Probability Density Estimation & Maximum Likelihood Estimation

Probability Density: Assume a random variable x that has a probability distribution p(x). The relationship between the outcomes of a random variable and its probability is referred to as the probability density.

The problem is that we don't always know the full probability distribution for a random variable. This is because we only use a small subset of observations to derive the outcome. This problem is referred to as **Probability Density Estimation** as we use only a random sample of observations to find the general density of the whole sample space.

Probability Density Function (PDF)

A PDF is a function that tells the probability of the random variable from a sub-sample space falling within a particular range of values and not just one value. It tells the likelihood of the range of values in the random variable sub- space being the same as that of the whole sample.

By definition, if X is any continuous random variable, then the function f(x) is called a probability density function if:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

```
where,
a -> lower limit
b -> upper limit
X -> continuous random variable
f(x) -> probability density function
```

Steps Involved:

```
Step 1 - Create a histogram for the random set of observations to understand the
         density of the random sample.

Step 2 - Create the probability density function and fit it on the random sample.
         Observe how it fits the histogram plot.

Step 3 - Now iterate steps 1 and 2 in the following manner:
         3.1 - Calculate the distribution parameters.
          3.2 - Calculate the PDF for the random sample distribution.
         3.3 - Observe the resulting PDF against the data.
         3.4 - Transform the data to until it best fits the distribution.
```

Most of the histogram of the different random sample after fitting should match the histogram plot of the whole population.

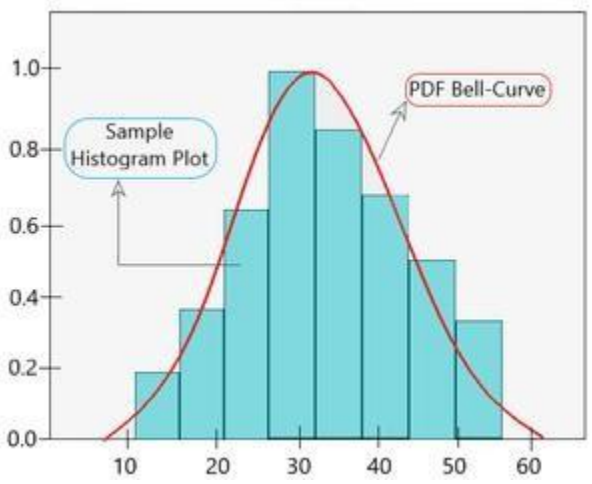
Density Estimation: It is the process of finding out the density of the whole population by examining a random sample of data from that population. One of the best ways to achieve a density estimate is by using a histogram plot.

Parametric Density Estimation

A normal distribution has two given parameters, mean and standard deviation. We calculate the sample mean and standard deviation of the random sample taken from this population to estimate the density of the random sample. The reason it is termed as *'parametric'* is due to the fact that the relation between the observations and its probability can be different based on the values of the two parameters.

Now, it is important to understand that the mean and standard deviation of this random sample is not going to be the same as that of the whole population due to its small size. A sample plot for parametric density estimation is shown below.

below.



PDF fitted over histogram plot with one peak value

Nonparametric Density Estimation

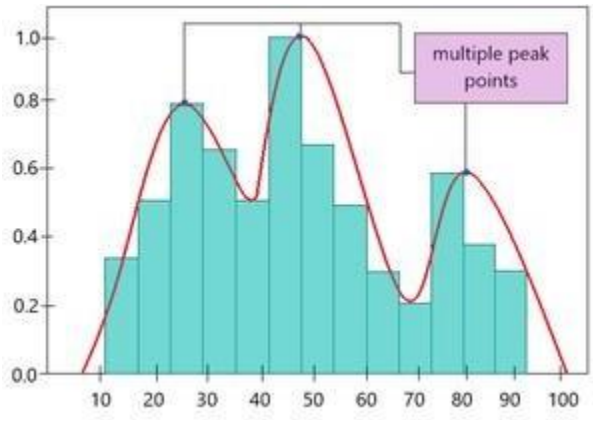
In some cases, the PDF may not fit the random sample as it doesn't follow a normal distribution (i.e instead of one peak there are multiple peaks in the graph). Here, instead of using distribution parameters like mean and standard deviation, a particular algorithm is used to estimate the probability distribution. Thus, it is known as a *'nonparametric density estimation'*.

