# Th5: Statistical Distributions

Federico Detomaso

1 Meaning

A statistical distribution, or probability distribution, describes how values are distributed for a field. In other words, the statistical distribution shows which values are common and uncommon.

There are many kinds of statistical distributions, including the bell-shaped normal distribution. We use a statistical distribution to determine how likely a particular value is. For example, if we have a chi-square value, we can use the chi-square distribution to determine how likely this chi-square value is.

2 Definition
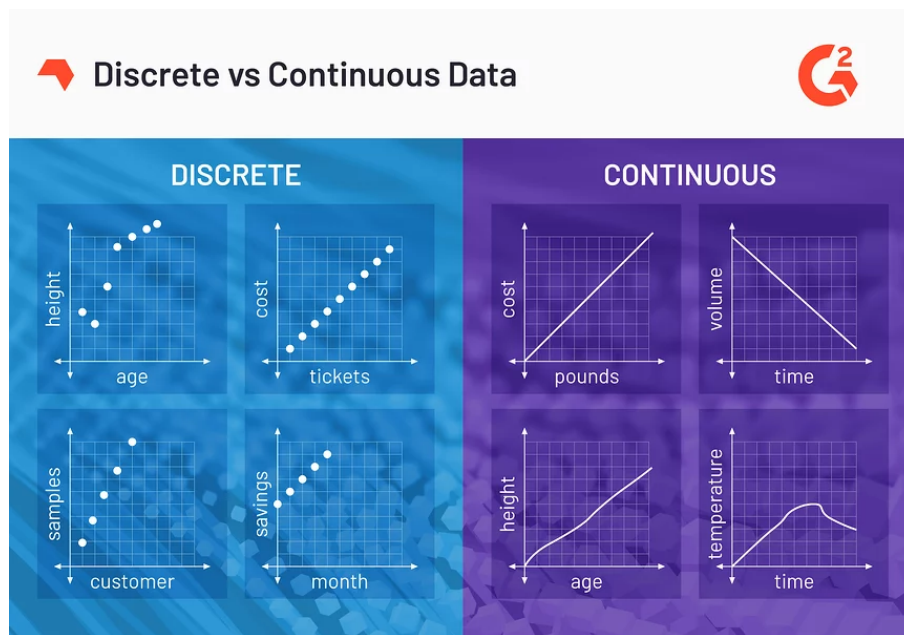
The first and most obvious categorization of data should be on whether the data is restricted to taking on only discrete values or if it is continuous.

2.1 Discrete Data

Data that can only take certain values is called discrete data or discrete values. This is data that can be counted and has a limited number of values. It usually comes in the form of whole numbers or integers. These values must fit into certain categories and can't be broken into smaller parts.

Discrete data examples include the following:

- The size of your department's workforce.

- How many new clients you brought on board in the previous quarter?

-How many items are currently kept in stock?

However, your daily total is a distinct, single amount. The relationship between the number of push-ups you can do each day and your fitness level is yet unknown. The more data you collect over time, the more conclusions you can draw.

2.1.1 Properties

- Takes particular countable values.

- Discrete data is information that has noticeable gaps between values.

- Discrete data is made up of discrete or distinct values.

- Discrete data can be counted.

- Bar graphs are a visual representation of discrete data.

- Ungrouped frequency distribution refers to the tabulation of discrete data against a single value.

2.2 Continuous Data

Continuous data are those that can be measured. It is conceivable for this data to take on an unlimited number of different values because its values are not fixed. There are smaller, individual components that make up these metrics as well.

Continuous data examples would include the following:

-An individual's stature or weight

-The temperature daily in your city

-How much time is required to finish an activity or project

The correctness of continuous data is everything. These data sets frequently contain variables with decimal points, with the rightmost number being as long as possible.

This level of information is required, for example, by scientists, physicians, and manufacturers.

2.2.1 Properties

- Takes any measured value within a given range.

- Continuous data is information that occurs in a continuous series.

