Artificial Intelligence Diploma

Math for Machine Learning Session 2

NeuroTech

Agenda

1. Probability Introduction
2. Probability Basic's
3. Bay's Theorem
4. Random Variable
5. Probability Distribution
6. Expectation and Variance

Probability

NeuroTech

Introduction to Probability

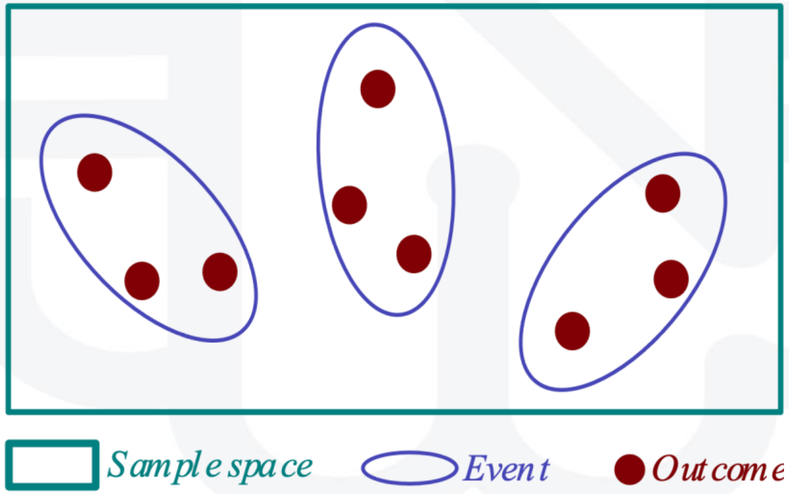
- **Definition of Probability:** Probability is a mathematical concept that quantifies the likelihood of events occurring in uncertain situations. It provides a framework for dealing with uncertainty and randomness.

- **Historical Background:** Probability theory has its roots in games of chance and gambling but has since evolved into a fundamental branch of mathematics with wide-ranging applications.

- **Importance in Real-World Applications:** Probability is used in various fields, including statistics, economics, physics, and machine learning, to make predictions, analyze data, and make informed decisions when faced with uncertainty.

Introduction to Probability

A sample space is the set of all possible outcomes (equally likely) of a probability experiment, typically denoted using set notation. Well-defined sample spaces are a key aspect of of a probabilistic model, along with well-defined event with assigned probabilities. The figure below represents a sample space:



NeuroTech

Introduction to Probability

- Each event has various possible outcomes with distinct probabilities, all of which are contained within the sample space of the experiment.
- A coin toss is an example of a simple experiment. When a coin is tossed, there are two possible outcomes: heads or tails. The sample space for the experiment of tossing a coin is {heads, tails}, or {H, T}.

- If there were 3 coins, and order were being considered, there would be 8 events in the ordered sample space: {HHH, HHT, HTH, HTT, TTT, TTH, THT, THH}. If order were not considered, then there would be 6 events in the unordered sample space would be {HHH, HHT, HTT, TTT, TTH, THH}. Each of the subsets of the sample spaces above is an event. For example, "HHH" is an event in either the ordered or unordered sample spaces above.

Introduction to Probability

- Probability is a measure of the likelihood of an event to occur. Many events cannot be predicted with total certainty. We can predict only the chance of an event to occur i.e., how likely they are going to happen, using it. Probability can range from 0 to 1, where 0 means the event to be an impossible one and 1 indicates a certain event. Probability for Class 10 is an important topic for the students which explains all the basic concepts of this topic. **The probability of all the events in a sample space adds up to 1.**

- **For example**, when we toss a coin, either we get Head OR Tail, only two possible outcomes are possible (H, T). But when two coins are tossed then there will be four possible outcomes, i.e {(H, H), (H, T), (T, H), (T, T)}.

Introduction to Probability

- **Formula for Probability**
    - Probability of event to happen P(E) = Number of favourable outcomes/Total Number of outcomes

- **Example:**

1) **There are 6 pillows in a bed, 3 are red, 2 are yellow and 1 is blue. What is the probability of picking a yellow pillow?**

    Ans: The probability is equal to the number of yellow pillows in the bed divided by the total number of pillows, i.e. 2/6 = 1/3.

Interception

- The notation $P(A \cap B)$ represents the probability of the intersection of two events, A and B. In other words, it signifies the likelihood that both events A and B will occur simultaneously.

- The formula for $P(A \cap B)$ is:
- $P(A \cap B)$=Probability that both A and B occur.

- For example, let's say we have two events:
  - Event A: Rolling a 6 on a fair six-sided die.
  - Event B: Drawing a red card from a standard deck of playing cards.

- If we want to find $P(A \cap B)$, we are calculating the probability that both event A (rolling a 6) and event B (drawing a red card) happen together. To do this, you would multiply the probability of A (1/6, since there is one favorable outcome out of six possibilities for rolling a 6) by the probability of B (26/52, since there are 26 red cards out of 52 cards in a standard deck).
- So, $P(A \cap B)=(1/6)*(26/52)=1/12$.
- In this case, $P(A \cap B)$ represents the probability of rolling a 6 on the die and drawing a red card from the deck on the same trial.

Conditional Probability

- Conditional probability is a concept in probability theory that quantifies the likelihood of one event occurring given that another event has already occurred. It represents the probability of an event A happening under the condition or assumption that event B has taken place. Conditional probability is denoted as P(A | B), which is read as "the probability of event A given event B.”

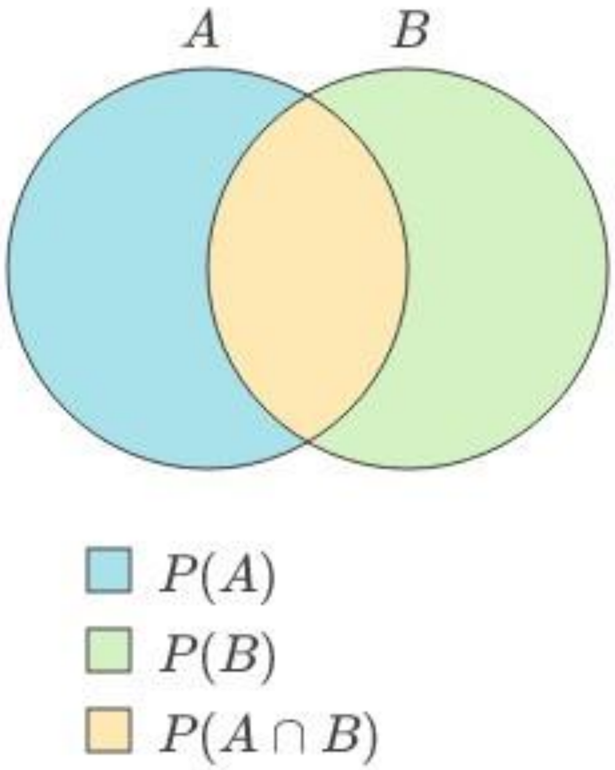
- The formula for conditional probability is:

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

- Where:

  - $P(A|B)$ is the conditional probability of event A given event B.
  - $P(A \cap B)$ is the probability of both events A and B occurring together.
  - $P(B)$ is the probability of event B occurring.

