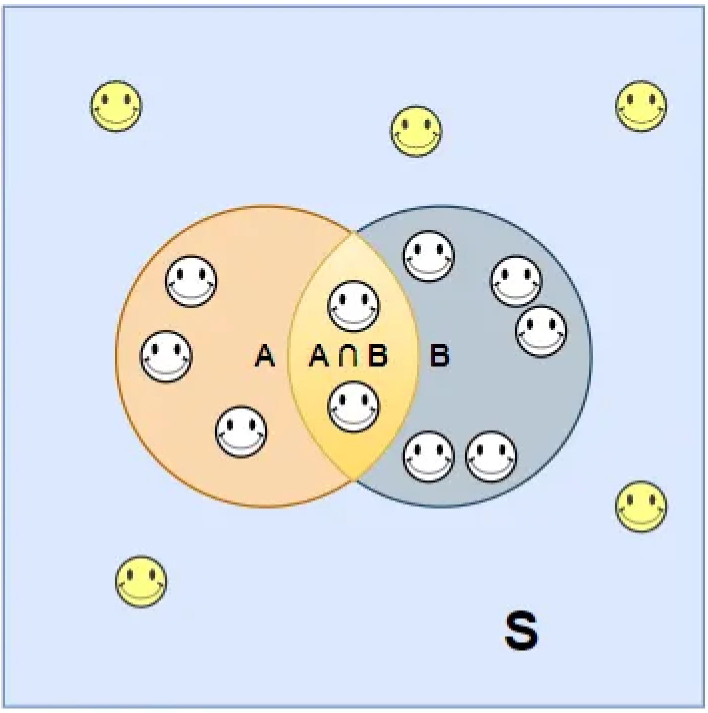
10. **Independent Events:** Two events A and B are independent if the occurrence of one event does not affect the probability of the other event. In other words, the probability of A and B occurring together is the product of their individual probabilities.

11. **Dependent Events:** Two events A and B are dependent if the occurrence or non-occurrence of one event affects the probability of the other event.

These basic terminologies provide a foundation for understanding probability and its applications.



This represents Union of two Events



This represents Intersection of two Events

Different types of probability:

1. **Classical Probability:** Classical probability, also known as theoretical probability, is based on the assumption of equally likely outcomes. It applies to situations where the sample space is finite and all outcomes are equally likely to occur. The probability of an event A is calculated by dividing the number of favorable outcomes for A by the total number of possible outcomes in the

sample space. ***For example***, when rolling a fair six-sided dice, each outcome (1, 2, 3, 4, 5, or 6) has an equal chance of occurring, and the probability of rolling a specific number, say 5, is 1/6.

2. **Empirical Probability**: Empirical probability, also known as experimental probability, is based on observations and data collected from experiments or real-life events. It involves estimating the probability of an event by conducting experiments and calculating the relative frequency of the event occurring. To determine the empirical probability, one performs repeated trials of an experiment and records the number of times the event of interest occurs. The empirical probability of an event A is calculated by dividing the number of times A occurs by the total number of trials. ***For example***, if you flip a coin 100 times and it lands on heads 55 times, the empirical probability of getting heads is 55/100 = 0.55.

3. **Subjective Probability**: Subjective probability is based on personal judgment, beliefs, or opinions about the likelihood of an event occurring. It is influenced by an individual's knowledge, experience, and subjective assessment of the situation. Subjective probability is often used in situations where there is insufficient data or when personal judgment is required. Subjective probabilities are expressed on a scale from 0 to 1, where 0 represents impossibility and 1 represents certainty. Different individuals may assign different subjective probabilities to the same event based on their own assessments. For example, a person might assign a subjective probability of 0.7 to the belief that it will rain tomorrow based on their interpretation of the weather forecast and their past experience.

It's important to note that while classical and empirical probabilities are based on objective measures and observations, subjective probability is inherently based on personal opinions and judgments.

Probability axioms and laws:

Probability axioms and laws provide the foundation for calculating probabilities and making probabilistic statements. Here's an explanation of probability axioms and the addition and multiplication rules:

Probability Axioms:

1. ***Non-Negativity Axiom***: The probability of any event is a non-negative value. That is, for any event A, **P(A) ≥ 0.**

2. ***Normalization Axiom***: The probability of the entire sample space is 1. That is, for the sample space S, **P(S) = 1.**

3. ***Additivity Axiom***: For any mutually exclusive events (events that cannot occur simultaneously), the probability of their union is equal to the sum of their individual probabilities. That is, for mutually exclusive events A and B, if A ∩ B = ∅ (empty set), then P(A ∪ B) = P(A) + P(B).

