

Seminar 1: Basics of Combinatorics

Carlos Buitrago & Anastasiya Chikalova

Table of contents

1	Fundamental Counting Principles	2
1.1	Sum Rule	2
1.1.1	Example	2
1.2	Product Rule	2
1.2.1	Example	2
1.3	Pigeonhole Principle	3
1.3.1	Example	3
1.4	Counting Formulas	3
1.4.1	How Fast Do Binomial Coefficients Grow?	3
1.4.2	How Many Possible Feature Subsets?	4
1.5	Binomial Theorem	4
2	Fundamental Binomial Identities	4
2.0.1	Symmetry Identity	4
2.0.2	Pascal Identity	4
2.0.3	Combinations with Repetition	4
2.0.4	Sum of Binomial Coefficients	4
2.0.5	Sum of Squares	5
2.0.6	Hockey-Stick Identity	5
2.1	Useful Consequences	5
3	Inclusion–Exclusion Principle	5
3.1	Derangements	5
3.1.1	Inclusion–Exclusion Formula	6
4	Why Combinatorics Matters in Probability	6
5	Monte Carlo Simulation	7
5.1	Basic Idea	7
5.2	Why Does It Work?	8

5.3 Example: Lucky Tram Tickets	8
6 The Matching Problem	9
7 Estimating π with Monte Carlo Simulation	10
8 Random Walk Simulation	12

1 Fundamental Counting Principles

1.1 Sum Rule

Suppose that a task can be performed in n_1 ways or in n_2 ways, and that the two possibilities are mutually exclusive. Then the total number of possible outcomes equals $n_1 + n_2$.

1.1.1 Example

A class consists of 15 boys and 12 girls. If we wish to choose a class representative, then the number of possible choices is

$$15 + 12 = 27.$$

1.2 Product Rule

Suppose that a procedure consists of k consecutive steps. If Step i can be performed in n_i ways for every $i = 1, \dots, k$, then the total number of possible outcomes equals

$$n_1 n_2 \cdots n_k.$$

1.2.1 Example

A password consists of three letters followed by two digits. Since each letter can be chosen in 26 ways and each digit in 10 ways, the total number of possible passwords equals

$$26^3 \cdot 10^2$$

1.3 Pigeonhole Principle

If $n + 1$ objects are placed into n boxes, then at least one box must contain at least two objects.

1.3.1 Example

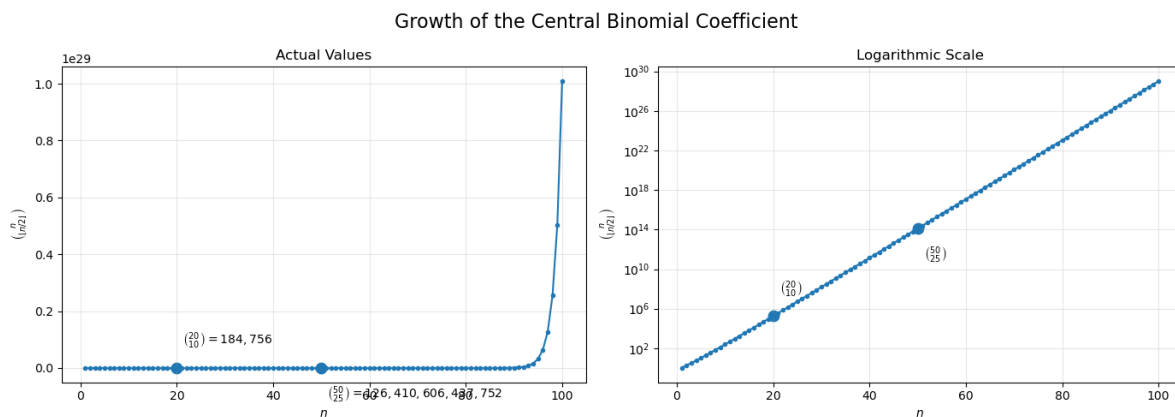
Among 13 people there are always at least two born in the same month. Indeed, the objects are the people and the boxes are the months of the year. Since $13 > 12$, the conclusion follows immediately.