Conditional Probability



Conditional Probability Formula

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Probability that A occurs given that B has already occured

Bayes' Theorem

NeuroTech

Bayes' Theorem

- To derive Bayes' Theorem, we can start from the definition of conditional probability and use basic probability rules. Here's how to derive the equation step by step:

- **Step 1: Definition of Conditional Probability**
  - The conditional probability of event A given event B, denoted as $P(A|B)$, is defined as:

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

  - This equation represents the probability that event A occurs given that event B has occurred.

- **Step 2: Rearrange the Equation**
  - Rearrange the definition of conditional probability to isolate $P(A \cap B)$:

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

NEUROTECH

Bayes' Theorem

- **Step 3: Apply Symmetry of Intersection**
  - The probability of the intersection of two events, $P(A \cap B)$, is the same as the probability of the intersection of B and A, so we can write:

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

- **Step 4: Equate the Two Expressions**
  - Now, equate the two expressions for $P(A \cap B)$ from Steps 2 and 3:

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

Bayes' Theorem

- **Step 5: Solve for *P*(*A*|*B*)**
  - To isolate $P(A|B)$, divide both sides of the equation by $P(B)$:

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

- And there you have it! This is Bayes' Theorem. It expresses the conditional probability of event A given event B in terms of the conditional probability of event B given event A, the prior probability of event A, and the prior probability of event B.

- In practical applications, Bayes' Theorem allows you to update your beliefs (represented by the prior probabilities) based on new evidence (represented by the conditional probabilities) to calculate the updated probabilities.

Bayes' Theorem

- Bayes' Theorem is a fundamental concept in probability theory and statistics that allows us to update our beliefs about the probability of an event based on new evidence or information. It provides a formal way to calculate conditional probabilities and is particularly useful in situations where we have prior knowledge and want to incorporate new data to make more informed decisions. Bayes' Theorem is named after Thomas Bayes, an 18th-century mathematician and statistician.

- The general form of Bayes' Theorem is expressed as follows:

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

- Where:
  - $P(A|B)$ is the conditional probability of event A occurring given that event B has occurred.
  - $P(B|A)$ is the conditional probability of event B occurring given that event A has occurred.
  - $P(A)$ is the prior probability of event A.
  - $P(B)$ is the prior probability of event B.

Bayes' Theorem

Let's break down each component of Bayes' Theorem and understand its significance:

1. ***P(A|B)*:** This is the updated or posterior probability of event A occurring, given that we have observed event B. It represents our new belief about the probability of A based on the evidence provided by B.

2. ***P(B|A)*:** This is the probability of observing event B, given that we already know event A has occurred. It represents the likelihood of our evidence (B) occurring if our hypothesis (A) is true.

3. ***P(A)*:** This is the prior probability of event A. It represents our initial belief about the probability of A before considering any new evidence (B). It's often referred to as the "prior belief" or "prior probability."

4. ***P(B)*:** This is the prior probability of event B. It represents our initial belief about the probability of B before considering any information related to A. It's often referred to as the "prior belief" or "prior probability."

Bayes' Theorem

The process of applying Bayes' Theorem typically involves the following steps:

1. Establish Prior Beliefs: Determine the prior probabilities $P(A)$ and $P(B)$ based on your initial knowledge or beliefs about the events A and B.

2. Gather New Evidence: Obtain new evidence or data related to the events A and B. This evidence is represented by the conditional probability $P(B|A)$.

3. Calculate the Updated Probability: Use Bayes' Theorem to calculate the updated or posterior probability $P(A|B)$ based on the prior probabilities and the new evidence.

4. Interpret the Result: The updated probability $P(A|B)$ represents your revised belief about event A given the new evidence B. It can be used for decision-making or drawing conclusions.

Bayes' Theorem

Applications of bay's theorem:

1. Medical Diagnosis:

1. Bayes' Theorem is widely used in medical diagnosis to assess the probability of a disease given certain symptoms and test results. It helps doctors make informed decisions about treatment and patient care.

2. Spam Filtering:

1. Email spam filters use Bayes' Theorem to classify incoming emails as spam or not spam based on the probability of certain words or phrases occurring in legitimate emails versus spam messages.

3. Natural Language Processing (NLP):

1. In NLP, Bayes' Theorem is used for tasks like text classification, sentiment analysis, and language modeling. It helps machines make predictions based on the probabilities of word sequences or patterns.

Bayes' Theorem

Applications of bay's theorem:

4. Finance and Investment:

1. Bayes' Theorem is applied in finance for risk assessment, portfolio optimization, and investment decisions. It helps in estimating the probabilities of various financial events and market movements.

5. Machine Learning and AI:

1. Bayesian methods are used in machine learning for probabilistic modeling, Bayesian networks, and Bayesian inference. It allows for uncertainty quantification and decision-making in complex systems.

6. Weather Forecasting:

1. Bayes' Theorem is used in meteorology to update weather predictions based on new observational data. It helps improve the accuracy of short-term and long-term weather forecasts.

Random Variable

NeuroTech

Random Variable

- A random variable is a fundamental concept in probability theory and statistics. It serves as a mathematical representation of uncertain or random phenomena, such as the outcome of a random experiment or a measurement in a probabilistic context. In essence, a random variable is a variable whose possible values are outcomes of a random process, and it associates a probability with each of these values.

- **Random Variable Definition:**

  - A random variable is a function that assigns a real number to each possible outcome (event) of a random experiment. It provides a way to quantify the uncertainty associated with various outcomes.
  - Random variables can be categorized into two main types: discrete random variables and continuous random variables.

Random Variable

1. **Discrete Random Variable:**

    - **Definition:** A discrete random variable is a random variable that takes on a countable number of distinct and separate values. These values are often whole numbers or integers.

    - **Examples:** The number of heads obtained when flipping a coin multiple times, the number of defective items in a batch of products, and the number of people in a queue are all examples of discrete random variables.

