Th5: Statistical Distributions

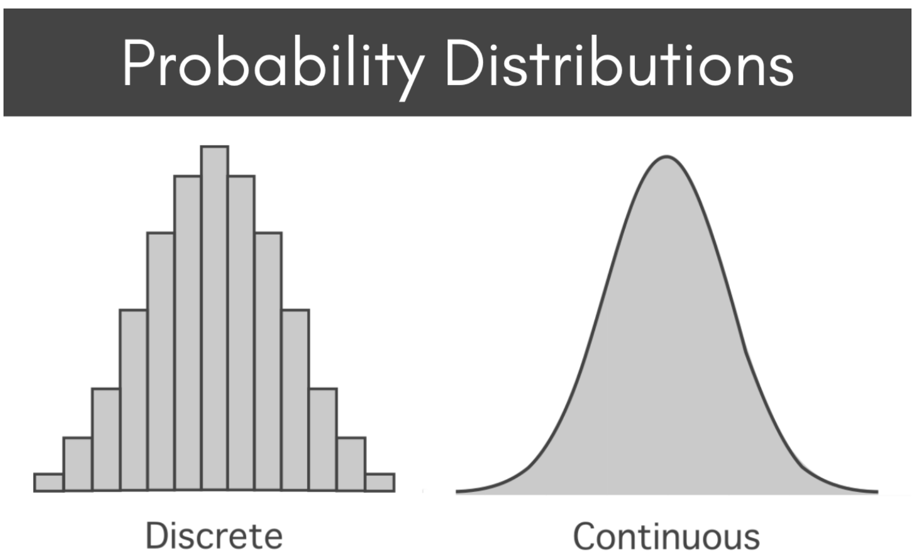
Federico Detomaso

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



2.1 Discrete Distribution

A discrete distribution describes the probability of occurrence of each value of a discrete random variable. A discrete random variable is a random variable that has countable values, such as a list of non-negative integers.

With a discrete probability distribution, each possible value of the discrete random variable can be associated with a non-zero probability. Thus, a discrete probability distribution is often presented in tabular form.

2.1.1 Properties

- **Probability Mass Function (PMF):**
  - Non-negativity: $P(X = x) \geq 0$ for all x in the sample space.
  - Summation: $\sum_x P(X = x) = 1$, where the sum is over all possible values of X.
- **Cumulative Distribution Function (CDF):**
  - Monotonicity: $F(x)$ is non-decreasing.
  - Limits: $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.
- **Expected Value and Variance:**
  - Expected Value: $E(X) = \sum_x x \cdot P(X = x)$.
  - Variance: $\mathrm{Var}(X) = \sum_x (x - E(X))^2 \cdot P(X = x)$.
- **Independence:**
  - Two random variables X and Y are independent if $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ for all x and y in their respective sample spaces.

2.2 Continuous Distribution

A continuous distribution describes the probabilities of the possible values of a continuous random variable. A continuous random variable is a random variable with a set of possible values (known as the range) that is infinite and uncountable.

Probabilities of continuous random variables (X) are defined as the area under the curve of its PDF. Thus, only ranges of values can have a nonzero probability. The probability that a continuous random variable equals some value is always zero.

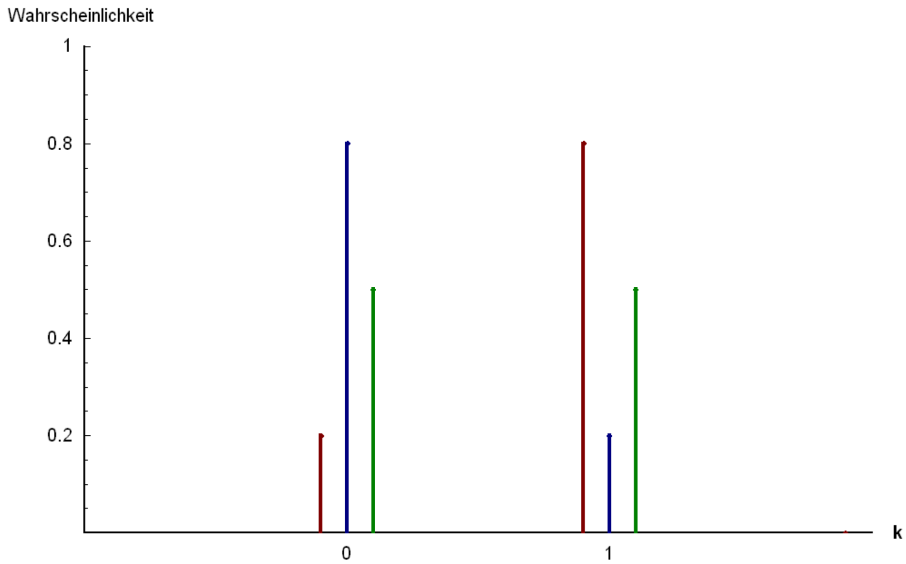
2.2.1 Properties

- **Probability Density Function (PDF):**
  - Non-negativity: $f(x) \geq 0$ for all x in the sample space.
  - Integration: $\int_{-\infty}^{\infty} f(x)\, dx = 1$ over the entire sample space.
- **Cumulative Distribution Function (CDF):**
  - Monotonicity: $F(x)$ is non-decreasing.
  - Limits: $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.
- **Expected Value and Variance:**
  - Expected Value: $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$.
  - Variance: $\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x)\, dx$.
- **Independence:**
  - Two random variables X and Y are independent if $P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$ for all x and y in their respective sample spaces.

