# The Game of Chance: A Blog Post on Probability and Statistics

## -KUNAL SINGH

Probability theory is nothing but common sense reduced to calculation.

That is **Pierre Laplace**, one of the greatest mathematicians who wrote a book, "Analytic theory of probabilities", which laid down many fundamental results in statistics and probability methods.

**Probability** word has a Latin origin- "*probabilities*," which means "likelihood". However, the concept of probability and its mathematical formalization evolved over time.

Probability can be defined as– the measure of how likely an event can occur. Suppose a fair die is thrown- then the probability of occurring an event of number 1 or 2 or any number is 1/6 as all are equally likely to happen. FYI, excavations in the Middle East and in India revealed that dice were already in use at least 14 centuries before Christ. I hope they did know about the equal chances and didn't bid their fortunes.

Key Concepts in Probability:

**Events** and **Outcomes**: An event is something that can happen, like tossing a coin. **Outcomes** are the all-possible results of an event like getting a head or tails.

**Probability distribution**: It describes the likelihood of each outcome in an event. PD ranges from uniform distribution to complex ones. For a fair die to be rolled – the probability distribution will be uniform distribution as all events are equally likely to happen.

To better understand the game of chance we also need to understand - **Statistics.**

In a very layman term – Statistics is mathematics making sense. When you take out maths from the notebooks and apply to the real-world data, suddenly the maths and data, both starts making sense.

Statistics is a discipline which involves collecting, analysing, interpreting and presenting data to uncover patterns, trends, and insights.

Now imagine you have a pile of information- numbers, measurements, observations. With a mere look, they might seem like a jumbled mess but statistics is the toolkit that allows you to organise and understand this information.

Now, let's better understand the terms surrounding statistics.

Statistics involves two main branches:

1. **Descriptive Statistics**: This branch focuses on summarizing and describing data. It uses techniques like measures of central tendency and measures of variability (we will see further these terms) to provide a clear overview of the data's characteristics.

2. **Inferential Statistics**: Once you've grasped the essence of your data, inferential statistics come into play. This branch helps us to draw conclusions about a larger population based on sample of data. It involves various complex techniques like regression analysis which we are not going into detail over here.

**Central Tendency** is a statistical concept that refers to the central value around which a set of data points tend to cluster or concentrate. There are three primary measures of central tendency.

1. **Mean**- It is also known Average. You may be familiar of this term as we use in day-to-day life. The mean is calculated by adding up all the values and then dividing by the no. of values. Here, it needs to be taken in mind that- the mean is sensitive to the outliers, which means that the extreme values can greatly influence the mean.
2. **Median**- The median is the middle value in a dataset when it is arranged in an order of magnitude. Unlike mean, the median is not affected by outliers and gives a better representation of the distribution.
3. **Mode**- The mode is the value that most frequently appears in a dataset. A dataset can have one mode (*unimodal*), multiple modes (*multimodal*) or no mode at all.

**Measures of variability,** also known as measures of dispersion, are statistical concepts that provide information about how dispersed the data points are within a dataset. These measures give you insights into how much the values deviate from the central tendency and how much variation exists in the data. Some common measures of variability are -

1. **Range**- It is the simplest measure of variability. Range is the difference between the maximum and minimum values in the data set.
2. **Variance**- It quantifies how much individual data points deviate from the mean. A higher variance indicates greater variability. Variance is calculated as:
3. **Standard Deviation**- Standard deviation is the square root of the variance. A low SD suggests that data points are closer to mean and vice versa.

**Correlation-** It is a statistical concept that measures the strength and direction of the linear relationship between two variables. It quantifies how closely the values of one variable are associated with the values of another variable. In simple terms- Correlation is used to understand whether and how changes in one variable might be related to the changes in another variable.

Now that we have understood what Probability and Statistics are, we will further learn how these entities help changed the world.

1. **Space Sector: Apollo 11 Moon Landing**
   Probability and statistics played a crucial role in ensuring the success of the mission, particularly during the lunar module's descent to the moon's surface. The onboard computer had to make critical decisions based on sensory data to safely land the module. The team used probabilistic algorithms to interpret sensor data, estimate the module's position and velocity, and make real-time adjustments to the landing trajectory. The use of statistical methods reduced the risks associated with landing in an unfamiliar and challenging environment.

2. **Internet Revolution: Google Search Algorithms**
   The internet revolution brought about by the rise of www created a need for effective search engines. Google developed PageRank algorithm which used Probability and Statistics to revolutionize web search. PageRank assigns importance scores to web pages based on the number and quality of links pointing to them. This approach effectively ranked web pages and improved search accuracy, forming the foundation of Google's search engine. The use of statistical principles to measure the importance of web pages had a profound impact on how we navigate and access information on the internet.

3. **Medicine and Healthcare:**

   Clinical Trials: Randomized controlled trials (RCTs) use probability theory to design experiments, and statistical methods (e.g., hypothesis testing) to analyze the efficacy of new treatments.

   Epidemiology: Probability models, such as the Poisson and logistic regression, are used to study disease patterns, identify risk factors, and predict disease outbreaks.

4. **Manufacturing and Quality Control:**

   Six Sigma: This methodology combines probability and statistical tools to reduce defects and improve manufacturing processes.

   Statistical Process Control (SPC): Control charts and process capability analysis are used to monitor and maintain quality in manufacturing.

5. **Machine Learning and AI:**

   Supervised Learning: Probability and statistics underpin algorithms like linear regression, decision trees, and neural networks.

   Natural Language Processing (NLP): Hidden Markov Models (HMMs) and probabilistic graphical models are used in language modeling and sentiment analysis.

Thus, we see how crucial role Probability and Statistics play in our day-to-day life and hence we should study them better for the betterment of human civilization.

*/KUNAL SINGH*