    - **Probability Distribution:** Discrete random variables are associated with probability mass functions (PMFs), which describe the probability of each possible value. The PMF provides a discrete probability distribution.

Random Variable

2. Continuous Random Variable:

- **Definition:** A continuous random variable is a random variable that can take on an uncountable number of possible values within a given range. These values are often real numbers.

- **Examples:** The height of a person, the time it takes for an event to occur, and the temperature at a specific location are examples of continuous random variables.

- **Probability Distribution:** Continuous random variables are associated with probability density functions (PDFs), which describe the probability density at various points within a range. The PDF provides a continuous probability distribution.

Random Variable

Key Differences Between Discrete and Continuous Random Variables:

1. **Nature of Values:**
   1. Discrete Random Variable: Takes on distinct, separate values.
   2. Continuous Random Variable: Takes on a continuous range of values.
2. **Probability Distribution:**
   1. Discrete Random Variable: Associated with a probability mass function (PMF).
   2. Continuous Random Variable: Associated with a probability density function (PDF).
3. **Examples:**
   1. Discrete Random Variable Examples: Counting events (e.g., number of cars passing by), integer-valued measurements (e.g., number of customer arrivals), and outcomes involving a finite set of possibilities (e.g., rolling a die).
   2. Continuous Random Variable Examples: Measurements with infinite possibilities (e.g., weight, time, temperature), physical characteristics, and real-valued outcomes (e.g., height).
4. **Notation:**
   1. Discrete Random Variable: Often denoted using uppercase letters (e.g., X, Y).
   2. Continuous Random Variable: Also denoted using uppercase letters (e.g., X, Y).

Probability Distribution

NeuroTech

Discrete Probability Distribution

NeuroTech

Discrete Probability Distribution

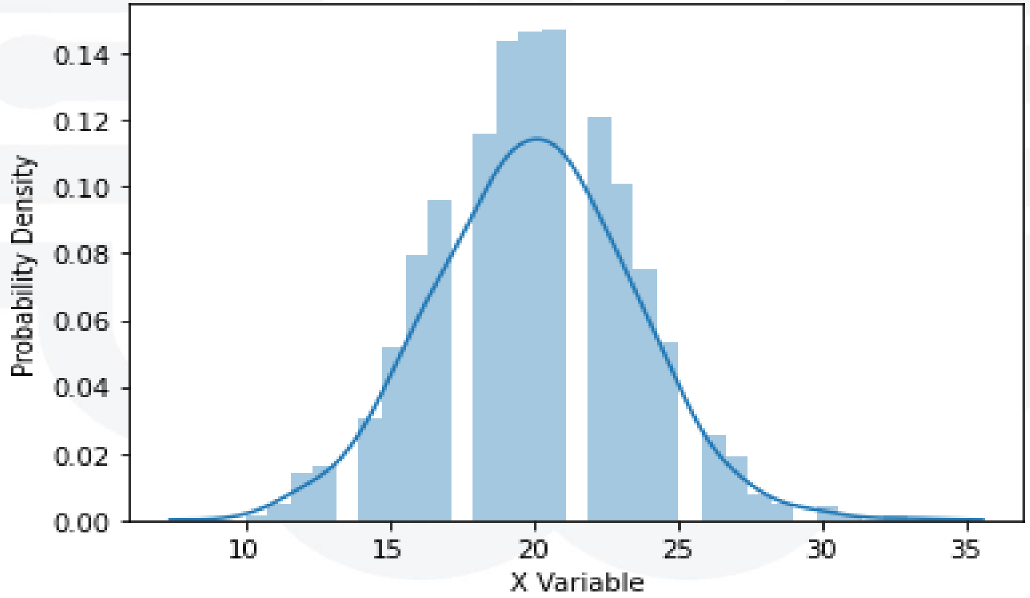
What is Discrete Probability Distributions?

- Discrete probability distributions represent the likelihood of different outcomes in a discrete set, such as the results of rolling a dice or the number of successes in a fixed number of trials. Each outcome is associated with a probability, and when graphed, these probabilities create a distribution. Common examples include the binomial distribution for binary events and the Poisson distribution for rare events. Such distributions are essential in statistics and probability theory for modeling and analyzing discrete random variables.

- Discrete probability distributions are graphs of the outcomes of test results, such as a value of 1, 2, 3, true, false, success, or failure. Investors use discrete probability distributions to estimate the chances that a particular investing outcome is more or less likely to happen.

Discrete Probability Distribution

Let X be a random variable that has more than one possible outcome. Plot the probability on the y-axis and the outcome on the x-axis. If we repeat the experiment many times and plot the probability of each possible outcome, we get a plot that represents the probabilities. This plot is called the probability distribution (PD). The height of the graph for X gives the probability of that outcome.



Discrete Probability Distribution

Discrete Probability Distributions

- There are many discrete probability distributions to be used in different scenarios. We will discuss Discrete distributions in this post. Binomial and Poisson distributions are the most discussed ones in the following list.

  1. Bernoulli Distribution
  2. Binomial Distribution
  3. Hypergeometric Distribution
  4. Geometric Distribution
  5. Poisson Distribution
  6. Multinomial Distribution

