

Chapter 2: Distributions

Distribution is the function that describes the relationship between observations in a sample. In other words, a distribution is simply a collection of data, or scores, on a variable. Usually, these scores are arranged in order from smallest to largest and then they can be presented graphically.

Density functions are functions that describe how the proportion of data or likelihood of the proportion of observations changes over the range of the distribution. There are two types of Density functions, which are:

- Probability Density function: calculates the probability of observing a given value.
- Cumulative Density function: calculates the probability of an observation equal or less than a value.

Gaussian distribution/Normal distribution

It is named after Carl Friedrich Gauss, which is also known as Normal distribution.

In the normal distribution, the mean of the data is equal to median. When the data is plotted into the graph, we will get a bell-shaped curve where in the values are clustered at the central region of the x-axis. The variables which include about the physical world, human society or social relationship are normally or approximately normally distributed. Examples: height, birth rate, job satisfaction or scores. With the help of understanding the properties of the normal distribution, we can compare different groups and make estimates about population by using the samples.

Properties of normal distribution

- The means, median and mode are the same.
- The distribution is symmetric about the mean. (bell-shaped curve)

With the help of Scipy module, we can work with Gaussian distribution. The `norm.pdf()` function can be used to create a Gaussian probability density function with a given sample, mean and standard deviation.

```
pdf(sample, mean, standard deviation)
```

`cdf()` function can be used to create cumulative density function with a given sample.

```
cdf(sample)
```

Bernoulli's Distribution

Only one trial will be executed in which you will get two outcomes which are success or failure. Hence, we will be calculating the probability of 1 or the probability of 0.

For example, there are two candidates for the selection of student representative who are James and John. Let's say, the probability of success $p(1) = 0.6$ for James, then the probability of failure $p(0) = 1 - 0.6 = 0.4$.

Probability mass function (pmf):

A probability mass function of a discrete random variable X assigns probabilities to each of the possible values of the random variable. By using PMF, we can get the probabilities of each random variable.

Binomial Distribution

A binomial distribution is one of the types of distribution that is two possible outcomes success or failure which is repeated multiple outcomes.

For example, let's say you are rolling a dice to know the probability of getting 1. Now, if we roll the dice 30 times then the probability of rolling the dice is $1/6$. Hence, you have a binomial distribution of $(n=30, p=1/6)$.

Criteria:

Binomial distribution must meet the following criteria:

1. The number of observations or trials is fixed.
2. Each observation or trial is independent.
3. The probability of success is exactly the same from one trial to another.

Uniform Distribution

Uniform distribution is the type of probability distribution in which all outcomes are equally likely. For example, flipping a coin has a uniform distribution as the probability; or withdrawing a card from a deck which have the probability of spade, a club, a diamond or a heart.

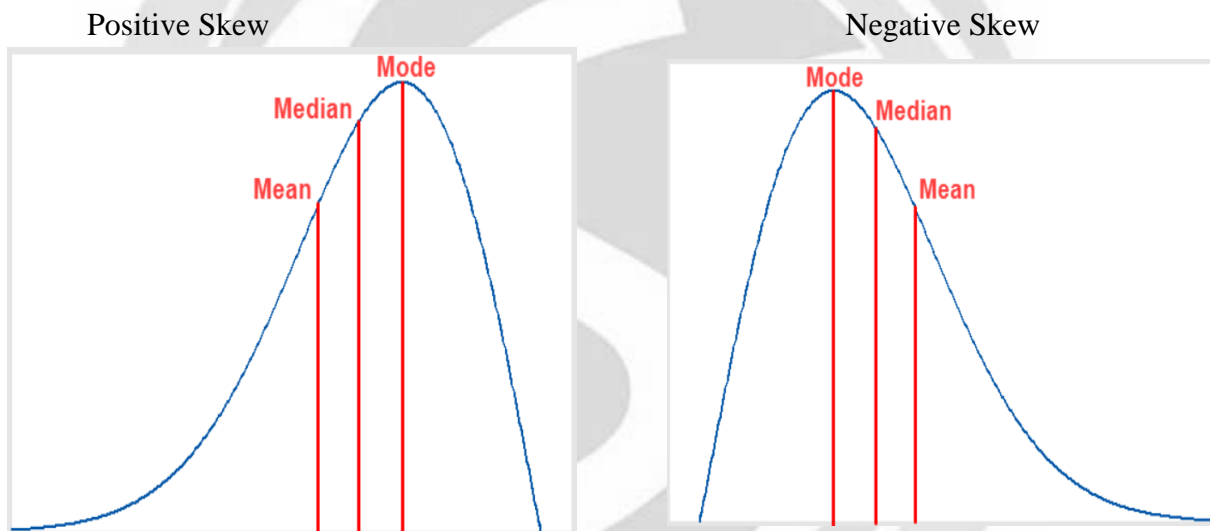
The uniform distribution can be visualized and we always get a straight horizontal line. In a discrete uniform distribution, outcomes are discrete and have the same probability. In a continuous uniform distribution, outcomes are continuous and infinite.

Skewed Distribution skewness

Skewness is defining the asymmetry of a distribution.

There are three types of skewness:

1. A **left-skewed distribution** has a long left tail. Left-skewed distributions are also called negatively-skewed distributions. That's because there is a long tail in the negative direction on the number line. The mean is also to the left of the peak.
2. A **right-skewed distribution** has a long right tail. Right-skewed distributions are also called positive-skew distributions. That's because there is a long tail in the positive direction on the number line. The mean is also to the right of the peak.
3. **Normal distribution** is the symmetric distribution with no skew.



Poisson Distribution

A Poisson distribution is a tool that helps to predict the probability of certain events happening when you know how often the event has occurred. It gives us the probability of a given number of events happening in a fixed interval of time.

Poisson distribution is calculated based on the *pmf*.

Exponential Distribution

The exponential distribution is a continuous distribution that is commonly used to measure the expected time for an event to occur. For example, in physics it is often used to measure radioactive decay, in engineering it is used to measure the time associated with receiving a defective part on an assembly line, and in finance it is often used to measure the likelihood of the next default for a portfolio of financial assets. It can also be used to measure the likelihood of incurring a specified number of defaults within a specified time period.