1.4 Counting Formulas

	With Repetition	Without Repetition
Combinations	$\binom{n+k-1}{k}$	$\binom{n}{k}$
Arrangements	n^k	$\frac{n!}{(n-k)!}$

1.4.1 How Fast Do Binomial Coefficients Grow?

How large is the central binomial coefficient $\binom{n}{\lfloor n/2 \rfloor}$ for increasing values of n ?



$$C(20, 10) = 184,756$$

$$C(50, 25) = 126,410,606,437,752$$

$$C(100, 50) = 100,891,344,545,564,193,334,812,497,256$$

1.4.2 How Many Possible Feature Subsets?

Suppose a dataset has 100 features.

How many possible subsets of features exist?

Number of subsets = 1.267651×10^{30}

$\log_{10}(\text{Number of subsets}) = 30.10$

Even if a computer evaluated one billion feature subsets per second, it would still take much longer than the age of the universe to check them all.

1.5 Binomial Theorem

For every $n \in \mathbb{Z}_+$,

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}.$$

2 Fundamental Binomial Identities

2.0.1 Symmetry Identity

$$\binom{n}{k} = \binom{n}{n-k}.$$

2.0.2 Pascal Identity

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$

2.0.3 Combinations with Repetition

$$\binom{n+k-1}{k}.$$

2.0.4 Sum of Binomial Coefficients

$$\sum_{k=0}^n \binom{n}{k} = 2^n.$$

2.0.5 Sum of Squares

$$\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}.$$

2.0.6 Hockey-Stick Identity

$$\binom{n+m}{n} = \sum_{i=0}^m \binom{n+i-1}{n-1}.$$

2.1 Useful Consequences

From the Hockey-Stick Identity we obtain

$$\sum_{i=1}^n i = \frac{n(n+1)}{2}$$

and

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}.$$

3 Inclusion–Exclusion Principle

Suppose that $a_1, \dots, a_N$ are objects and $d_1, \dots, d_n$ are properties.

Let $N(d_i)$ denote the number of objects satisfying property d_i , and let $N(d_{i_1}, \dots, d_{i_k})$ denote the number of objects satisfying all properties $d_{i_1}, \dots, d_{i_k}$ simultaneously.

Then the number of objects satisfying none of the properties is given by

$$N - \sum_i N(d_i) + \sum_{i < j} N(d_i, d_j) - \sum_{i < j < k} N(d_i, d_j, d_k) + \dots + (-1)^n N(d_1, \dots, d_n).$$

3.1 Derangements

A **derangement** of n objects is a permutation in which no object remains in its original position.

Let D_n denote the number of derangements of an n -element set.

3.1.1 Inclusion–Exclusion Formula

Using the Inclusion–Exclusion Principle, one obtains

$$D_n = n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!} \right).$$

Equivalently,

$$D_n = n! \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

4 Why Combinatorics Matters in Probability

In many elementary probability problems, all outcomes are assumed to be equally likely. In that case, the probability of an event A is

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

Here Ω is the sample space, $|\Omega|$ is the total number of possible outcomes, and $|A|$ is the number of favorable outcomes.

