1. Introduction and Normal Distribution

Welcome to the series of E-learning modules on standard Univariate continuous distributions and their properties.
Here we shall discuss in brief, the different continuous distributions and their properties.

By the end of this session, you will be able to:

- Understand the uses, applications and properties of
  - Normal distributions
  - Exponential distribution
  - Beta distribution
  - Gamma distribution
  - Pareto distribution
  - Laplace distribution
  - Cauchy distribution and
  - Logistic distribution

We consider some univariate continuous distributions in this module like uniform distribution, normal distribution, gamma distribution, beta distribution, exponential distribution, Laplace, Weibul, Logistic and Cauchy distribution. We shall discuss these distributions in brief and in the coming modules we will discuss these distributions in detail.

Consider the normal distribution.
Normal distribution plays a very important role in statistical theory because of the following reasons.

i. Most of the distributions occurring in practice, e.g., Binomial, Poisson, Hypergeometric distributions etc. can be approximated by normal distribution.
ii. Moreover, many of the sampling distributions, e.g., Student's t, Snedecor's F and – chi square distributions, etc., tend to normality for large samples
iii. Even if a variable is not normally distributed, it can sometimes be brought to normal form by simple transformation of variable. For example, if the distribution of X is skewed, the distribution of square root of x might come out to be normal
iv. If X follows normal distribution with parameters mu and sigma square, then Probability of (mu minus three into sigma is less than or equal to X less than or equal to mu plus three into sigma) is equal to Probability of (minus three less than or equal to Z less than or equal to three) is equal to zero point nine, nine, seven, three. This property of normal distribution forms the basis of entire large sample theory
v. Many of the distributions of sample statistics (for example, the distributions of mean, sample variance etc.,) tend to normality for large samples and as such they can be best studied with the help of the normal curves.
vi. The entire theory of small sample tests, namely t, F, chi square tests, etc., is based on the fundamental assumption that the parent populations form which the samples have been drawn follow normal distribution
vii. Normal distribution finds large applications in Statistical Quality Control in industry for

setting control limits.

The following quote by Lipman rightly reveals the popularity and importance of normal distribution.
“Everybody believes in the law of errors (the normal curve), the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is experimental fact.”

W, J Youden of the National Bureau of Standards describes the importance of the Normal Distribution artistically in the following words.
The normal Law of errors Stands out in the Experience of mankind As one of the broadest Generalizations of natural Philosophy. It serves as the Guiding instrument in researches, In the physical and social sciences And in medicine, agriculture and Engineering. It is an indispensable tool for The analysis and the interpretation of the Basic data obtained by observation and experiment.

2. Properties of Normal Distribution

Now let us list out the properties of normal distribution.
The normal distribution with parameters mu and sigma square has the following properties.

- The curve of the distribution is bell-shaped and symmetrical about the line x is equal to mu
- Mean, median and mode of the distribution coincide.
- As x increases numerically, f of (x) decreases rapidly, the maximum probability occurring at the point x is equal to mu and is given by, one divided by sigma into square root of two into pi
- Beta one is equal to zero and beta two is equal to three
- All odd order central moments are equal to zero and even order central moments are given by, mu two r is equal to one into three into five into etc., into (2 r minus one) into sigma to the power two into r, where (r is equal to zero, one, two etc.)
- Since f of (x) being the probability, can never be negative, no portion of the curve lies below the x axis
- Linear combination of independent normal variates is also a normal variate
- x-axis is an asymptote to the curve
- The points of inflexion of the curve are x is equal to mu plus or minus sigma
- Mean deviation about mean is equal to sigma into square root of two divided by pi nearly equal to four divided by five into sigma
- Quartiles are given by, Q one is equal to mu minus zero point six seven four five sigma and Q three is equal to mu plus zero point six seven four five sigma
- Quartile deviation Q.D. is equal to Q three minus Q one whole divided by two nearly equal to two divided by three into sigma. We have approximately Q D is to M D is to S D is equal to two by three into sigma is to four by five into sigma is to sigma, which is same as two by three is to four by five is to one
- Area property of normal distribution is as follows
  Probability of mu minus sigma less than X less than mu plus sigma is equal to zero point six eight two six;
  Probability of mu minus two into sigma less than X less than mu plus two into sigma is equal to zero point nine five four four
  Probability of mu minus sigma less than X less than mu plus sigma is equal to zero point nine nine seven three

- If X and Y are independent standard normal variates, then it can be easily proved that U is equal to X plus Y and U is equal to X minus Y are independently distributed as normal distribution with parameters (zero, and two)
  The converse of this property is true which is given by D. Bernstein. That is if X and Y are independent and identically distributed random variables with finite variances and if U is equal to X plus Y and U is equal to X minus Y are independent, the all random variables X, Y, U and V are independent.

Some of the above properties will be proved in our discussions in the coming modules.