One of the most common nonparametric approach is known as **Kernel Density Estimation**. In this, the objective is to calculate the unknown density $f_h(x)$ using the equation given below:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x-x_i}{h})$$

```
where,
K -> kernel (non-negative function)
h -> bandwidth (smoothing parameter, h > 0)
Kh -> scaled kernel
fh(x) -> density (to calculate)
n -> no. of samples in random sample.
```

A sample plot for nonparametric density estimation is given below.



PDF plot over sample histogram plot based on KDE

Problems with Probability Distribution Estimation

Probability Distribution Estimation relies on finding the best PDF and determining its parameters accurately. But the random data sample that we consider, is very small. Hence, it becomes very difficult to determine what parameters and what probability distribution function to use. To tackle this problem, Maximum Likelihood Estimation is used.

Maximum Likelihood Estimation

It is a method of determining the parameters (mean, standard deviation, etc) of normally distributed random sample data or a method of finding the best fitting PDF over the random sample data. This is done by maximizing the likelihood function so that the PDF fitted over the random sample. Another way to look at it is that MLE function gives the mean, the standard deviation of the random sample is most similar to that of the whole sample**.**

> **NOTE:** MLE assumes that all PDFs are a likely candidate to being the best fitting curve. Hence, it is computationally expensive method.

Intuition:

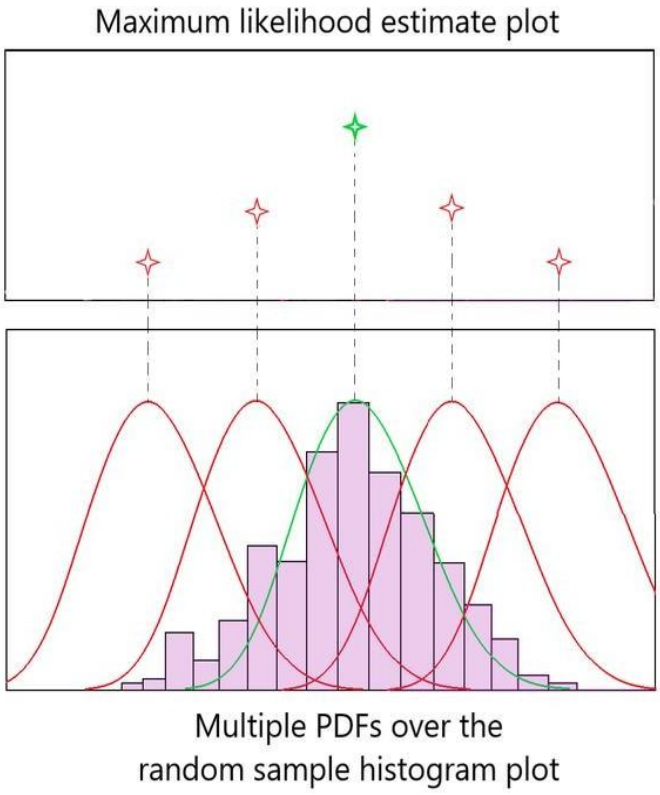


Fig 1 : MLE Intuition

Fig 1 shows multiple attempts at fitting the PDF bell curve over the random sample data. Red bell curves indicate poorly fitted PDF and the green bell curve shows the best fitting PDF over the data. We obtained the optimum bell curve by checking the values in Maximum Likelihood Estimate plot corresponding to each PDF.

As observed in Fig 1, the red plots poorly fit the normal distribution, hence their *'likelihood estimate'* is also lower. The green PDF curve has the maximum likelihood estimate as it fits the data perfectly. This is how the maximum likelihood estimate method works.