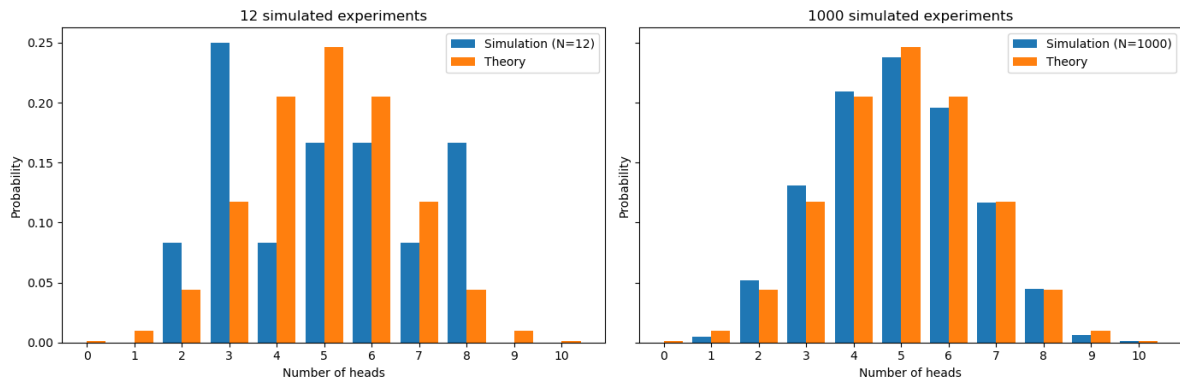
Therefore, probability often begins with counting.

For example, if we toss a fair coin 10 times, then there are 2^{10} possible outcomes. The probability of getting exactly 4 heads is

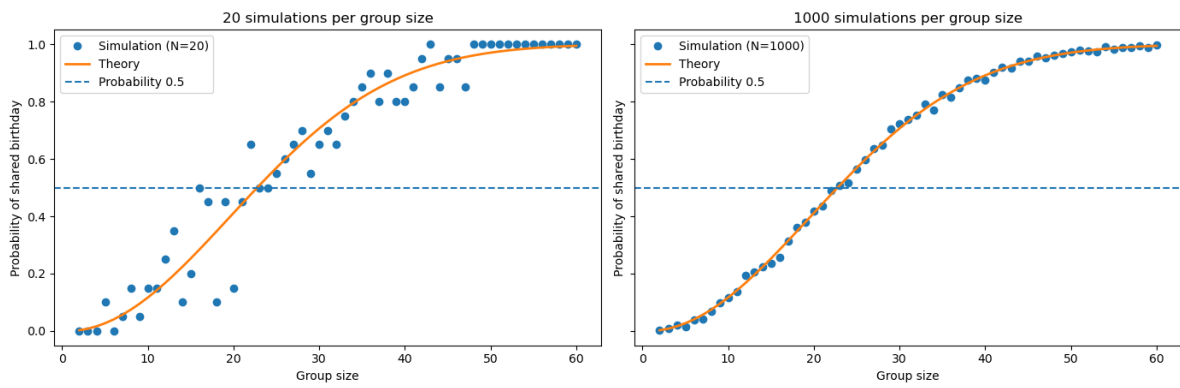
$$\mathbb{P}(\text{exactly 4 heads}) = \frac{\binom{10}{4}}{2^{10}}.$$

Combinatorics tells us how many outcomes are possible. Probability tells us how likely they are.

Binomial Distribution: Simulation vs Theory



Birthday Problem: Simulation vs Theory



5 Monte Carlo Simulation

In many probability problems, obtaining an exact answer can be difficult or even impossible. A powerful alternative is to estimate probabilities using random experiments performed by a computer.

This approach is known as the **Monte Carlo method**.

5.1 Basic Idea

Suppose that we want to estimate the probability of an event A .

If we repeat the experiment independently N times and observe that the event occurs M times, then

$$P(A) \approx \frac{M}{N}.$$

As N becomes large, this approximation becomes increasingly accurate.

This phenomenon is a consequence of the **Law of Large Numbers**, which states that the empirical frequency of an event converges to its true probability.

5.2 Why Does It Work?

Suppose that the true probability of an event is p .

If we perform N independent experiments and count the number of successes M , then

$$\frac{M}{N} \rightarrow p \quad \text{as} \quad N \rightarrow \infty.$$

Thus, by generating enough random samples, we can estimate probabilities with arbitrary precision.

5.3 Example: Lucky Tram Tickets

A six-digit ticket is called **lucky** if the sum of its first three digits equals the sum of its last three digits. Computing the exact probability requires a nontrivial combinatorial argument.

