

Monte Carlo simulation

Christian, Jonathan, Marcus, Puk & Sofia

March 19, 2025

Contents

| | | |
|-----------|---------------------------------------|-----------|
| 1 | Introduction | 3 |
| 2 | Statistical theory | 4 |
| 2.1 | Probability space | 4 |
| 2.2 | Probability distribution | 5 |
| 2.2.1 | Normal distribution | 5 |
| 2.2.2 | The central limit theorem | 5 |
| 2.2.3 | The t-distribution | 6 |
| 2.3 | Statistical methods | 6 |
| 3 | Polynomial regression | 7 |
| 3.1 | Assumptions | 7 |
| 4 | Pseudo random number generator | 8 |
| 5 | Monte Carlo | 9 |
| 5.1 | Assumptions | 9 |
| 6 | Problem statement | 10 |
| 7 | Classical regression | 11 |
| 8 | Monte Carlo regression | 12 |
| 9 | Comparison between regressions | 13 |
| 10 | Discussion | 14 |
| 11 | Conclusion | 15 |
| 12 | Litteratur | 16 |

1 Introduction

2 Statistical theory

2.1 Probability space

The **sample space**, S , is the set of all possible outcomes.

If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a **discrete sample space**.

Example: When rolling a standard six-sided die form the discrete sample space, the possible outcomes are $S = 1, 2, 3, 4, 5, 6$

If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a **continuous sample space**.

Example: Measuring the heights of people in a population. This is a continuous sample space, because height can take any real value within a given range.

An **event** is a subset, $A \subseteq S$, of the sample space. The event is the amount that contains all possible events. An example of a discrete event could be rolling a die and getting an uneven number, this would be the event $A = 1, 3, 5$.

For the continuous event, it could be that a person is between 160 cm and 170 cm tall.

The probability of an event A , $P(A)$, is the sum of the weights of all sample points in A . The probability of the whole sample space is 1, $P(S) = 1$ The probability of any event being between 0 and 1, $0 < P(A) < 1$ The probability of the empty set being 0, $P(\emptyset) = 0$

Probability of mutually exclusive events

If A and B are mutually exclusive, $A \cap B = \emptyset$, then

$$P(A \cup B) = P(A) + P(B)$$

Where A and B never occur at the same time, so their union is equal to the two events added together.

Probability of union

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Here, the union of the two events is A added to B , but minus their common event, since it otherwise would be added twice.

Two events A and B are independent, if

$$P(A|B) = P(A)$$

The equivalent definition to this is:

Two events A and B are independent if and only if

$P(A \cap B) = P(A)P(B)$ This says that the probability of both event A and B happening, is equal to the product of the two events.

2.2 Probability distribution

Data can come in various distributions depending on different parameters such as degrees of freedom. The distribution is the shape of the data and it will have an effect on statistical models. Therefore it is important to have an understanding of distributions.

2.2.1 Normal distribution

In the world of statistics, the most common distribution is the normal distribution. It is constructed as a bell shape. The normal distribution is a continuous distribution, with this density function:

$$n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The distribution is dependent on the mean(μ) and the standard deviation(σ), where changes to the mean will result in a change in the positioning of the normal distribution. Whereas a change in the standard deviation will change the spread of the curve. The normal distribution also always contains an area under the curve that is equal to one. This is to ensure that the normal distribution correctly models probability.

There is a special case of the normal distribution, called the standard normal distribution, where the mean is zero and the standard deviation is one. All variations of a normal distribution can be standardized by a transformation of the distribution, using the Z-score formula.

$$Z = \frac{X - \mu}{\sigma}$$

Z in the Z-score represents the amount of standard deviations a given X value, deviates from the mean.

2.2.2 The central limit theorem

A very effective theorem in statistics is the central limit theorem. This theorem states that if a random sample \bar{X} , with the size n , is taken from a population with a mean and a finite variance, then as n goes towards infinity, the distribution will resemble a normal distribution. If used with the Z-score formula, the distribution will resemble a standard normal distribution. The formula for the Z-score, when in conjunction with the central limit theorem, looks like this:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Where \bar{X} is a random sample of size n and μ is the mean of the true population. The standard error is represented by σ/\sqrt{n} , where σ is the standard deviation and n is the sample size. Usually the standard deviation is unknown, for these

situations it's possible to use the estimator S^2 . This estimates the variance of the population from the variance of the sample, by this formula:

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$$

The square root of the variance is the standard deviation, therefore the square root of the estimator S^2 would be the estimated standard deviation. The problem with using the estimator S^2 , is that with small samples the variance is small and therefore it contains a lot of bias. In this situation the t-distribution would be used instead of the normal distribution, because the t-distribution takes the bias into account the bias of the standard deviation. It does this by having thicker tails, meaning that the probability of more extreme values are higher.

2.2.3 The t-distribution

The t-distribution is shaped as the standard normal distribution, in a bell shape and symmetrical around the mean of zero, the difference is that the t-distribution is more variable. This comes from the fact that the t-distribution is dependent on the degrees of freedom. When the degrees of freedom surpasses 30, the rule of thumb is that the distribution will resemble a normal distribution. So before 30 degrees of freedom, the distribution contains more variance. The t-distribution will come to resemble the standard normal distribution, when it surpasses 30 degrees of freedom, this makes sense, since the two distributions have the same formula:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

The only difference is the estimated standard deviation S .

2.3 Statistical methods

2.3.1 Hejsa

3 Polynomial regression

3.1 Assumptions

4 Pseudo random number generator

5 Monte Carlo

5.1 Assumptions

6 Problem statement

7 Classical regression

8 Monte Carlo regression

9 Comparison between regressions

10 Discussion

11 Conclusion

12 Litteratur