National Technology of Mexico

Technological Institute of Tijuana

ACADEMIC SUBDIRECTORATE

Systems and Computing Department

SEMESTER

January - June 2021

RACE

Information and Communication Technologies Engineer

SUBJECT AND KEY:

Data Mining BDD-1703TI9A

STUDENT'S NAME AND REGISTRATION:

Camacho Manabe Juan Daniel 17210534

NAME OF THE JOB:

Statistical distributions

UNIT TO BE EVALUATED

Unit 2

TEACHER'S NAME:

Mc José Christian Romero Hernández

What is a distribution?

A probability distribution of a random variable is a function that assigns to each event defined on the variable the probability that the event will occur. The probability distribution is defined over the set of all events and each one of the events is the range of values of the random variable. It can also be said that it has a close relationship with the frequency distributions. In fact, a probability distribution can be understood as a theoretical frequency, since it describes how the results are expected to vary.

There are three types of variables.

1. **Random variable:** It is one whose value is the result of a random event. This means that they are the results that are presented at random in any event or experiment.

2. **Discrete random variable:** It is one that only takes certain values (frequently integers) and that results mainly from the count carried out.

3. **Continuous random variable:** It is one that generally results from the measurement and can take any value within a given interval.

Distribution division

This division is carried out depending on the type of variable to be studied. The four main ones (from which all the others are born) are:

a) If the variable is a discrete variable (integer values), a discrete distribution will correspond, of which there are:

1. **Binomial distribution** (independent events).

2. **Poisson distribution** (independent events)

3. **Hypergeometric distribution** (dependent events).

b) If the variable is continuous, this means that it can take any value within an interval, the distribution that will be generated will be a continuous distribution, also called normal or Gaussian distribution.

Discrete variable distribution is called the one whose probability function only takes positive values in a set of values of X finite or countable infinite. This function is called the probability mass function. In this case the probability distribution is the sum of the mass function.

Types of Discrete Variable Distributions

Defined over an infinite domain

● · The binomial distribution, which describes the number of hits in a series of n independent experiments with possible binary outcomes, that is, "yes" or "no", all of them with probability of success p and probability of failure q = 1 - p.

● · The Bernoulli distribution, the classical binomial, which takes values "1", with probability p, or "0", with probability q = 1 - p (Bernoulli test).

● · The Rademacher distribution, which takes values «1» or «-1» with probability 1/2 each.

● · The beta-binomial distribution, which describes the number of hits in a series of n independent experiments with possible "yes" or "no" results, each with a variable probability of success defined by a beta.

● · The degenerate distribution in x0, in which X takes the value x0 with probability 1. Although it does not seem like a random variable, the distribution satisfies all the requirements to be considered as such.

● · The discrete uniform distribution, which collects a finite set of values that are all equally probable. This distribution describes, for example, the random behavior of a balanced coin, dice, or casino roulette wheel (without bias).

● · The hypergeometric distribution, which measures the probability of obtaining x ($0 \leq x \leq d$) elements of a certain class formed by d elements belonging to a population of N elements, taking a sample of n elements from the population without replacement.

● · Fisher's non-central hypergeometric distribution.

● · The non-central hypergeometric Wallenius distribution.

● · Benford's law, which describes the frequency of the first digit of a set of numbers in decimal notation.

Defined over an infinite domain

● · The negative binomial distribution or Pascal distribution, which describes the number of independent Bernoulli trials necessary to achieve n hits, given a constant individual probability of success p.