3 Simulations

3.1 Discrete

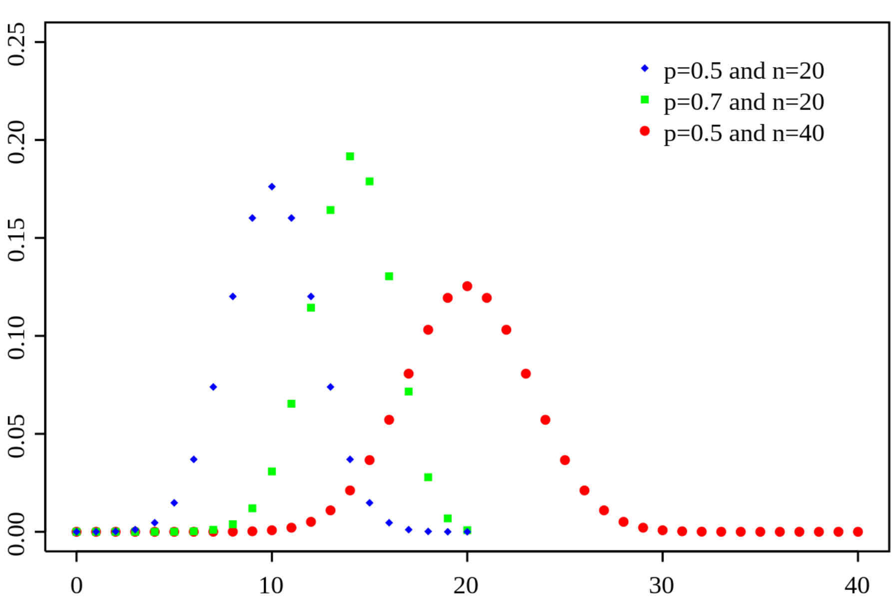
3.1.1 Bernoulli distribution



Three examples of Bernoulli distribution:

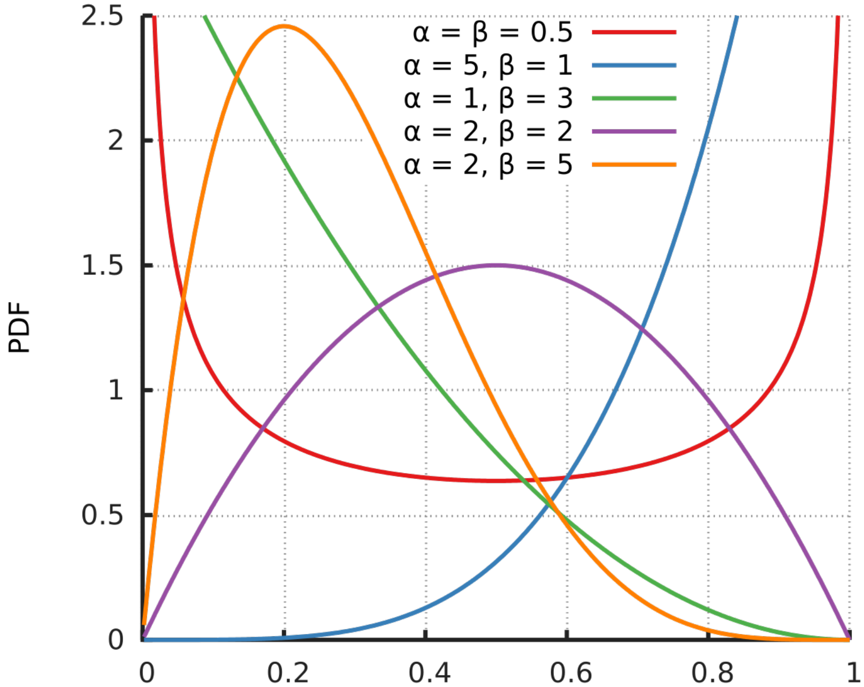
- RED: $Pr(x = 0) = 0.2$ and $Pr(x = 1) = 0.8$.
- BLUE: $Pr(x = 0) = 0.8$ and $Pr(x = 1) = 0.2$.
- GREEN: $Pr(x = 0) = 0.5$ and $Pr(x = 1) = 0.5$.

3.1.2 Binomial distribution

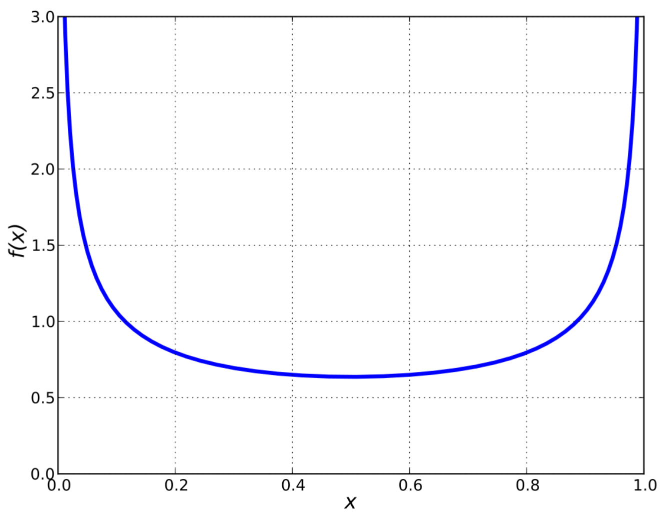


3.2 Continuous

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.investopedia.com/terms/d/discrete-distribution.asp#toc-discrete-distribution-vs-continous-distribution

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