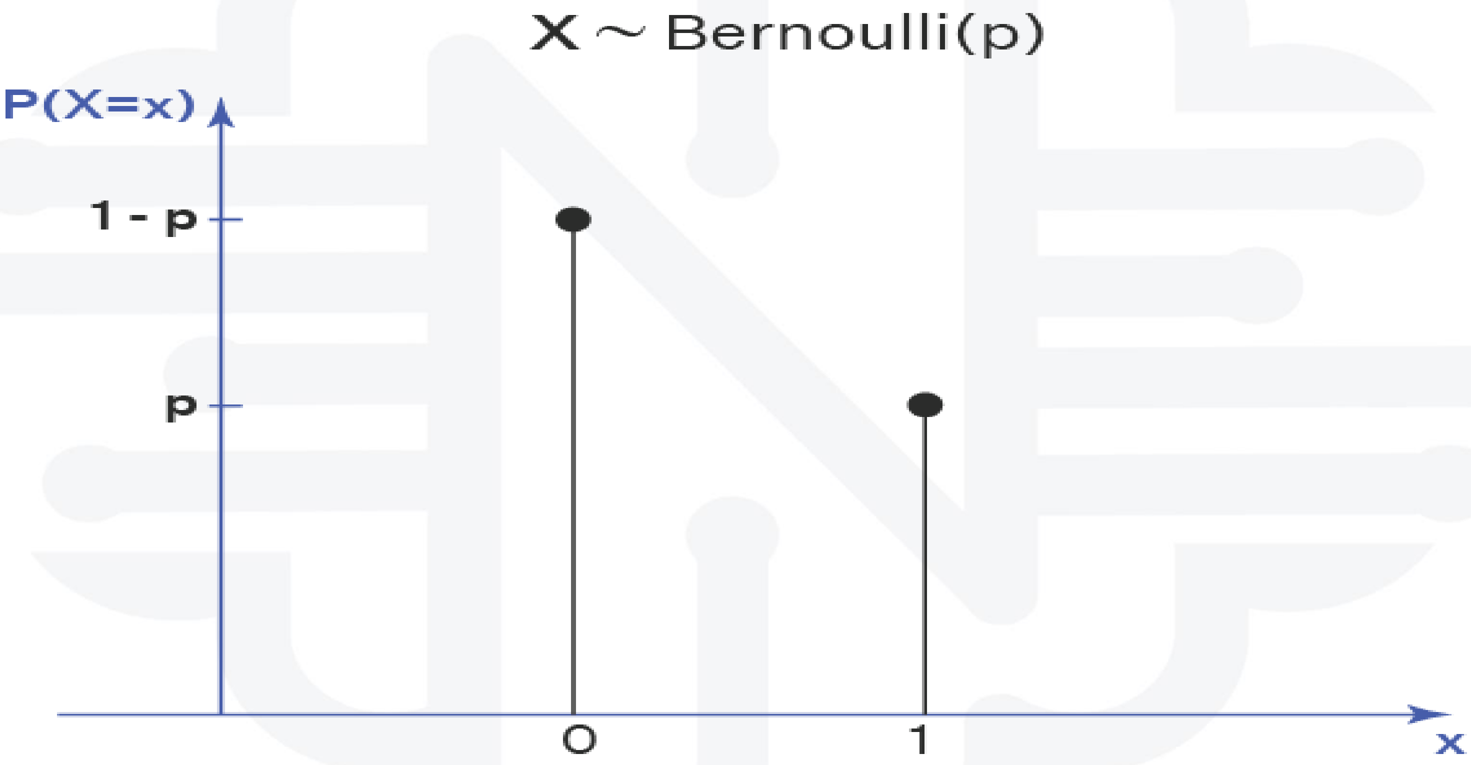
Bernoulli Distribution

- This distribution is generated when we perform an experiment once and it has only two possible outcomes – success and failure. The trials of this type are called Bernoulli trials, which form the basis for many distributions discussed below. Let p be the probability of success and 1 – p is the probability of failure.
- The PMF is given as:

$$\text{PMF} = \begin{cases} p, & \text{Success} \\ 1 - p, & \text{Failure} \end{cases}$$

- One example of this would be flipping a coin once. p is the probability of getting ahead and 1 – p is the probability of getting a tail. Please note down that success and failure are subjective and are defined by us depending on the context.

Bernoulli Distribution



Binomial Probability Distribution

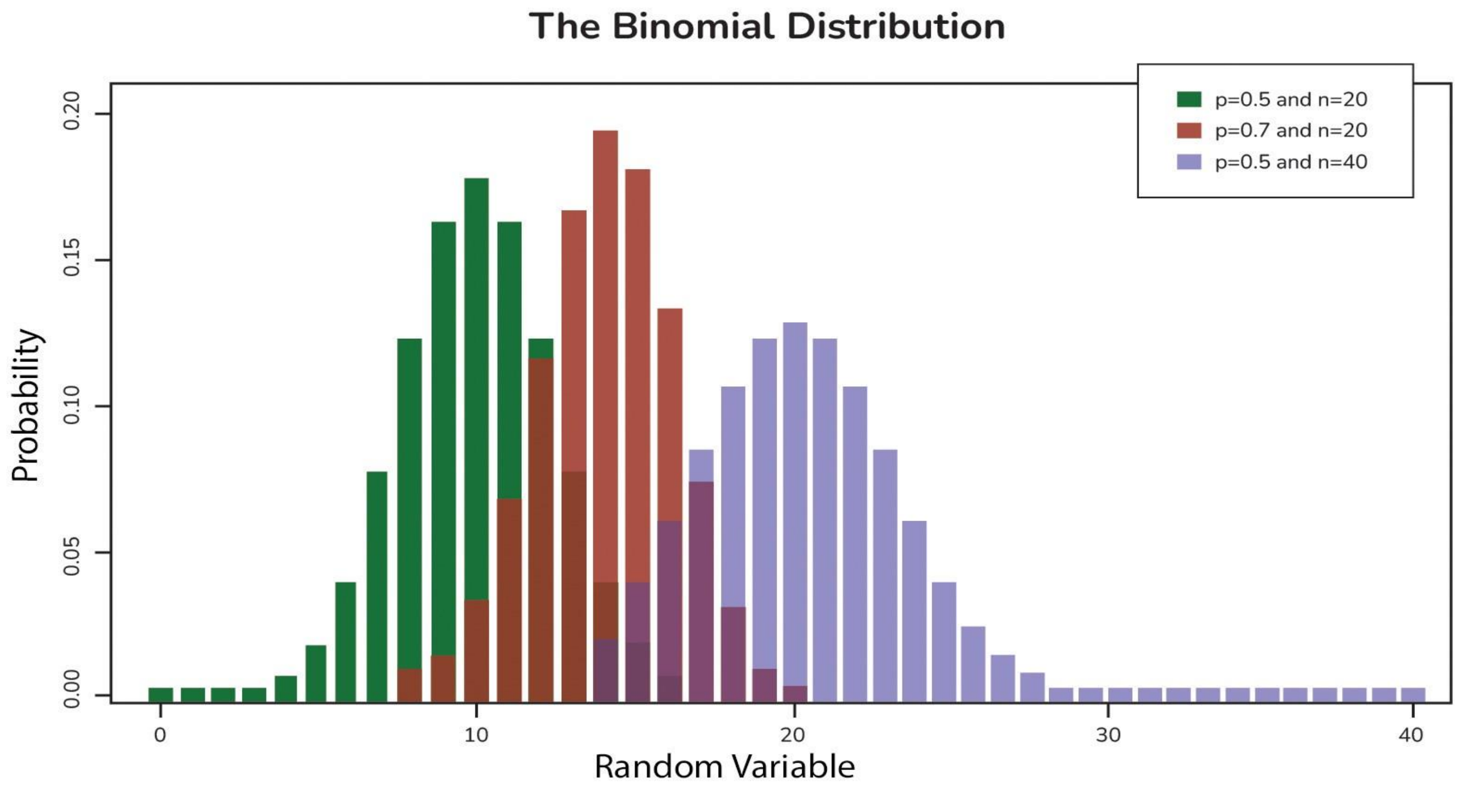
- This is generated for random variables with only two possible outcomes. Let p denote the probability of an event is a success which implies 1 – p is the probability of the event being a failure. Performing the experiment repeatedly and plotting the probability each time gives us the Binomial distribution.
- The most common example given for Binomial distribution is that of flipping a coin n number of times and calculating the probabilities of getting a particular number of heads. More real-world examples include the number of successful sales calls for a company or whether a drug works for a disease or not.
- The PMF is given as,

$$^{n}C_{x}p^{x}(1-p)^{x}$$

- where p is the probability of success, n is the number of trials and x is the number of times we obtain a success.