Mathematics Involved

In the intuition, we discussed the role that Likelihood value plays in determining the optimum PDF curve. Let us understand the math involved in MLE method.

We calculate Likelihood based on conditional probabilities. See the

$$L = F([X_1 = x_1], [X_2 = x_2], \ldots, [X_n = x_n] \mid P) = \Pi_{i=1}^{n} P^{x_i}(1-P)^{1-x_i}$$

where,
L -> Likelihood value
F -> Probability distribution function
P -> Probability
$X_1, X_2, \ldots X_n$ -> random sample of size n taken from the whole population.
$x_1, x_2, \ldots x_n$ -> values that these random sample (Xi) takes when determining the PDF.
Π -> product from 1 to n.

In the above-given equation, we are trying to determine the likelihood value by calculating the joint probability of each X_i taking a specific value x_i involved in a particular PDF. Now, since we are looking for the maximum likelihood value, we differentiate the likelihood function w.r.t P and set it to 0 as given below.

$$\frac{\partial L}{\partial P} = 0$$

This way, we can obtain the PDF curve that has the maximum likelihood of fit over the random sample data.

But, if you observe carefully, differentiating L w.r.t P is not an easy task as all the probabilities in the likelihood function is a product. Hence, the calculation becomes computationally expensive. To solve this, we take the log of the Likelihood function L.

Log Likelihood

$$\log(L) = \log(\Pi_{i=1}^{n} P^{x_i}(1-P)^{1-x_i})$$

Taking the log of likelihood function gives the same result as before due to the increasing nature of Log function. But now, it becomes less computational due to the property of logarithm:

$$\log(a) + \log(b)$$

Thus, the equation becomes:

$$\log(L) = \log[\Pi_{i=1}^{n} P^{x_i}(1-P)^{1-x_i}]$$
$$= \Sigma_{i=1}^{n} \log[P^{x_i}(1-P)^{1-x_i}]$$

Now, we can easily differentiate log L wrt P and obtain the desired result.

Covariance and Correlation

Covariance and correlation are the two **key concepts in Statistics** that help us **analyze the relationship between two variables.** Covariance measures how two variables change together, **indicating whether they move in the same or opposite directions**.

However, its magnitude can be difficult to interpret because it's not standardized. Correlation, refines this measure by normalizing covariance, Correlation explains the proportion in which the second variable change. **Correlation varies between -1 to +1.** If the correlation value is 0 then it means there is no Linear Relationship between variables however other functional relationship may exist. This allows for a clearer understanding of both the strength and direction of the relationship between variables.

What is Covariance?

Covariance is a statistical measure that **indicates the direction of the linear relationship between two variables**. It assesses how much two variables change together from their mean values.

Types of Covariance:

- **Positive Covariance**: When one variable increases, the other variable tends to increase as well, and vice versa.
- **Negative Covariance**: When one variable increases, the other variable tends to decrease.
- **Zero Covariance**: There is no linear relationship between the two variables; they move independently of each other.

Covariance is calculated by taking the average of the product of the deviations of each variable from their respective means. It is useful for understanding the direction of the relationship but not its strength, as its magnitude depends on the units of the variables.

It is an essential tool for understanding how variables change together and are widely used in various fields, including finance, economics, and science.

Covariance:

1. It is the relationship between a pair of random variables where change in one variable causes change in another variable.
2. It can take any value between – infinity to +infinity, where the negative value represents the negative relationship whereas a positive value represents the positive relationship.
3. It is used for the linear relationship between variables.
4. It gives the direction of relationship between variables.

Covariance Formula

For Population:

$$Covri(x,y) = \frac{\sum_{i=1}^{n}(x_i - x')\,(y_i - y')}{n}$$

For Sample:

$$Covari(x,y) = \frac{\sum_{i=1}^{n}(x_i - x')\,(y_i - y')}{n-1}$$

Here, x' and y' = mean of given sample set n = total no of sample xi and yi = individual sample of set

Example –

What is Correlation?