Instead, we can estimate the probability by generating a large number of random tickets and checking how often the event occurs. For example, after simulating one million random tickets, we obtain

$$P(\text{lucky ticket}) \approx 0.055.$$

The exact value is

$$P(\text{lucky ticket}) = \frac{55252}{10^6} = 0.055252.$$

The simulation provides an excellent approximation.

Estimated probability = 0.055233

Estimated number of lucky tickets = 55233

6 The Matching Problem

Suppose that n people check their coats at a restaurant.

At the end of the evening, the coats are returned completely at random.

Let X denote the number of people who receive their own coat.

Questions:

1. What is the expected value of X ?
2. What is the probability that nobody receives their own coat?
3. What does the distribution of X look like for large n ?

A permutation of the coats corresponds to a random assignment. A person receives their own coat if and only if they are a fixed point of the permutation. Thus, the event

$$X = 0$$

corresponds exactly to a derangement. Therefore $P(X = 0) = \frac{D_n}{n!}$.

Since $D_n \sim \frac{n!}{e}$, we obtain

$$P(X = 0) \rightarrow \frac{1}{e} \approx 0.3679.$$

Surprisingly, even for very large groups there is still roughly a 37% chance that nobody gets their own coat.

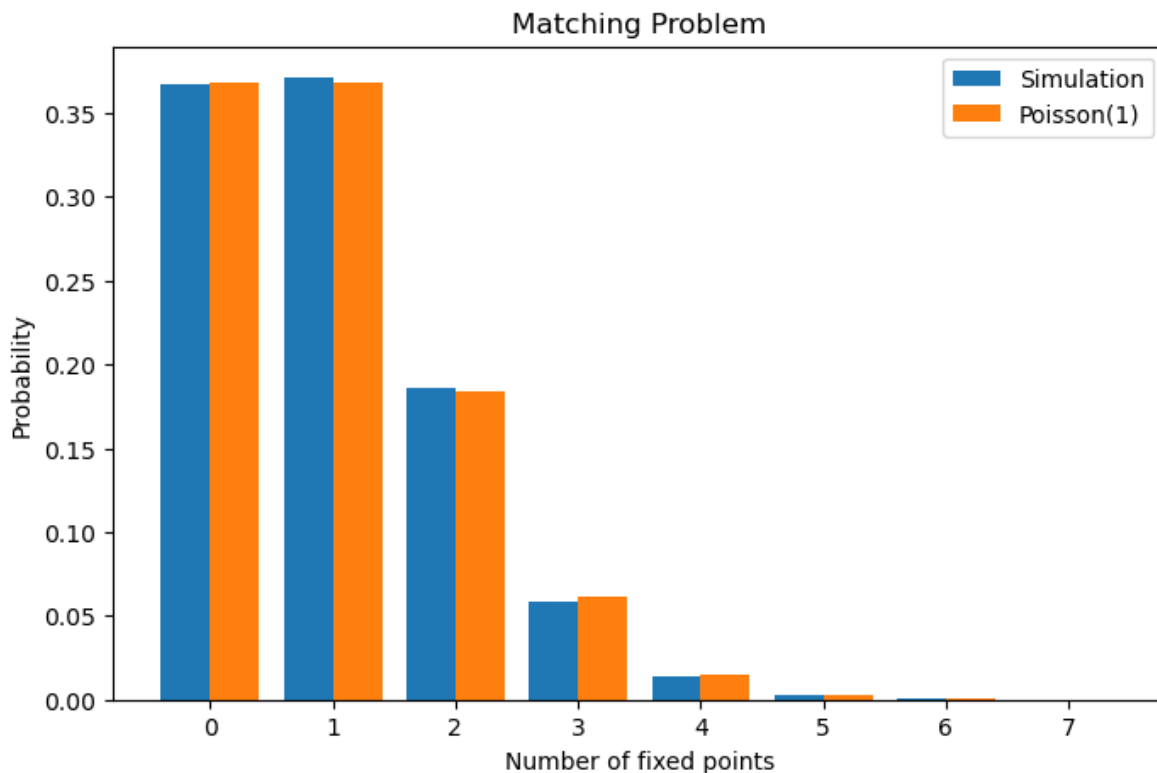
Average number of fixed points: 0.993

The simulation suggests that the average number of people receiving their own coat is approximately

$$E[X] = 1.$$

Remarkably, this is true for every value of n .