NeuroTech

Binomial Probability Distribution



NEUROTECH

Hypergeometric Probability Distribution

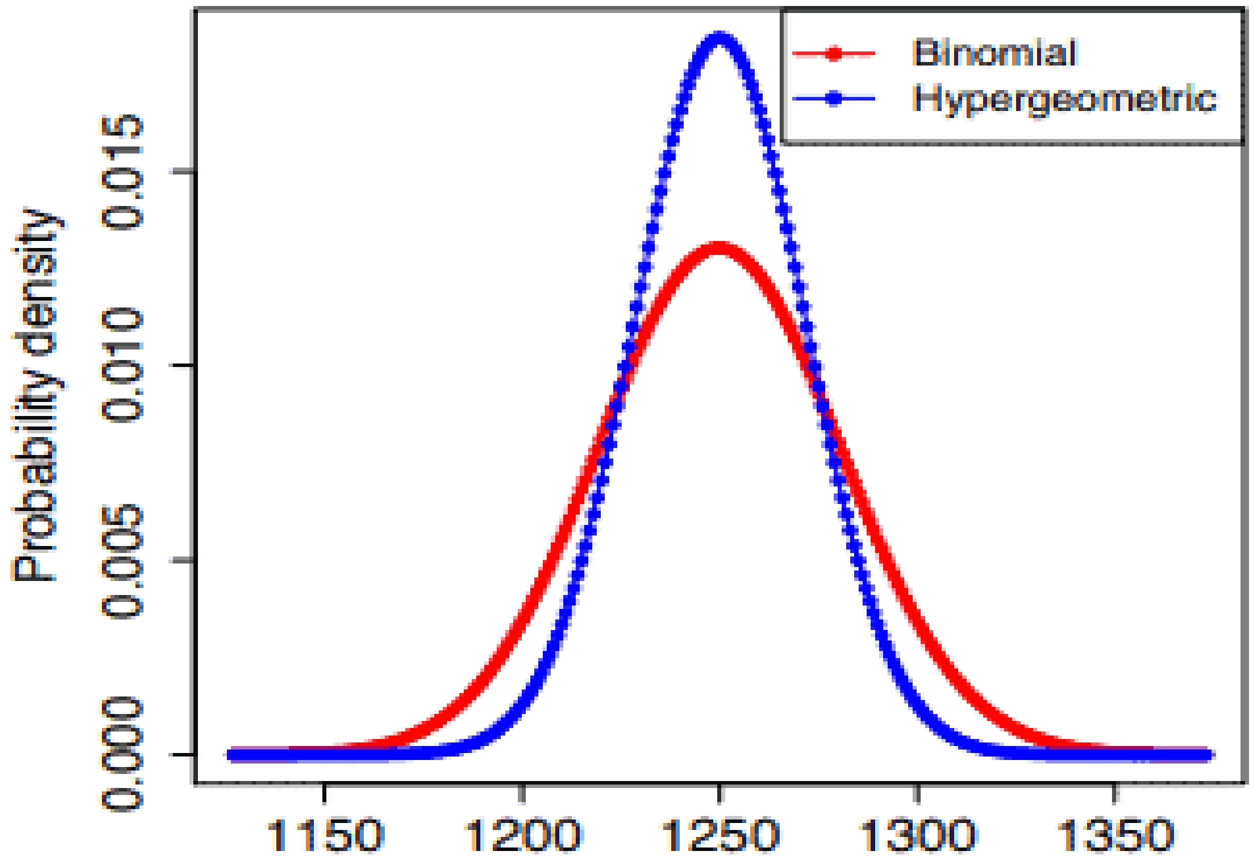
- Consider an event of drawing a red marble from a box of marbles with different colors. The event of drawing a red ball is a success and not drawing it is a failure. But each time a marble is drawn it is not returned to the box and hence this affects the probability of drawing a ball in the next trial. The hypergeometric distribution models the probability of k successes over n trials where each trial is conducted without replacement. This is unlike the binomial distribution where the probability remains constant through the trials.

- The PMF is given as,

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

- where k is the number of possible successes, x is the desired number of successes, N is the size of the population and n is the number of trials.

Hypergeometric Probability Distribution



NEUROTECH

Geometric Probability Distribution

- This is a special case of the negative binomial distribution where the desired number of successes is 1. It measures the number of failures we get before one success. Using the same example given in the previous section, we would like to know the number of failures we see before we get the first 4 on rolling the dice.

$$P(X = k) = (1 - p)^{k-1}\, p$$

- where p is the probability of success and k is the number of failures. Here, r = 1.

Poisson Probability Distribution

- This distribution describes the events that occur in a fixed interval of time or space. An example might make this clear. Consider the case of the number of calls received by a customer care center per hour. We can estimate the average number of calls per hour but we cannot determine the exact number and the exact time at which there is a call. Each occurrence of an event is independent of the other occurrences.

- The PMF is given as,

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

- where λ is the average number of times the event has occurred in a certain period of time, x is the desired outcome and e is the Euler's number.

Multinomial Probability Distribution

- In the Previous distributions, there are only two possible outcomes – success and failure. The multinomial distribution, however, describes the random variables with many possible outcomes. This is also sometimes referred to as categorical distribution as each possible outcome is treated as a separate category. Consider the scenario of playing a game n number of times. Multinomial distribution helps us to determine the combined probability that player 1 will win x_1 times, player 2 will win x_2 times and player k wins x_k times.

- The PMF is given as,

$$P(X = x_1, X = x_2, \dots . X = x_k) = \frac{\text{n!}}{x_1!\, x_2! \dots . x_k!} {p_1}^{x_1} {p_2}^{x_2} \dots . {p_k}^{x_k}$$

- where n is the number of trials, $p_{1,} \dots\dots p_k$ denote the probabilities of the outcomes $x_1 \dots\dots x_k$ respectively.