Addition Rule: The addition rule allows us to calculate the probability of the union of two events. It can be stated as follows:

P(A ∪ B) = P(A) + P(B) — P(A ∩ B)

The rule subtracts the probability of the intersection of the two events to avoid double-counting. If the events are mutually exclusive (i.e., A ∩ B = ∅), then the intersection probability is zero, and the rule simplifies to:

P(A ∪ B) = P(A) + P(B)

Multiplication Rule: The multiplication rule allows us to calculate the probability of the intersection of two events. It can be stated in two forms: the general multiplication rule and the multiplication rule for independent events.

General Multiplication Rule: For any two events A and B, the probability of their intersection is given by:

P(A ∩ B) = P(A) * P(B|A) *,where P(B|A) denotes the conditional probability of event B given that event A has occurred.*

Multiplication Rule for Independent Events: If two events A and B are independent, meaning that the occurrence of one event does not affect the probability of the other event, the multiplication rule simplifies to:

P(A ∩ B) = P(A) * P(B)

This rule states that the probability of the intersection of two independent events is equal to the product of their individual probabilities.

Conditional probability:

Conditional Probability: Conditional probability is the probability of an event A occurring given that another event B has already occurred. It is denoted as **P(A|B)**, read as "*the probability of A given B*." The formula for conditional probability is:

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

In words, the conditional probability of A given B is equal to the probability of both A and B occurring divided by the probability of B occurring. The occurrence of event B serves as new information or a condition that affects the probability of event A. Sometimes $P(A \cap B)$ is also referred as P(A and B) representing the joint probability of events A and B occurring together.

For example, consider drawing two cards successively from a standard deck of 52 playing cards. If we know that the first card drawn is a heart, the probability of drawing a second heart from the remaining deck will be different than if we didn't have this information. This change in probability based on new information is captured by conditional probability.

Bayes' theorem

Bayes' theorem is a fundamental concept in probability theory that allows us to update the probability of an event based on new evidence or information. It provides a framework for incorporating prior knowledge or beliefs into the calculation of probabilities.

Bayes' theorem is named after Reverend Thomas Bayes, an 18th-century mathematician and Presbyterian minister who first formulated the theorem. It is mathematically represented as:

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

In this formula:

- P(A|B) represents the conditional probability of event A occurring given that event B has occurred.
- P(B|A) represents the conditional probability of event B occurring given that event A has occurred.
- P(A) and P(B) represent the probabilities of events A and B occurring independently, without considering each other.

Here's a step-by-step explanation of how to apply Bayes' theorem:

1. ***Start with the prior probability***: Begin with an initial probability, P(A), which represents the prior belief or prior knowledge about the probability of event A occurring.

2. ***Collect new evidence***: Observe or obtain new evidence or information, typically in the form of the conditional probability P(B|A), which represents the probability of observing event B given that event A has occurred.

3. ***Update the probability***: Use Bayes' theorem to update the probability of event A given the new evidence. Multiply the prior probability by the likelihood ratio (P(B|A) / P(B)) and normalize the result.

- ***Multiply***: Multiply the prior probability P(A) by the conditional probability P(B|A). This represents the joint probability of events A and B occurring together.

- ***Divide***: Divide the result by the probability of event B, P(B), which acts as a normalization factor. It ensures that the updated probability is appropriately scaled and sums to 1.

4. ***Interpret the updated probability***: The resulting value, P(A|B), represents the updated probability of event A occurring given the new evidence or information.

Bayes' theorem is widely used in various fields, including statistics, machine learning, medical diagnosis, and decision-making. It enables the incorporation of new information into the analysis, allowing for more accurate and informed probabilistic reasoning. By iteratively applying Bayes' theorem, probabilities can be continuously refined as more evidence becomes available.

A widely popular Machine Learning algorithm Naïve Bayes classifier is based on this theorem. It is a supervised machine learning algorithm, which is used for classification tasks, like text classification.

These are all the fundamental concepts of Probability theory which can get you started with the Statistics and/or Machine Learning. These are also the foundational concepts you need to grasp if you want to learnmore advanced topics in Probability theory.

I tried to explain all the concepts in a very precise manner. I hope that helps you. In case you are stuck at some part or you want me to cover some more advanced topics do let me know in the comment section.

Happy (machine) Learning!!

Probability Theory | Probability | Statistics | Machine Learning | Mathematics

Written by Diptaraj Sen

6 Followers

I love telling stories from Data.

More from Diptaraj Sen