3. Uniform, Exponential and Beta Distribution

Uniform distribution is also known as a rectangular distribution. Since the curve of uniform distribution describes a rectangle over the X axis and between the ordinates at x is equal to ‘a’ and ‘x’ is equal to ‘b’ if X follows uniform distribution over the interval (a, b).

The mean of the distribution is usually the mean of the limits of the range of the distribution. That is if X has uniform distribution over the interval (a, b) then the mean of the distribution is given by, (a plus b) divided by two. Also for uniform distribution, the mean is equal to median. Here mode is ill-defined.

Now let us discuss some of the properties of uniform distribution.
The probability that a uniformly distributed random variable falls within any interval of fixed length is

- Independent of the location of the interval itself
- But it is dependent on the interval size, so long as the interval is contained in the distribution's support
- This distribution can be generalized to more complicated sets than intervals.
- If *S* is a Borel set of positive, finite measure, the uniform probability distribution on *S* can be specified by defining the probability density function to be zero outside *S* and constantly equal to 1divided by *K* on *S*, where *K* is the Lebesgue measure of *S*

Now let us look at the properties of exponential distribution.
An important property of the exponential distribution is that it is memoryless. This means that if a random variable *T* is exponentially distributed, its conditional probability obeys
Probability that T greater than s plus t given T greater than s is equal to probability of T greater than t for all s, t greater than or equal to zero.

This says that for conditional probability we need to wait, for example, more than another ten seconds before the first arrival, given that the first arrival has not yet happened after thirty seconds, is equal to the initial probability that we need to wait more than ten seconds for the first arrival

So, if we waited for thirty seconds and the first arrival didn't happen (*T* is greater than thirty), probability that we'll need to wait another ten seconds for the first arrival (*T* is greater than thirty plus ten) is the same as the initial probability that we need to wait more than ten seconds for the first arrival (*T* is greater than ten). The fact that Probability that (*T* is greater than forty given *T* is greater than thirty) is equal to Probability that (*T* greater than ten) does *not* mean that the events *T* greater than forty and *T* greater than thirty are independent.

Now let us consider beta distribution of first kind. Here the parameters are taken as alpha and beta.
If one is less than alpha, less than beta then mode is less than or equal to median, which is

less than or equal to mean. Expressing the mode (only for alpha greater than one and beta greater than one), and the mean in terms of alpha and beta:
Alpha minus one divided by alpha plus beta minus two is less than or equal to median which is less than or equal to alpha divided by alpha plus beta.

If one is less than beta less than alpha then the order of the inequalities are reversed. For alpha greater than one and β greater than one the absolute distance between the mean and the median is less than five per cent of the distance between the maximum and minimum values of *x*.

On the other hand, the absolute distance between the mean and the mode can reach fifty per cent of the distance between the maximum and minimum values of *x*, for the (pathological) case of *alpha nearly equal to* one and;
beta nearly equal to one (for which values the beta distribution approaches the uniform distribution and the differential entropy approaches its maximum value, and hence maximum "disorder").

4. Gamma and Pareto Distribution

Now let us consider the gamma distribution.
If a gamma distribution with parameter α and β are considered, then

- Gamma distribution is considered as the distribution of alpha independently identically distributed exponential variates with parameter beta
- The gamma distribution has been used to model the size of insurance claims and rainfalls. This means that aggregate insurance claims and the amount of rainfall accumulated in a reservoir are modeled by a gamma process
- The gamma distribution is also used to model errors in multi-level Poisson regression models, because the combination of the Poisson distribution and a gamma distribution is a negative binomial distribution

- In neuroscience, the gamma distribution is often used to describe the distribution of inter-spike intervals. Although in practice the gamma distribution often provides a good fit, there is no underlying biophysical motivation for using it
- In bacterial gene expression, the copy number of a constitutively expressed protein often follows the gamma distribution, where the scale and shape parameter are, respectively, the mean number of bursts per cell cycle and the mean number of protein molecules produced by a single mRNA during its lifetime.
- The gamma distribution is widely used as a conjugate prior in Bayesian statistics. It is the conjugate prior for the precision (i.e. inverse of the variance) of a normal distribution. It is also the conjugate prior for the exponential distribution.

Now let us consider Pareto distribution.
The Pareto distribution is a skewed, heavy-tailed distribution that is sometimes used to model the distribution of incomes and other financial variables.
Pareto originally used this distribution to describe the allocation of wealth among individuals.

It seemed to show rather well the way that a larger portion of the wealth of any society is owned by a smaller percentage of the people in that society. He also used it to describe distribution of income. This idea is sometimes expressed more simply as the Pareto principle or the "eighty-twenty rule" which says that twenty percent of the population controls eighty percent of the wealth.

- However, the eighty twenty rule corresponds to a particular value of α, and in fact, Pareto's data on British income taxes in his *Cours d'économie politique* indicates that about thirty percent of the population had about seventy percent of the income.
- The probability density function graph at the beginning of this article shows that the "probability" or fraction of the population that owns a small amount of wealth per person is rather high, and then decreases steadily as wealth increases.