NEUROTECH

Continuous Probability Distribution

NeuroTech

Continuous Probability Distribution

What is Continuous Probability Distributions?

- If you recall from our previous discussion, continuous random variables can take an infinite number of values over a given interval. For example, in the interval [2, 3] there are infinite values between 2 and 3. Continuous distributions are defined by the Probability Density Functions(PDF) instead of Probability Mass Functions. The probability that a continuous random variable is equal to an exact value is always equal to zero. Continuous probabilities are defined over an interval. For instance, $P(X = 3) = 0$ but $P(2.99 < X < 3.01)$ can be calculated by integrating the PDF over the interval [2.99, 3.01].

Continuous Probability Distribution

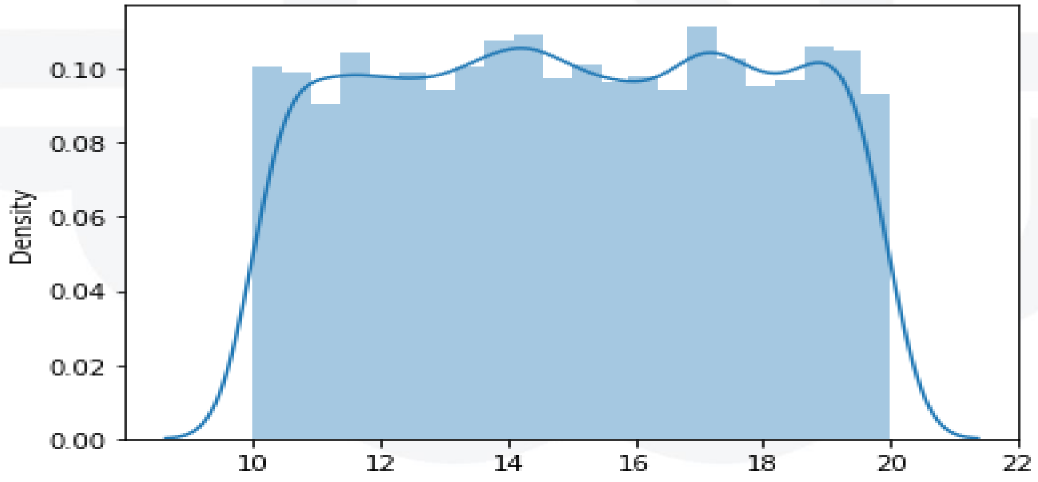
Continuous Probability Distributions:

1. Continuous Uniform Distribution
2. Normal Distribution
3. Log-normal Distribution
4. Student's T Distribution
5. Chi-square Distribution
6. Exponential Distribution

Continuous Uniform Distribution

- Uniform distribution has both continuous and discrete forms. Here, we discuss the continuous one. This distribution plots the random variables whose values have equal probabilities of occurring. The most common example is flipping a fair die. Here, all 6 outcomes are equally likely to happen. Hence, the probability is constant.

- Consider the example where a = 10 and b = 20, the distribution looks like this:



NeuroTech

Continuous Uniform Distribution

- The PDF is given by,

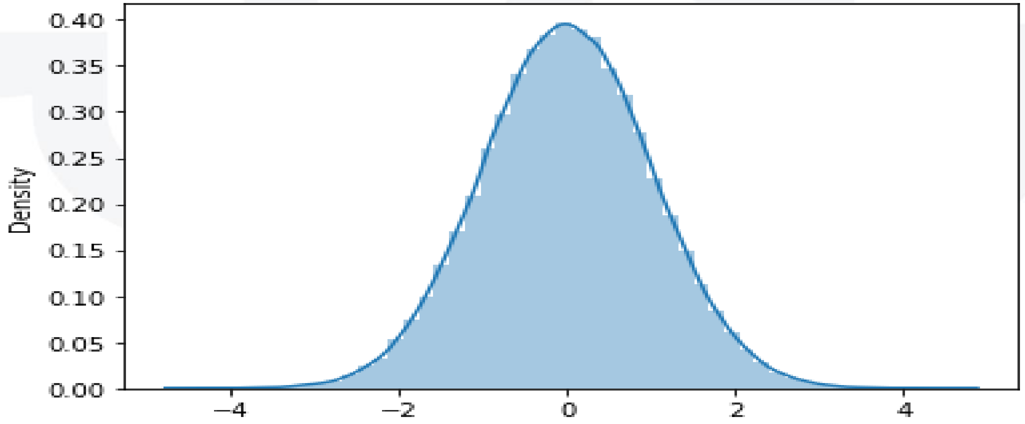
$$f(x) = \begin{cases} \dfrac{1}{b - a}, & x \in [a, b] \\ 0, & otherwise \end{cases}$$

- where a is the minimum value and b is the maximum value.

Normal Distribution

- This is the most commonly discussed distribution and most ojen found in the real world. Many continuous distributions ojen reach normal distribution given a large enough sample. This has two parameters namely mean and standard deviation.

- This distribution has many interesting properties. The mean has the highest probability and all other values are distributed equally on either side of the mean in a symmetric fashion. The standard normal distribution is a special case where the mean is 0 and the standard deviation of 1.



Normal Distribution

- It also follows the empirical formula that 68% of the values are 1 standard deviation away, 95% percent of them are 2 standard deviations away, and 99.7% are 3 standard deviations away from the mean.

