Statistical Distribution

Chiara Iannicelli

November 2023

1 Introduction

A statistical distribution, often simply referred to as a probability distribution, describes the likelihood of different outcomes in a sample space. It provides a way to model and understand the variability and uncertainty associated with random variables in probability theory and statistics.

The formula for a probability distribution depends on the type of distribution. There are several common probability distributions, each with its own formula. Here follows few examples.

2 Discrete Uniform Distribution

A discrete uniform distribution models outcomes where each value in a range is equally likely.
The probability mass function (PMF) is given by:

$$P(X = x) = \frac{1}{N}$$

where N s the number of possible outcomes.

3 Bernoulli Distribution

A Bernoulli distribution models a binary outcome (success or failure).
The probability mass function is given by:

$$P(X = x) = p^x(1-p)^(1-x)$$

where p is the probability of success (and $1-p$ is the probability of failure), and x is the outcome (0 or 1).

4 Normal Distribution

The normal distribution (also known as the gaussian distribution) is a continuous distribution that is symmetric and bell-shaped.
The probability density function (PDF) is given by:

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

where μ is the mean and σ^2 is the variance.

5 Exponential distribution

The exponential distribution models the time between events in a Poisson process.
The probability density function is given by:

$$f(x \mid \lambda) = \lambda \cdot e^{-\lambda x}$$

where λ is the rate parameter.

6 Graphic representations

These are just a few examples, and there are many other probability distributions, each with its own formula.
The choice of distribution depends on the nature of the data and the underlying statistical assumptions.

Choosing the right probability distribution for specific data is a crucial step in statistical analysis. The choice depends on various factors, and making an informed decision requires a combination of statistical knowledge, understanding of the data, and sometimes empirical testing.
Here are some guidelines to help making the right choice:

- **Understand the Nature of the Data:** Consider whether the data is discrete or continuous. Examine the range and possible values of your data. Look for symmetry, skewness, or multimodality in the data distribution.
- **Consider the Characteristics of the Distribution:** Different distributions have different characteristics. For example, the normal distribution is symmetric and bell-shaped, while the exponential distribution is skewed.
  Consider whether the distribution should be unimodal or multimodal.

- **Review Data Descriptive Statistics:** Calculate descriptive statistics such as mean, median, and standard deviation.
  These statistics can provide insights into the central tendency and variability of the data, helping you identify potential distributional fits.

- **Explore Empirical Data:** Create visualizations such as histograms, probability plots, or Q-Q plots to explore the empirical distribution of your data.
  Compare these plots to the characteristic shapes of common distributions.

- **Consider the Purpose of the Analysis:** The choice of distribution may be influenced by the specific goals of your analysis. For example, if you're modeling waiting times between events, the exponential distribution might be appropriate.

- **Use Statistical Tests:** There are statistical tests, such as the Kolmogorov-Smirnov test or Anderson-Darling test, that can help assess goodness-of-fit between your data and a hypothesized distribution. Keep in mind that these tests have limitations, and their results should be interpreted cautiously.

- **Model Assumptions:** Be aware of the assumptions associated with different distributions. For example, normality is often assumed in many statistical tests, so if your data deviates significantly from normality, you may need to consider alternative approaches or transformations.

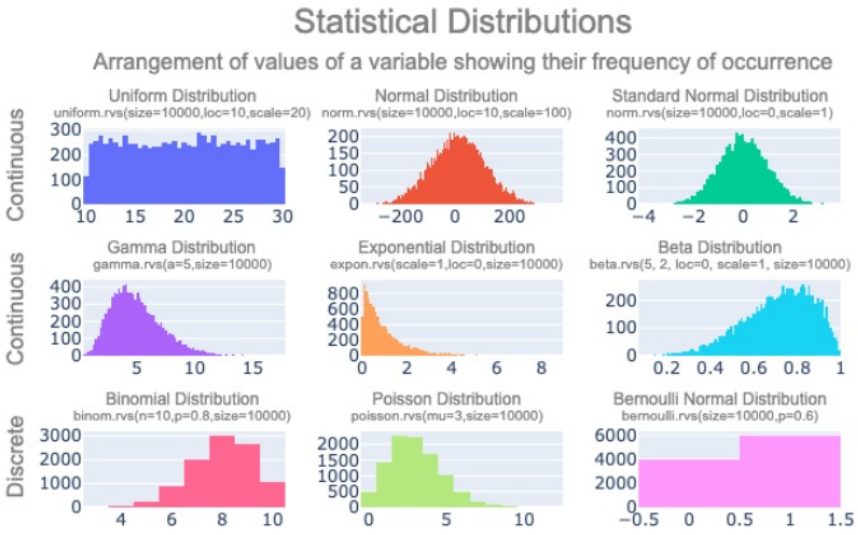

Figure 1: Different distributions