- Directly in opposition, continuous data includes any value that falls inside a range.

- Continuous data is quantifiable.

- Continuous data are graphically represented using a histogram.

- The tabulation of continuous data performed against a set of values is called grouped frequency distribution.
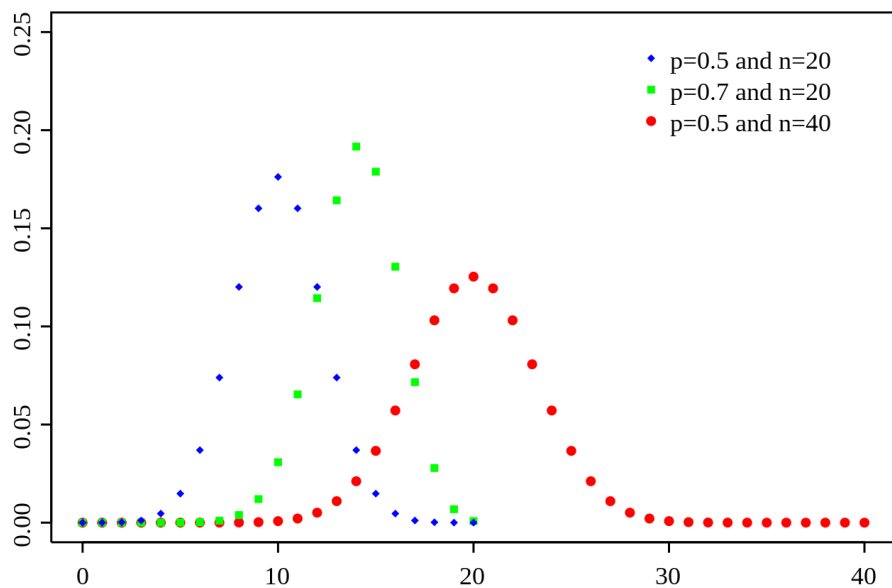
3 Simulations