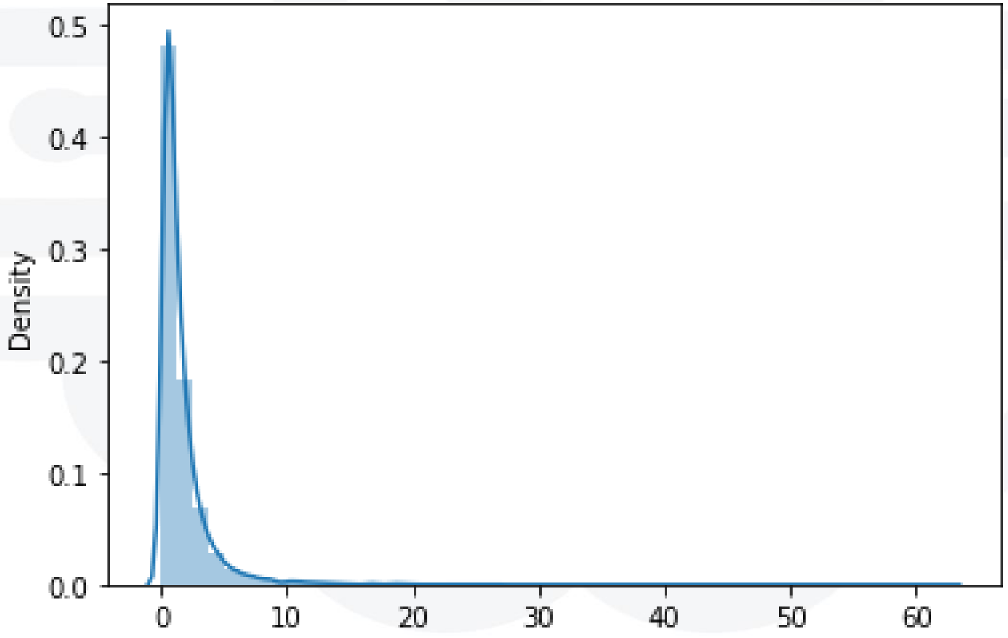
- The PDF is given by,

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

where μ is the mean of the random variable X and σ is the standard deviation.

Long normal Distribution

- This distribution is used to plot the random variables whose logarithm values follow a normal distribution. Consider the random variables X and Y. Y = ln(X) is the variable that is represented in this distribution, where ln denotes the natural logarithm of values of X.



Long normal Distribution

- The PDF is given by,

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma} exp\left(-\frac{1}{2}\left(\frac{lnx - \mu}{\sigma}\right)^2\right)$$

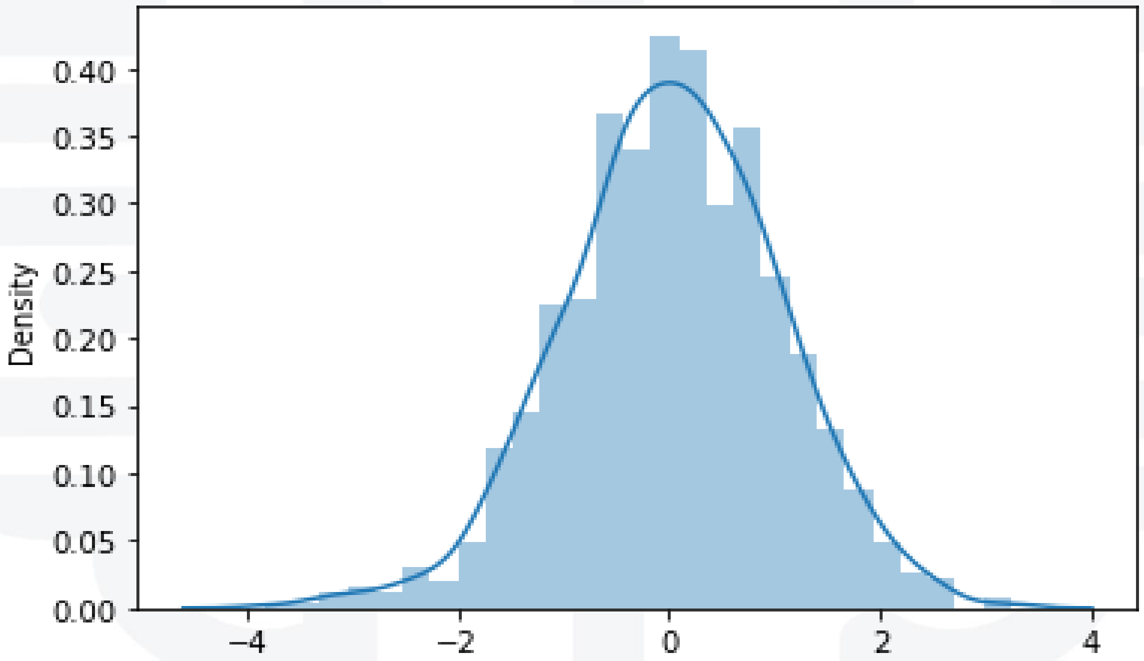
- where μ is the mean of Y and σ is the standard deviation of Y.

Student's T Distribution

- The student's t distribution is similar to the normal distribution. The difference is that the tails of the distribution are thicker. This is used when the sample size is small and the population variance is not known. This distribution is defined by the degrees of freedom(p) which is calculated as the sample size minus 1(n – 1).

- As the sample size increases, degrees of freedom increases the t-distribution approaches the normal distribution and the tails become narrower and the curve gets closer to the mean. This distribution is used to test estimates of the population mean when the sample size is less than 30 and population variance is unknown. The sample variance/standard deviation is used to calculate the t-value.

Student’s T Distribution



Student's T Distribution

- The PDF is given by,

$$f(t) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{p\pi}\,\Gamma\left(\frac{p}{2}\right)}\left(1 + \frac{t^2}{p}\right)^{-\left(\frac{p+1}{2}\right)}$$

- where p is the degrees of freedom and Γ is the gamma function. Check this link for a brief description of the gamma function.

- The t-statistic used in hypothesis testing is calculated as follows,

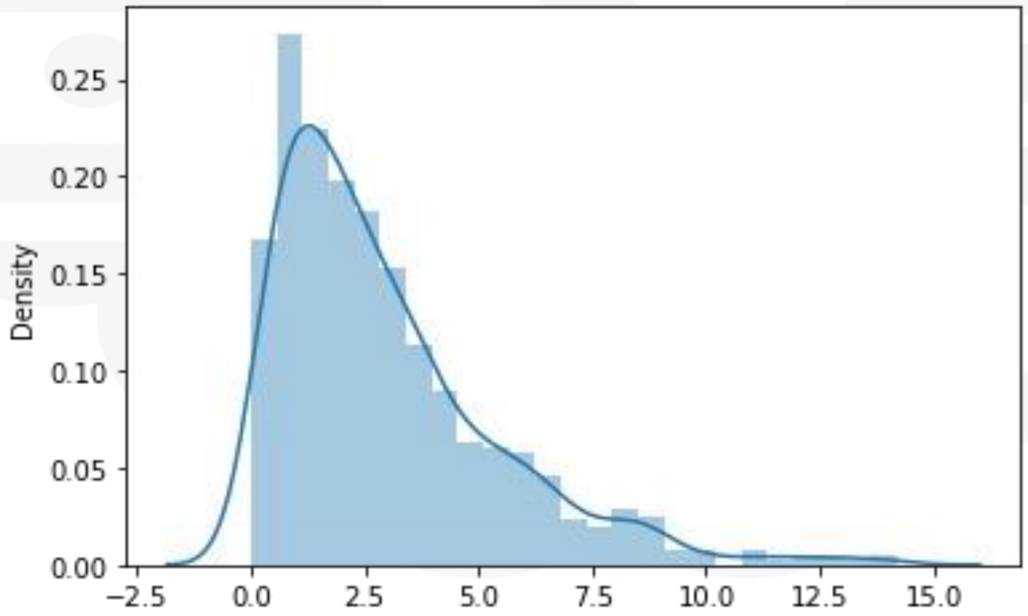
$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

- where x̄ is the sample mean, μ the population mean and s is the sample variance.