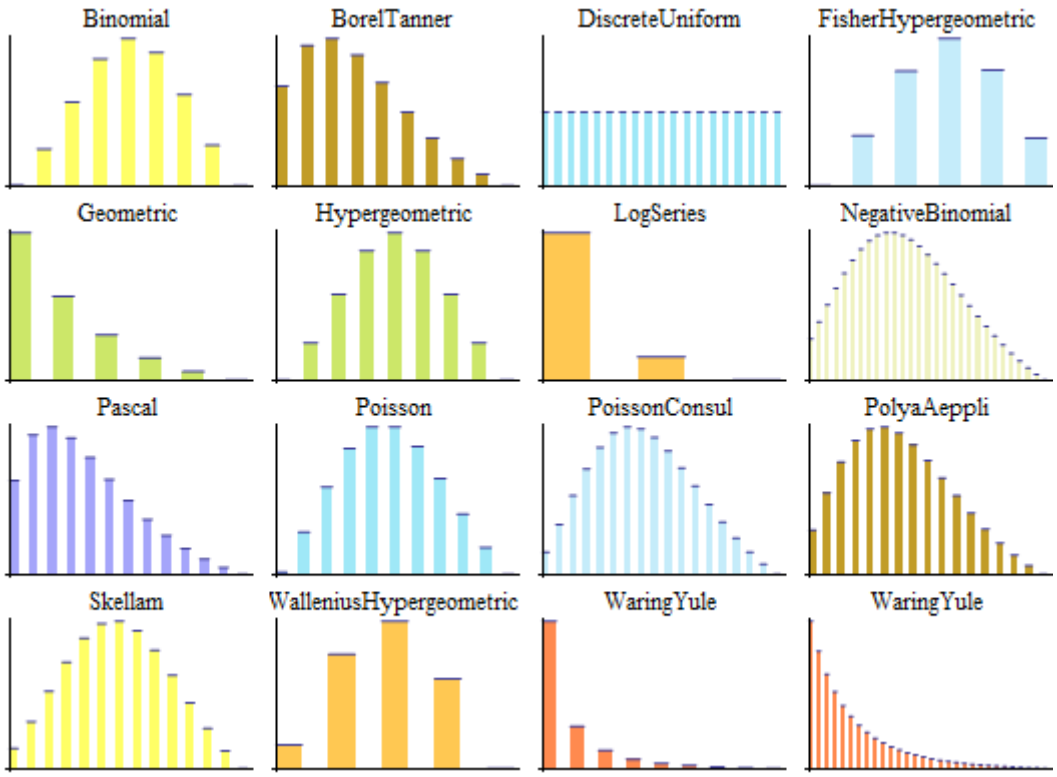
● · The geometric distribution, which describes the number of attempts required to get the first hit.

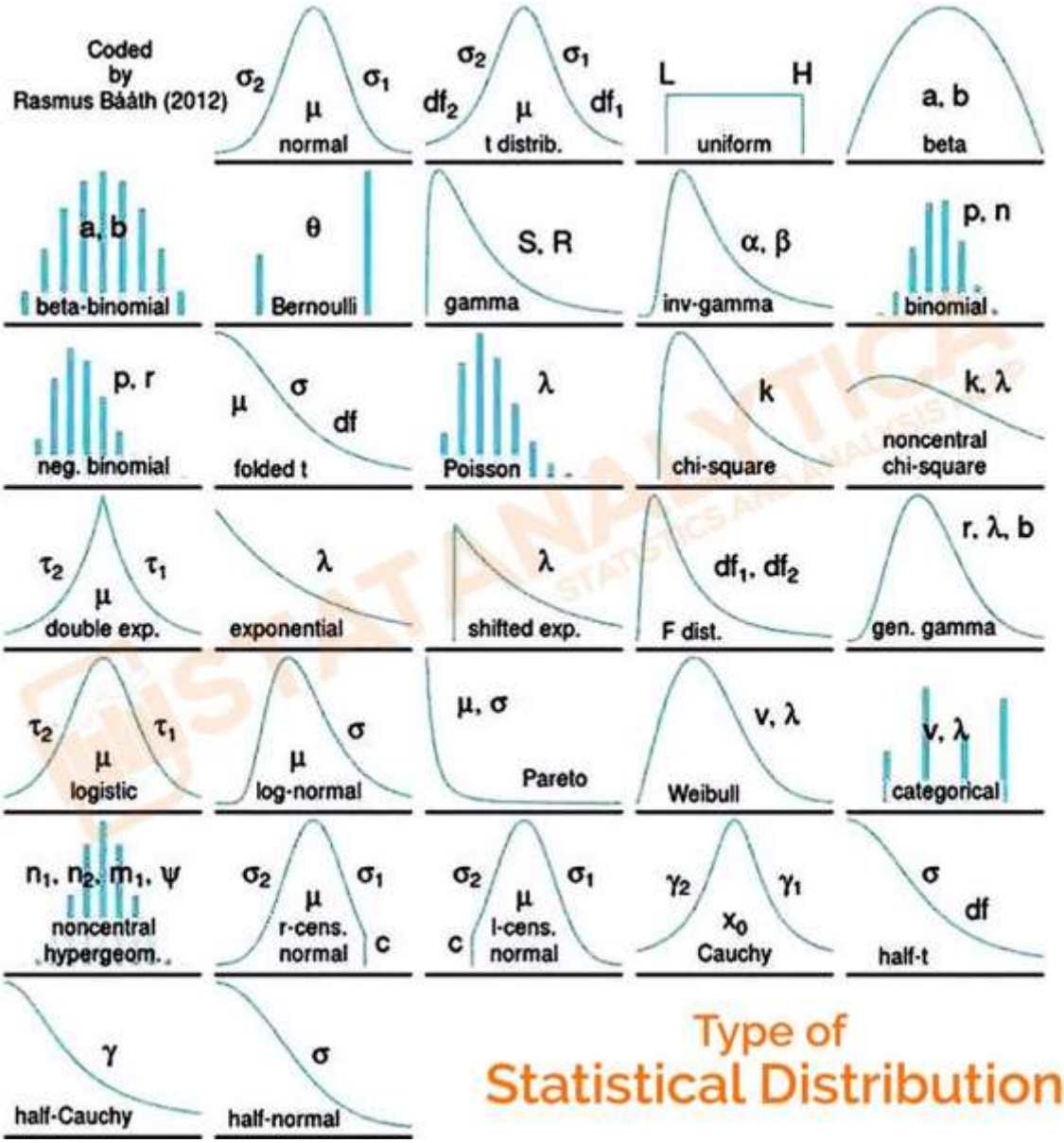
● · The negative binomial-beta distribution, which describes the number of "yes / no" experiments necessary to achieve n hits, when the probability of success of each of the attempts is distributed according to a beta.

