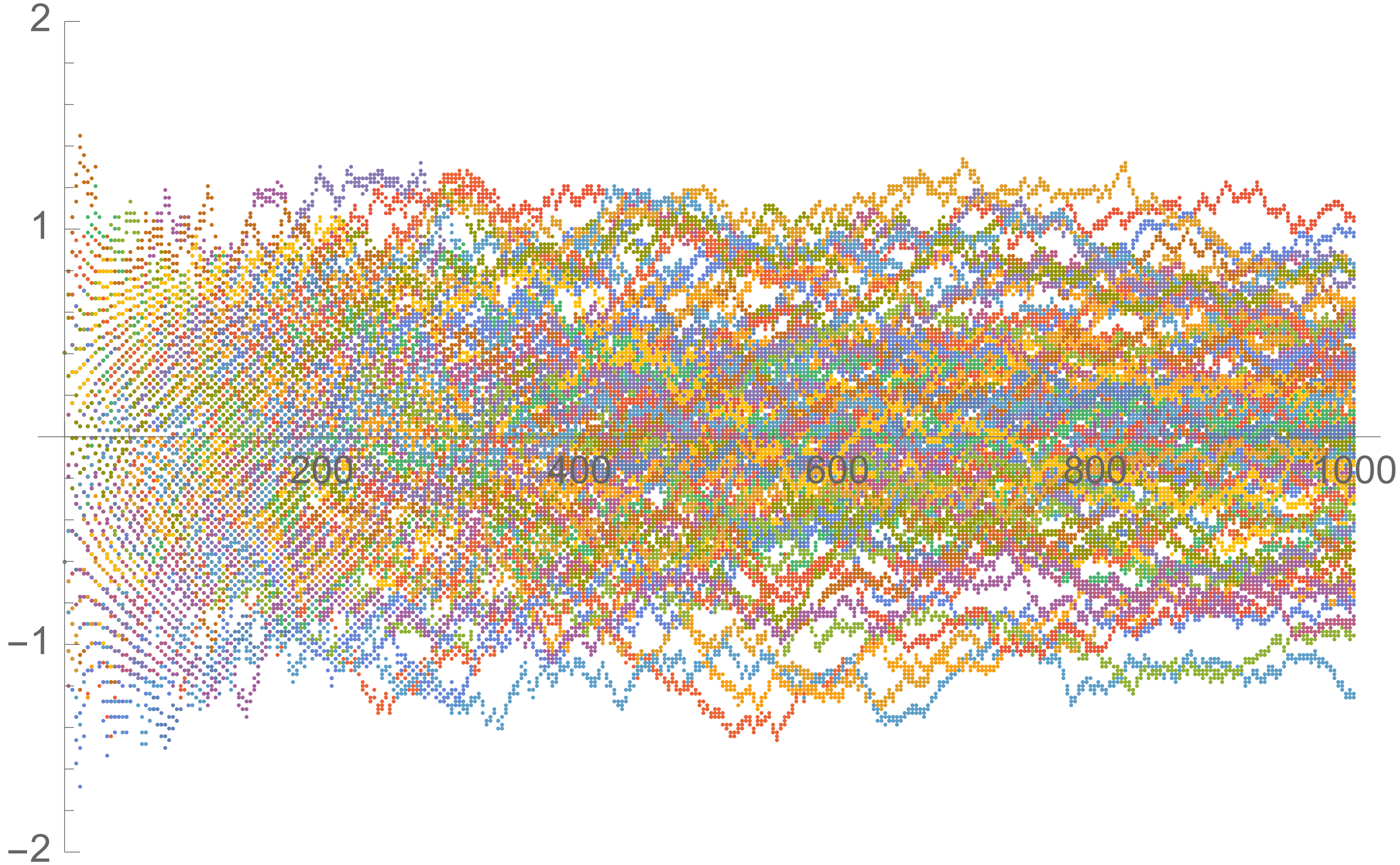
Structure in Randomness



Discrete mathematics MAA103

I. Law of large numbers

II. Central limit theorem

The law of large numbers

Fix $p \in (0, 1)$. Throw n times in a row a coin which has a probability p of giving heads.