Note that the Pareto distribution is not realistic for wealth for the lower end. In fact, net worth may even be negative.
This distribution is not limited to describing wealth or income, but to many situations in which an equilibrium is found in the distribution of the "small" to the "large".

The following examples are sometimes seen as approximately Pareto-distributed:

- The sizes of human settlements (few cities, many hamlets/villages)
- File size distribution of Internet traffic which uses the TCP protocol (many smaller files, few larger ones)
- Hard disk drive error rates
- Clusters of Bose–Einstein condensate near absolute zero
- The values of oil reserves in oil fields (a few large fields, many small fields)
- The length of distribution in jobs assigned to supercomputers (a few large ones, many small ones)
- The standardized price returns on individual stocks
- Sizes of sand particles
- Sizes of meteorites
- Numbers of species per genus (There is subjectivity involved: The tendency to divide a genus into two or more increases with the number of species in it)
- Areas burnt in forest fires
- Severity of large casualty losses for certain lines of business such as general liability, commercial auto, and workers compensation
- In hydrology the Pareto distribution is applied to extreme events such as annually maximum one-day rainfalls and river discharges

5. Laplace, Cauchy and Logistic Distribution

Now let us consider the Laplace distribution

- The Laplacian distribution has been used in speech recognition to model priors on DFT coefficients.
- The addition of noise drawn from a Laplacian distribution, with scaling parameter appropriate to a function's sensitivity, to the output of a statistical database query is the most common means to provide differential privacy in statistical databases.
- The Laplace distribution has found a variety of very specific uses, but they nearly all relate to the fact that it has long tails compared to the Normal distribution.
- It has recently become quite popular in modeling financial variables (Brownian Laplace motion) like stock returns because of the greater tails. The Laplace distribution is very extensively reviewed in the monograph Kotz et al (2001).

Consider Cauchy distribution.

- The Cauchy distribution is often used in statistics as the canonical example of a "pathological" distribution.
- Both its mean and its variance are undefined
- The Cauchy distribution does not have finite moments of order greater than or equal to one; only fractional absolute moments exist
- The Cauchy distribution has no moment generating function
- The Cauchy distribution is an example of a distribution which has no mean, variance or higher moments defined.
- Its mode and median are well defined and are both equal to x naught
- The Cauchy distribution is an infinitely divisible probability distribution. It is also a strictly stable distribution
- When U and V are two independent normally distributed random variables with expected value zero and variance one, then the ratio U divided by V has the standard Cauchy distribution.
- It is also an example of a more generalized version of the central limit theorem that is characteristic of all stable distributions, of which the Cauchy distribution is a special case.

Now let us consider logistic distribution
The logistic distribution — and the S-shaped pattern of its cumulative distribution function (the logistic function) and quantile function (the logit function) — have been extensively used in many different areas.

One of the most common applications is in logistic regression, which is used for modeling categorical dependent variables (e.g. Yes-no choices or a choice of three or four possibilities), much as standard linear regression is used for modeling continuous variables (e.g. income or population).

Specifically, logistic regression models can be phrased as latent variable models with error variables following a logistic distribution. This phrasing is common in the theory of discrete choice models, where the logistic distribution plays the same role in logistic regression as the normal distribution does in probit regression.

The logistic and normal distributions have a quite similar shape. However, the logistic distribution has heavier tails, which often increases the robustness of analyses based on it compared with using the normal distribution.

Some of the other applications are,

- In Biology, to describe how species populations grow in competition
- In Epidemiology, to describe the spreading of epidemics
- In Psychology, to describe learning
- Technology , to describe how new technologies diffuse and substitute for each other
- In Marketing – the diffusion of new-product sales
- In Energy – the diffusion and substitution of primary energy sources, as in the Hubbert curve
- In Hydrology - In hydrology the distribution of long duration river discharge and rainfall (e.g. monthly and yearly totals, consisting of the sum of respectively thirty and three hundred and sixty daily values) is often thought to be almost normal according to the central limit theorem.

The normal distribution, however, needs a numeric approximation. As the logistic distribution, which can be solved analytically, is similar to the normal distribution, it can be used instead.

- In Physics - the cdf of this distribution describes a Fermi gas and more specifically the number of electrons within a metal that can be expected to occupy a given quantum state. Its range is between 0 and 1, reflecting the Pauli exclusion principle. The value is given as a function of the kinetic energy corresponding to that state and is parameterized by the Fermi energy and also the temperature (and Boltzmann constant).

  - By changing the sign in front of the "one" in the denominator, one goes from Fermi–Dirac statistics to Bose–Einstein statistics. In this case, the expected number of particles (bosons) in a given state *can* exceed unity, which is indeed the case for systems such as lasers.

Here's a summary of our learning in this session where we have:

- Understood the uses, applications and properties of
  - Normal distributions
  - Exponential distribution
  - Beta distribution
  - Gamma distribution
  - Pareto distribution
  - Laplace distribution
  - Cauchy distribution and
  - Logistic distribution