● · The extended negative binomial distribution.

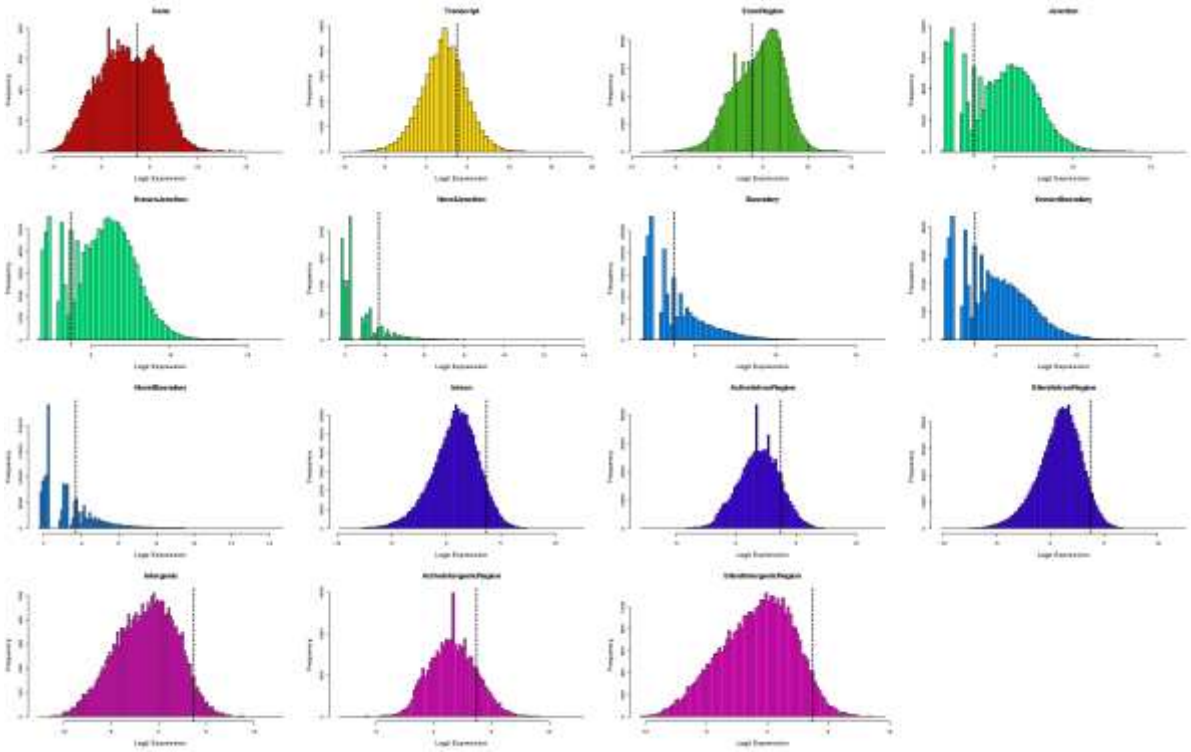
The Boltzmann distribution, important in statistical mechanics, which describes the occupation of discrete energy levels in a system in thermal equilibrium.