The law of large numbers

Fix $p \in (0, 1)$. Throw n times in a row a coin which has a probability p of giving heads.

As $n \to \infty$, how does evolve the proportion of heads?

The law of large numbers

A bit more formally, for $i \geqslant 1$ set $X_i = 1$ if the i-th throw is heads (happens with probability p) and 0 otherwise (happens with probability $1-p$). Set $S_n = X_1 + \cdots + X_n$.

The law of large numbers

A bit more formally, for $i \geqslant 1$ set $X_i = 1$ if the i-th throw is heads (happens with probability p) and 0 otherwise (happens with probability $1-p$). Set $S_n = X_1 + \cdots + X_n$.

⤳ What is the behavior of $\frac{S_n}{n}$ as $n \to \infty$?

The law of large numbers

A bit more formally, for $i \geqslant 1$ set $X_i = 1$ if the i-th throw is heads (happens with probability p) and 0 otherwise (happens with probability $1-p$). Set $S_n = X_1 + \cdots + X_n$.

↝ What is the behavior of $\frac{S_n}{n}$ as $n \to \infty$?

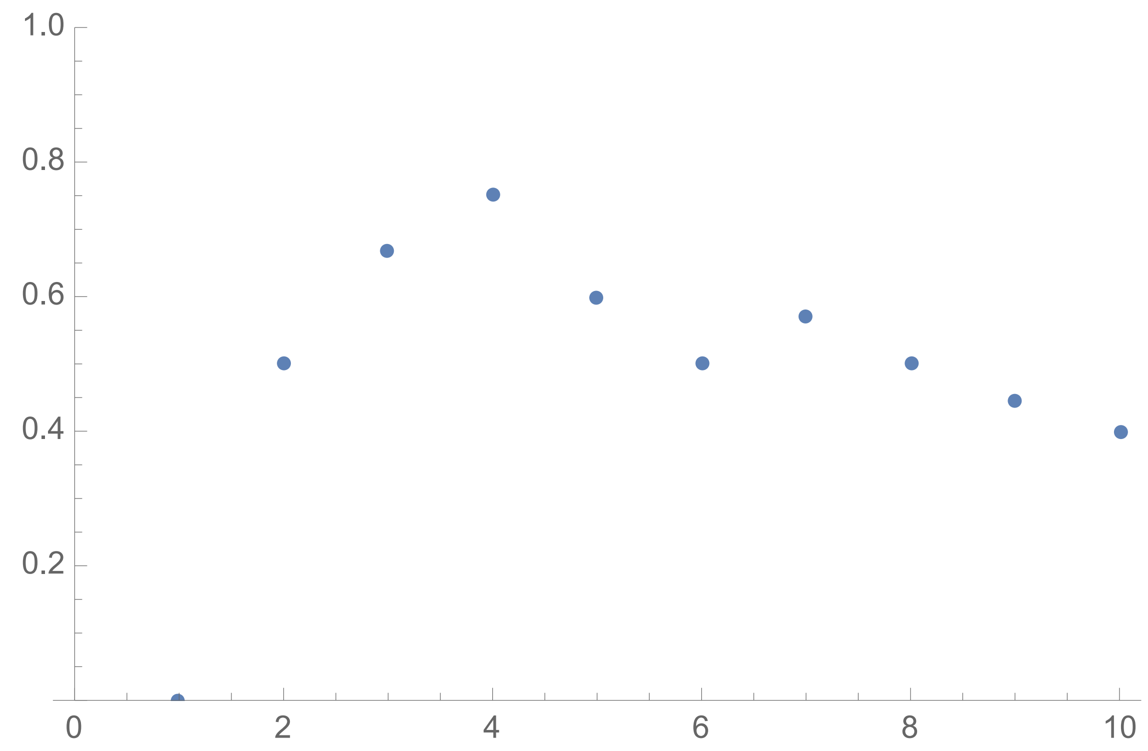


Figure: Simulation of $\left(\frac{S_n}{n} : 1 \leqslant n \leqslant 10\right)$.