Correlation is a standardized measure of the strength and direction of the linear relationship between two variables. It is derived from covariance and ranges between -1 and 1. Unlike covariance, which only indicates the direction of the relationship, correlation provides a standardized measure.

- Positive Correlation (close to +1): As one variable increases, the other variable also tends to increase.
- Negative Correlation (close to -1): As one variable increases, the other variable tends to decrease.

- Zero Correlation: There is no linear relationship between the variables. The correlation coefficient ρ\rhoρ (rho) for variables X and Y is defined as:

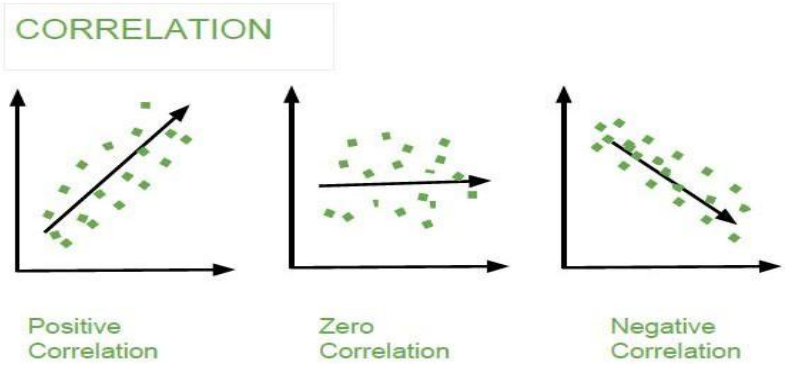
1. It show whether and how strongly pairs of variables are related to each other.
2. Correlation takes values between -1 to +1, wherein values close to +1 represents strong positive correlation and values close to -1 represents strong negative correlation.
3. In this variable are indirectly related to each other.
4. It gives the direction and strength of relationship between variables.

Correlation Formula

$$Corr(x, y) = \frac{\sum_{i=1}^{n}(x_i - x')\,(y_i - y')}{\sqrt{\sum_{i=1}^{n}(x_i - x')^2 \sum_{i=1}^{n}(y_i - y')^2}}$$

Here, x' and y' = mean of given sample set n = total no of sample xi and yi = individual sample of set

Example –



Difference between Covariance and Correlation

This table shows the difference between Covariance and Covariance:

Covariance	Correlation
Covariance is a measure of how much two random variables vary together	Correlation is a statistical measure that indicates how strongly two variables are related.
Involves the relationship between two variables or data sets	Involves the relationship between multiple variables as well
Lie between -infinity and +infinity	Lie between -1 and +1
Measure of correlation	Scaled version of covariance
Provides direction of relationship	Provides direction and strength of relationship
Dependent on scale of variable	Independent on scale of variable Have
dimensions	Dimensionless

Applications of Covariance and Correlation

Applications of Covariance

- Portfolio Management in Finance: Covariance is used to measure how different stocks or financial assets move together, aiding in portfolio diversification to minimize risk.
- Genetics: In genetics, covariance can help understand the relationship between different genetic traits and how they vary together.

- Econometrics: Covariance is employed to study the relationship between different economic indicators, such as the relationship between GDP growth and inflation rates.
- Signal Processing: Covariance is used to analyze and filter signals in various forms, including audio and image signals.
- Environmental Science: Covariance is applied to study relationships between environmental variables, such as temperature and humidity changes over time.

Applications of Correlation

- Market Research: Correlation is used to identify relationships between consumer behavior and sales trends, helping businesses make informed marketing decisions.
- Medical Research: Correlation helps in understanding the relationship between different health indicators, such as the correlation between blood pressure and cholesterol levels.
- Weather Forecasting: Correlation is used to analyze the relationship between various meteorological variables, such as temperature and humidity, to improve weather predictions.
- Machine Learning: Correlation analysis is used in feature selection to identify which variables have strong relationships with the target variable, improving model accuracy.