Several special cases are:



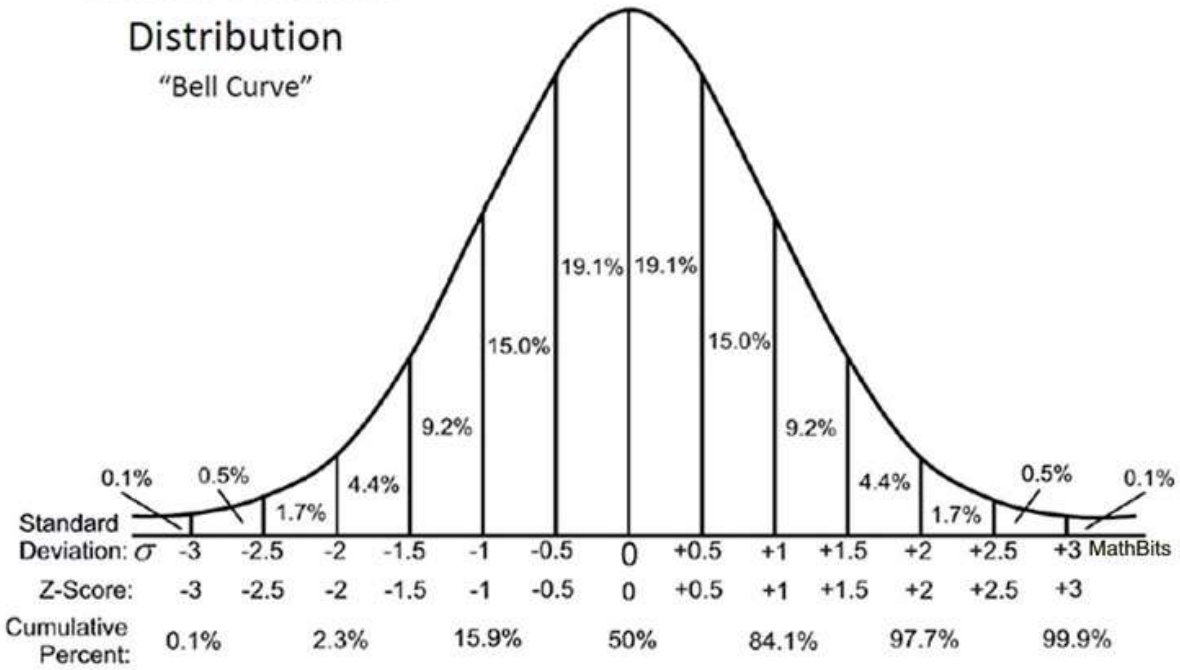


Type of
Statistical Distribution



Standard Normal Distribution

"Bell Curve"



Bibliography

1. *Solano, H. L., & Álvarez, C. R. (2005). Estadística descriptiva y distribuciones de probabilidad. Universidad del Norte.*

2. *Pedrosa, I., Juarros-Basterretxea, J., Robles-Fernández, A., Basteiro, J., & García-Cueto, E. (2015). Pruebas de bondad de ajuste en distribuciones simétricas,¿ qué estadístico utilizar?. Universitas psychologica, 14(1), 245-254.*

3. *Horváth, J. (1974). Distribuciones definidas por prolongación analítica. Revista Colombiana de Matemáticas, 8(2-3), 47-95.*

4. *Larranaga, P., Lozano, J. A., & Mühlenbein, H. (2003). Algoritmos de estimación de distribuciones en problemas de optimización combinatoria. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, 7(19), 0.*

5. *Hoel, P. G. (1968). Estadística elemental (No. 519.5 H645.). Cecsa.*

6. *González, P., & Mª E y Sanz Donaire, J. J. (2000). Distribuciones estadísticas ajustadas a las series temporales de totales anuales de precipitación sobre España: aspectos geográficos. Geographicalia, 38, 11-29.*

7. *Spiegel, M. R., & Losada Villasante, A. (1991). Estadística.*