Even if one thousand people check their coats, the expected number of correct matches is still exactly one.



The histogram is extremely close to a Poisson distribution with parameter $\lambda = 1$.

In fact, one can prove that

$$X \xrightarrow{d} \text{Poisson}(1) \quad (n \rightarrow \infty).$$

As a consequence,

$$P(X = 0) = e^{-1} \approx 0.3679,$$

which explains why the probability of a derangement converges to $1/e$.

7 Estimating π with Monte Carlo Simulation

Consider the square $[-1, 1] \times [-1, 1]$. Its area is 4. Inside the square, consider the unit disk

$$x^2 + y^2 \leq 1.$$

The area of the disk is π . Therefore,

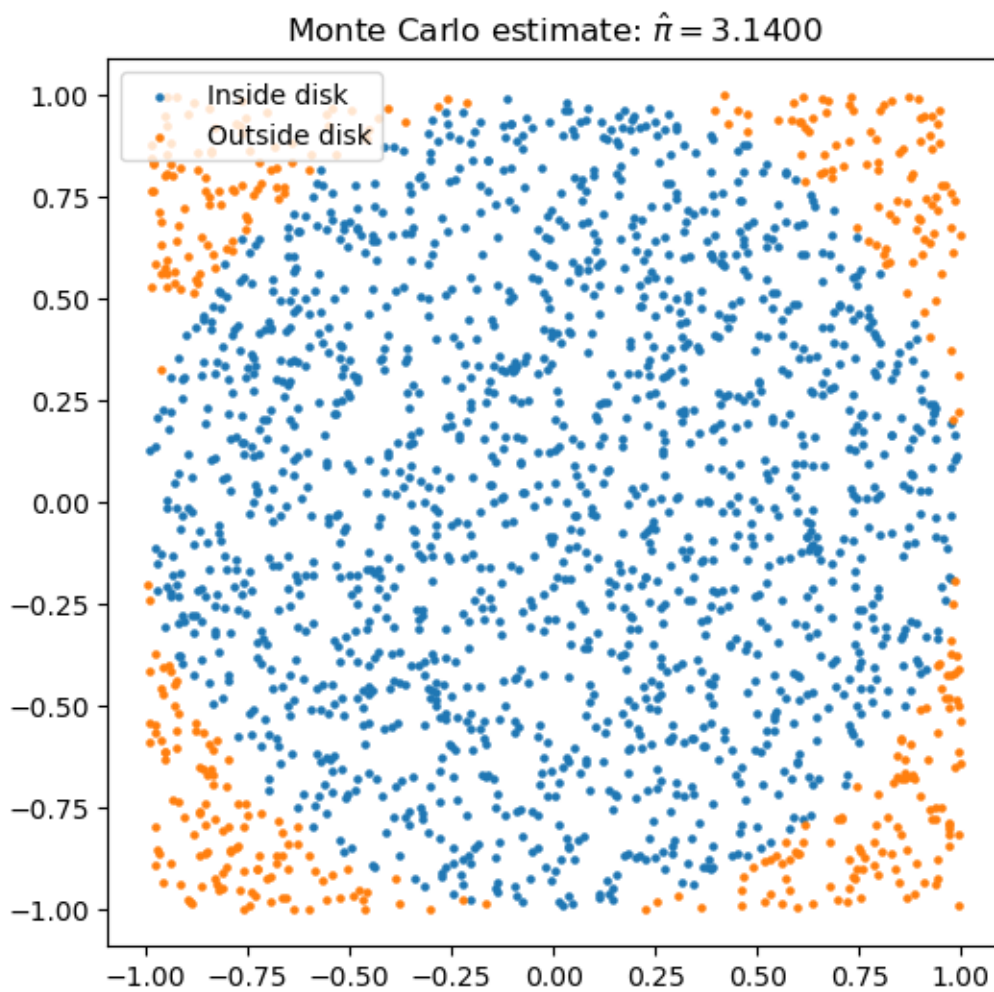
$$\frac{\text{Area of disk}}{\text{Area of square}} = \frac{\pi}{4}.$$

If we generate points uniformly at random inside the square, then $P((X, Y) \in \text{disk}) = \frac{\pi}{4}$.

