Th5: Statistical Distributions: Continuous, Discrete, Properties and simulations
Statistical distributions

In probability theory and statistics, a probability distribution is the mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment. It is a mathematical description of a random phenomenon in terms of its sample space and the probabilities of events (subsets of the sample space).

Probability distributions are used to compare the relative occurrence of many different random values. Probability distributions can be defined in different ways and for discrete or for continuous variables.

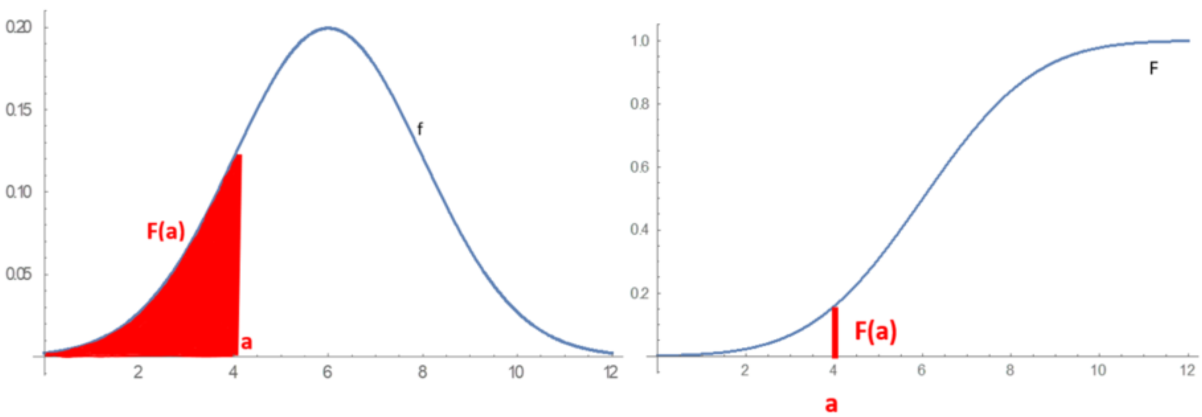
The sample space, often denoted by Ω , is the set of all possible outcomes of a random phenomenon being observed; it may be any set: a set of real numbers, a set of vectors, a set of arbitrary non-numerical values, etc.

To define probability distributions for the specific case of random variables (so the sample space can be seen as a numeric set), it is common to distinguish between discrete and absolutely continuous random variables.

In the discrete case, it is sufficient to specify a probability mass function P assigning a probability to each possible outcome. The probability of an event is then defined to be the sum of the probabilities of the outcomes that satisfy the event.

In contrast, when a random variable takes values from a continuum then typically, any individual outcome has probability zero and only events that include infinitely many outcomes, such as intervals, can have positive probability.

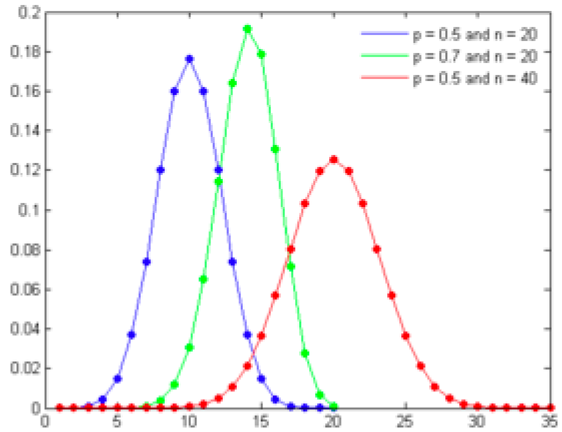
Absolutely continuous probability distributions can be described in several ways. The probability density function describes the infinitesimal probability of any given value, and the probability that the outcome lies in a given interval can be computed by integrating the probability density function over that interval. An alternative description of the distribution is by means of the cumulative distribution function, which describes the probability that the random variable is no larger than a given value
The cumulative distribution function is the area under the probability density function from
$-\infty$ to x, as described by the picture to the right.



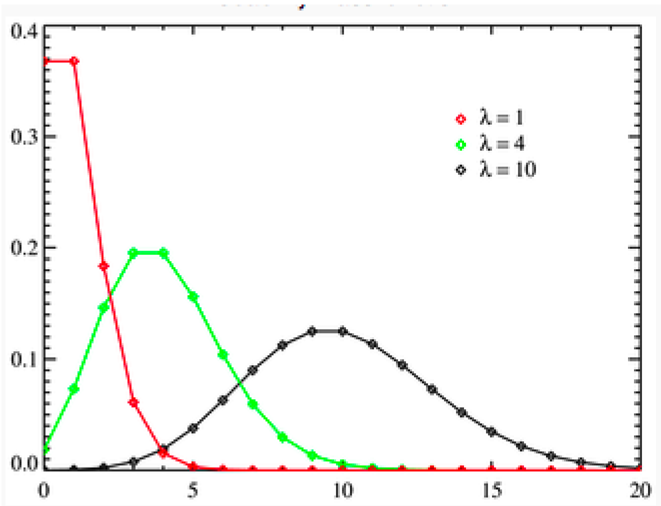
The left graph shows a probability density function. The right graph shows the cumulative distribution function, for which the value at a equals the area under the probability density curve to the left of a.