NeuroTech

Chi-square Distribution

- This distribution is equal to the sum of squares of p normal random variables. p is the number of degrees of freedom. Like the t-distribution, as the degrees of freedom increase, the distribution gradually approaches the normal distribution. Below is a chi-square distribution with three degrees of freedom.



Chi-square Distribution

The PDF is given by,

$$f(x) = \frac{\left(x^{\frac{p}{2}-1} e^{-\frac{x}{2}}\right)}{2^{p/2}\, \Gamma\left(\frac{p}{2}\right)}$$

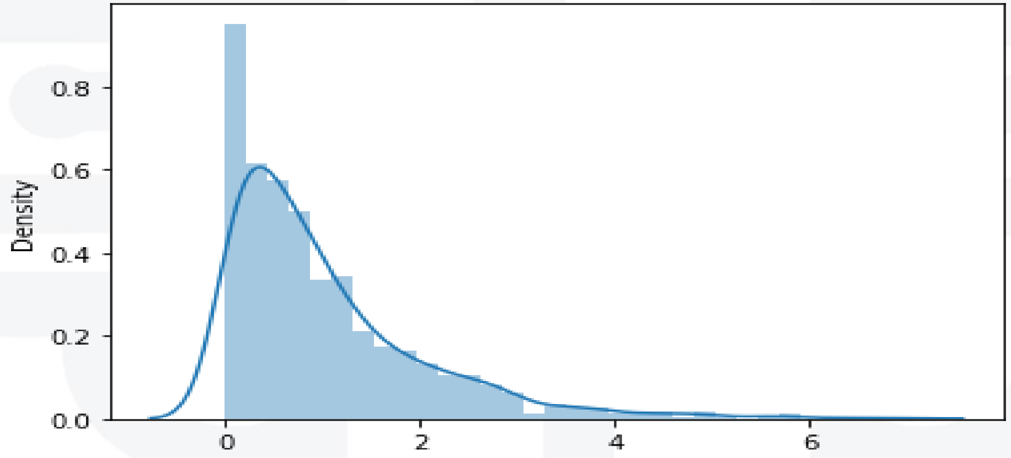
where p is the degrees of freedom and Γ is the gamma function.
The chi-square value is calculated as follows:

$$\chi^2 = \sum \frac{(o_i - E_i)^2}{E_i}$$

where o is the observed value and E represents the expected value. This is used in hypothesis testing to draw inferences about the population variance of normal distributions.

Exponential Distribution

- Recall the discrete probability distribution we have discussed in the Discrete Probability post. In the Poisson distribution, we took the example of calls received by the customer care center. In that example, we considered the average number of calls per hour. Now, in this distribution, the time between successive calls is explained.



- The exponential distribution can be seen as an inverse of the Poisson distribution. The events in consideration are independent of each other.

Exponential Distribution

- The PDF is given by,

$$f(x) = \lambda e^{-\lambda x}$$

- where λ is the rate parameter. λ = 1/(average time between events).
- To conclude, we have very briefly discussed different continuous probability distributions in this post. Feel free to add any comments or suggestions below.

Expectation and Variance

Expectation (Expected Value):

- Expectation, denoted as E[X], represents the long-term average or mean value of a random variable X. It is a fundamental concept in probability and statistics and provides insight into the central or typical value of the random variable. The expectation is a way to summarize the behavior of the random variable over many trials or observations.

For a Discrete Random Variable X:

The expectation of a discrete random variable X is calculated as follows:

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

Where:

- *E*[*X*] is the expected value of X.
- *x* represents each possible value of the random variable X.
- *P*(*X*=*x*) is the probability of X taking on the value *x*.

The formula involves multiplying each possible value *x* by its corresponding probability and summing up these products. The result is the expected value, which is a single number representing the mean of the random variable

Expectation (Expected Value):

For a Continuous Random Variable X:

The expectation of a continuous random variable X is calculated as follows:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Where:

- $E[X]$ is the expected value of X.
- x represents the variable itself.
- $f(x)$ is the probability density function (PDF) of the continuous random variable.

In this case, the integral represents the weighted average of the variable *x*, with the weight given by the PDF. The result is the expected value of X.

Interpretation:

Expectation represents the average value you would expect to obtain when you repeatedly sample or observe the random variable X over a large number of trials or observations. It provides a measure of the "center" or "typical" value of the random variable.

Variance:

- Variance, denoted as Var(X), measures the spread or variability of the values of a random variable X around its expected value. It quantifies how much individual values deviate from the mean and provides a measure of dispersion.

- **For a Discrete Random Variable X:**
    - The variance of a discrete random variable X is calculated as follows:

$$Var(X) = E[(X - E[X])^2] = \sum_x (x - E[X])^2 \cdot P(X = x)$$

Where:

- $Var(X)$ is the variance of X.
- $E[X]$ is the expected value of X.
- x represents each possible value of X.
- $P(X=x)$ is the probability of X taking on the value x.

Variance:

For a Continuous Random Variable X:

The variance of a continuous random variable X is calculated similarly:

$$Var(X) = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 \cdot f(x)\, dx$$

Where:

- *Var*(*X*) is the variance of X.
- *E*[*X*] is the expected value of X.
- *x* represents the variable itself.
- *f*(*x*) is the probability density function (PDF) of X.

The integral calculates the weighted average of the squared differences between the variable *x* and its expected value, using the PDF as the weighting function.

Variance:

- **Interpretation:**
  - Variance quantifies the amount of dispersion or spread in the values of the random variable. A higher variance indicates greater variability or spread, while a lower variance indicates less variability around the mean. Variance is a key measure of the degree of uncertainty or risk associated with a random variable.

- In summary, expectation (expected value) provides the central or typical value of a random variable, while variance measures the spread or variability of the values around that central value. Together, they are important tools for describing and understanding the behavior of random variables and probability distributions in statistics and probability theory.

Any Question

NeuroTech

Thanks

NeuroTech