Consequently,

$$\pi = 4P((X, Y) \in \text{disk}).$$

Thus, we can estimate π by repeatedly generating random points and counting how many fall inside the disk.



Estimate of pi: 3.14

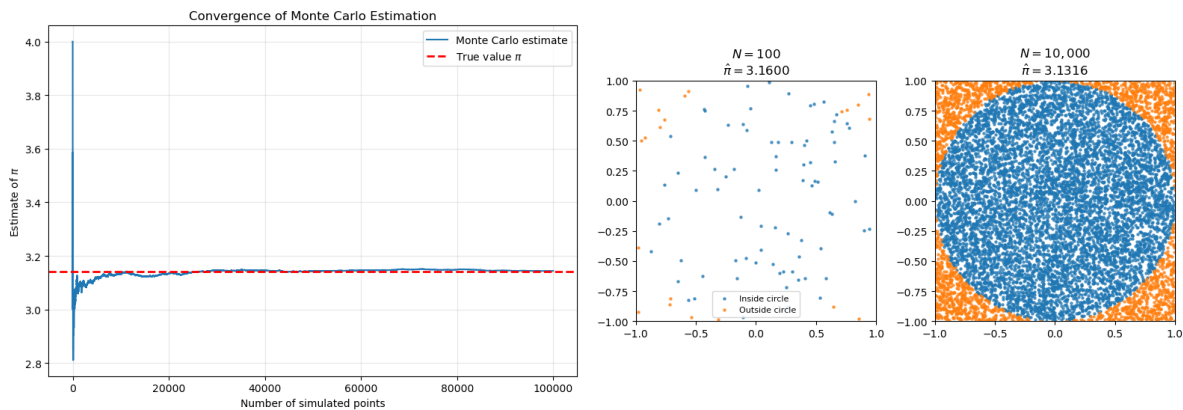
As the number of simulated points increases, the estimate approaches the true value

$$\pi \approx 3.14159.$$

This is another illustration of the Law of Large Numbers.

The Monte Carlo estimate becomes increasingly accurate as the sample size grows.

Monte Carlo Estimation of π



8 Random Walk Simulation

A simple random walk starts at position 0.

At each step, the particle moves

- one unit to the right with probability $\frac{1}{2}$;
- one unit to the left with probability $\frac{1}{2}$.

If the steps are denoted by $X_1, X_2, \dots, X_n$, where each X_i is either +1 or -1, then the position after n steps is