3.1 Discrete

3.1.1 Bernoulli distribution
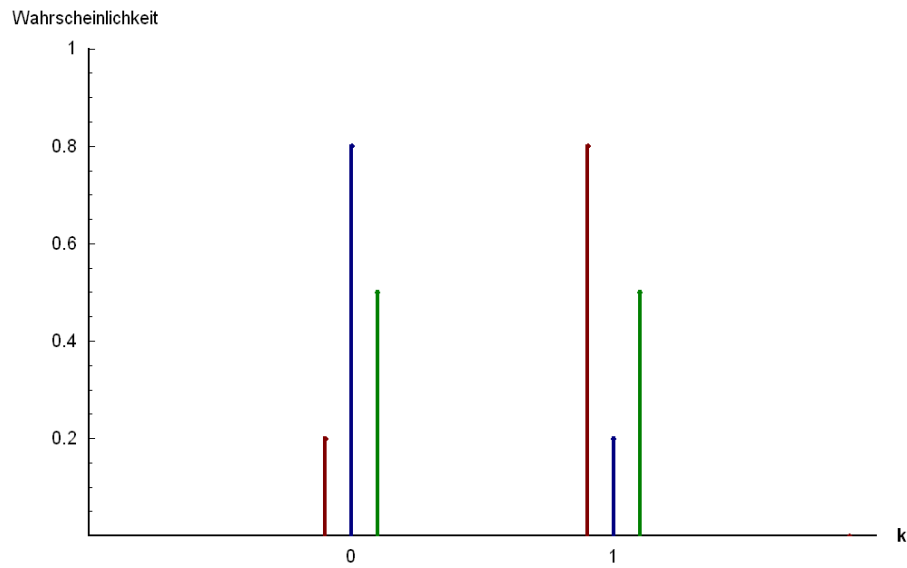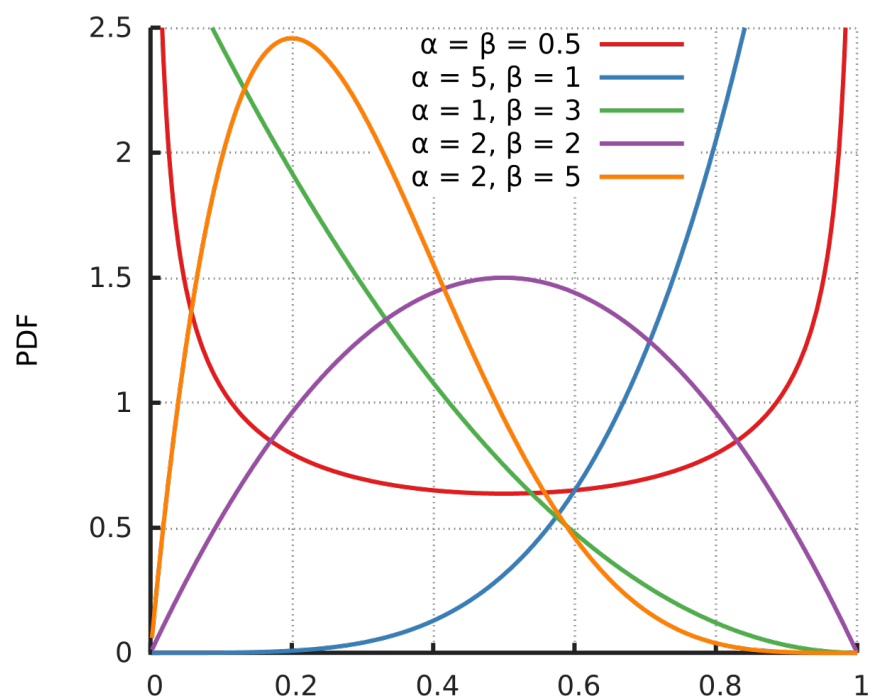

Three examples of Bernoulli distribution:

- RED: $Pr(x = 0) = 0.2$ and $Pr(x = 1) = 0.8$.

- BLUE: $Pr(x = 0) = 0.8$ and $Pr(x = 1) = 0.2$.

- GREEN: $Pr(x = 0) = 0.5$ and $Pr(x = 1) = 0.5$.
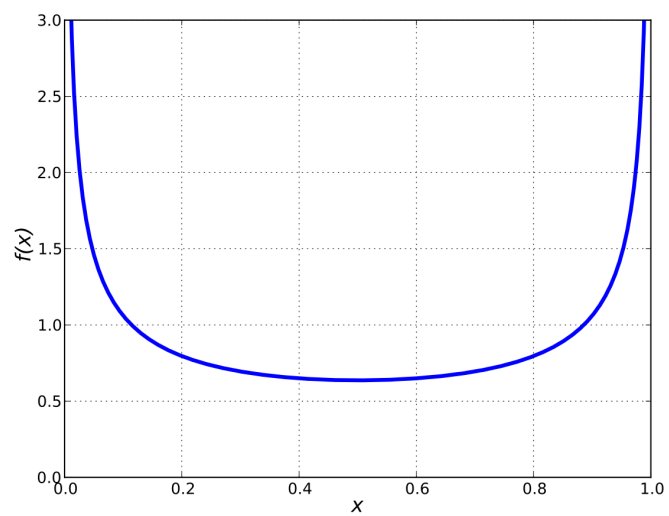
3.1.2 Binomial distribution

3.2 Continuous

3.2.1 Beta distribution

3.2.2 Arcsine distribution

References

[1] https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=
    terms-statistical-distribution

[2] https://datasciencedojo.com/blog/types-of-statistical-distributions-in-ml/
    #Common_types_of_data

[3] https://www.questionpro.com/blog/discrete-data-vs-continuous-data/#:~:text=
    Discrete%20data%20is%20information%20that,that%20falls%20inside%20a%20range.

[4] https://en.wikipedia.org/wiki/List_of_probability_distributions