To better understand the type of data that we are facing, the first categorization of data should be on whether the data is restricted to taking on only discrete values or if it is continuous.
With discrete data, the entire distribution can either be developed from scratch or the data can be fitted to a pre-specified discrete distribution. With the former, there are two steps to building the distribution. The first is identifying the possible outcomes and the second is to estimate probabilities to each outcome. This process is relatively simple to accomplish when there are a few outcomes with a well-established basis for estimating probabilities but becomes more tedious as the number of outcomes increases. If it is difficult or impossible to build up a customized distribution, it may still be possible fit the data to one of the following discrete distributions:

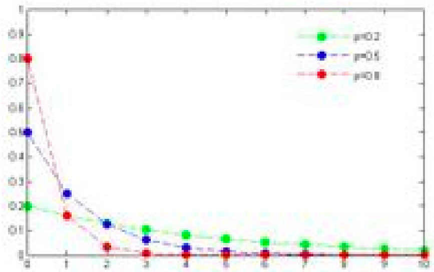
-Binomial distribution: The binomial distribution measures the probabilities of the number of successes over a given number of trials with a specified probability of success in each try.



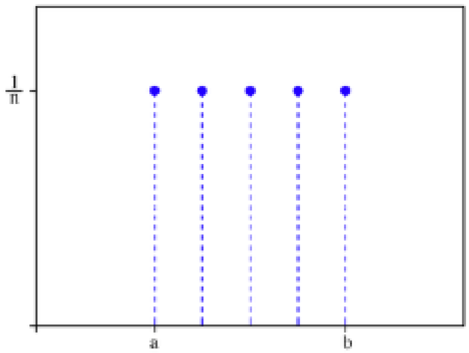
-Poisson distribution: The Poisson distribution measures the likelihood of a number of events occurring within a given time interval, where the key parameter that is required is the average number of events in the given interval lambda.



-Geometric distribution: The Geometric distribution measures the likelihood of when the first success will occur.

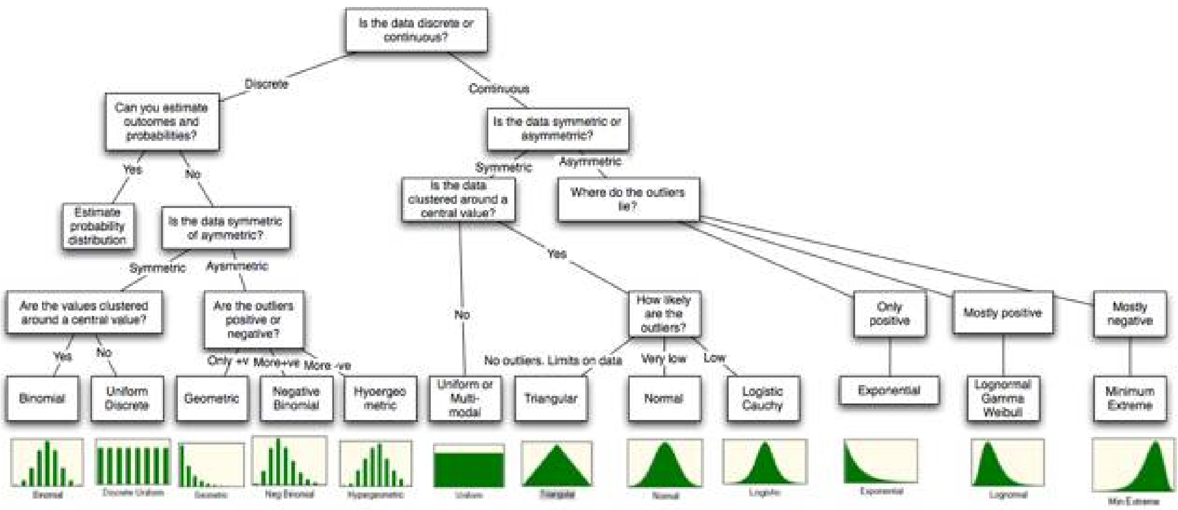


-Discrete uniform distribution: The Discrete uniform distribution applies when all of the outcomes have an equal probability of occurring.



With continuous data, we cannot specify all possible outcomes, since they are too numerous to list, but we have two choices. The first is to convert the continuous data into a discrete form dividing them into classes and then go through the same process that we went through for discrete distributions of estimating probabilities. The second is to find a continuous distribution that best fits the data and to specify the parameters of the distribution.

To visually represent in which distribution the data can be represented we can see the following graph.



Bibliography

-https://en.wikipedia.org/wiki/Probability_distribution

-https://pages.stern.nyu.edu/~adamodar/New_Home_Page/StatFile/statdistns.htm