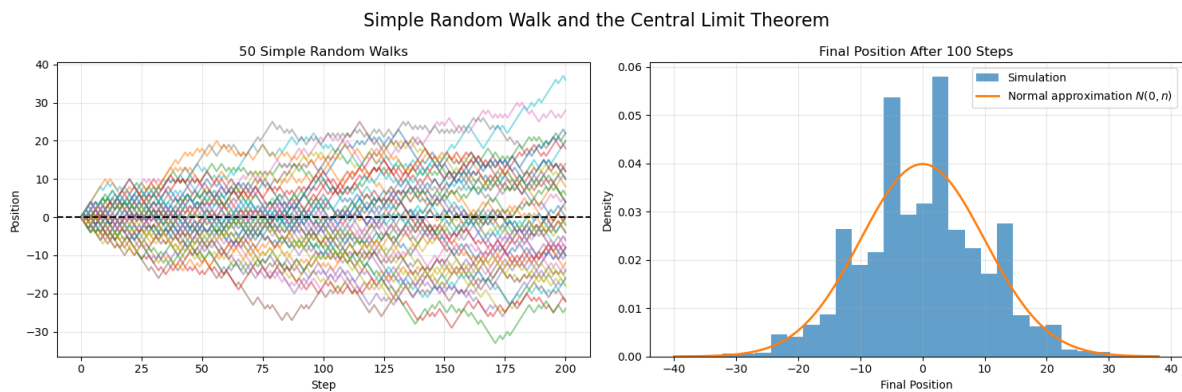
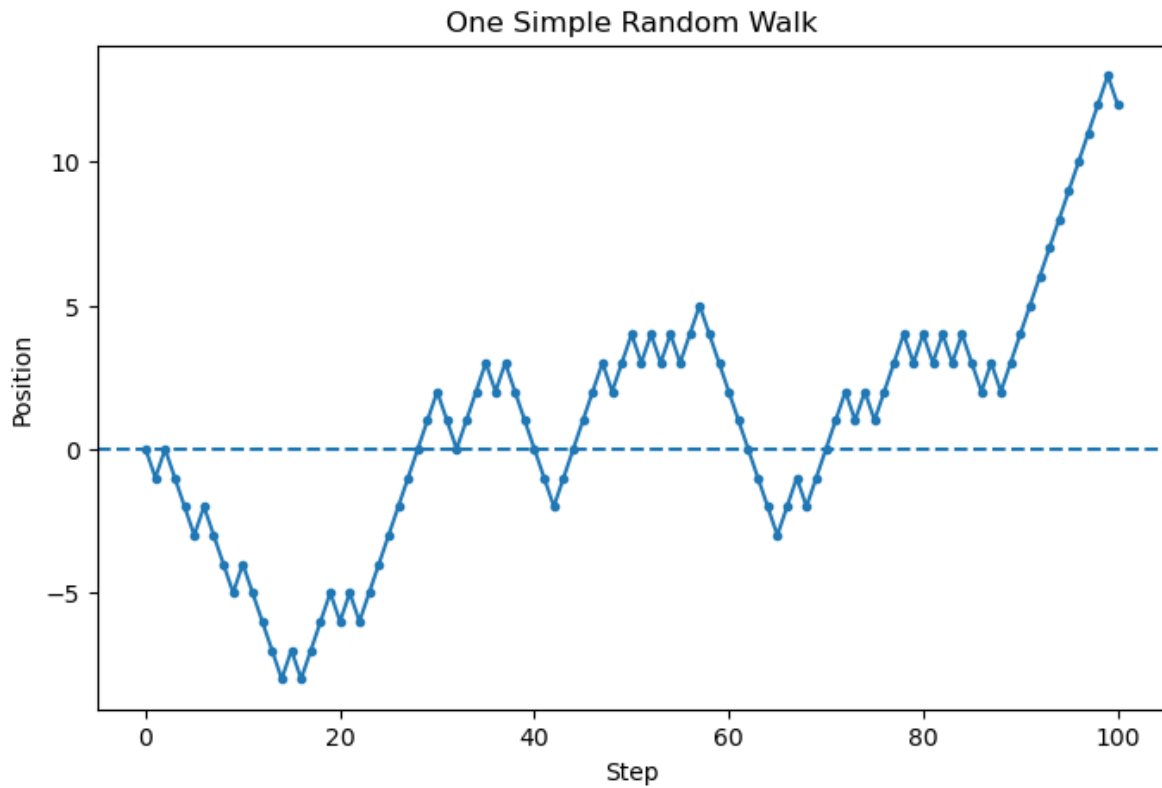
$$S_n = X_1 + \dots + X_n.$$

This model appears in many areas:

- gambling;
- stock prices;

- diffusion;
- physics;
- Markov chains;
- stochastic processes.

The random walk is one of the simplest examples of a stochastic process.



Although each walk is random, we can already observe some structure.

Most paths stay relatively close to 0, but some paths move far away.

After n steps, the typical distance from the origin is not of order n , but rather of order $\sqrt{n}$.

This is one of the first hints of the Central Limit Theorem.