The law of large numbers

A bit more formally, for $i \geqslant 1$ set $X_i = 1$ if the i-th throw is heads (happens with probability p) and 0 otherwise (happens with probability $1-p$). Set $S_n = X_1 + \cdots + X_n$.

↝ What is the behavior of $\frac{S_n}{n}$ as $n \to \infty$?

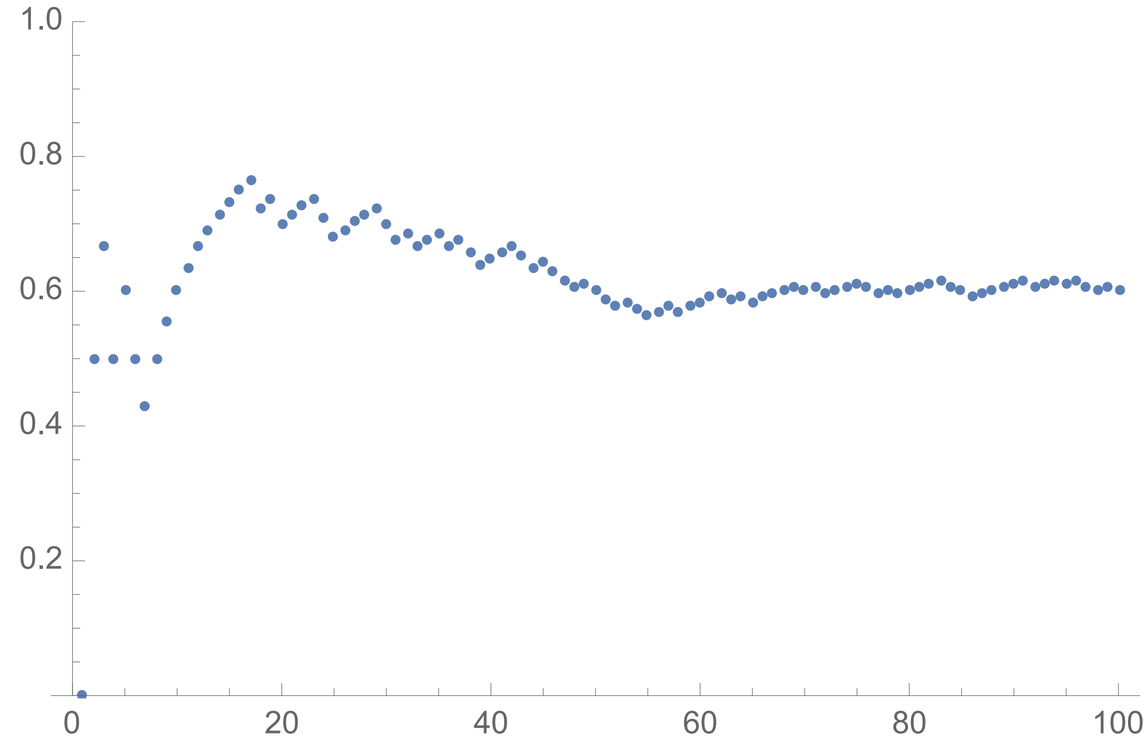


Figure: Simulation of $\left(\frac{S_n}{n} : 1 \leqslant n \leqslant 100\right)$.

The law of large numbers

A bit more formally, for $i \geqslant 1$ set $X_i = 1$ if the i-th throw is heads (happens with probability p) and 0 otherwise (happens with probability $1 - p$). Set $S_n = X_1 + \cdots + X_n$.

↝ What is the behavior of $\frac{S_n}{n}$ as $n \to \infty$?

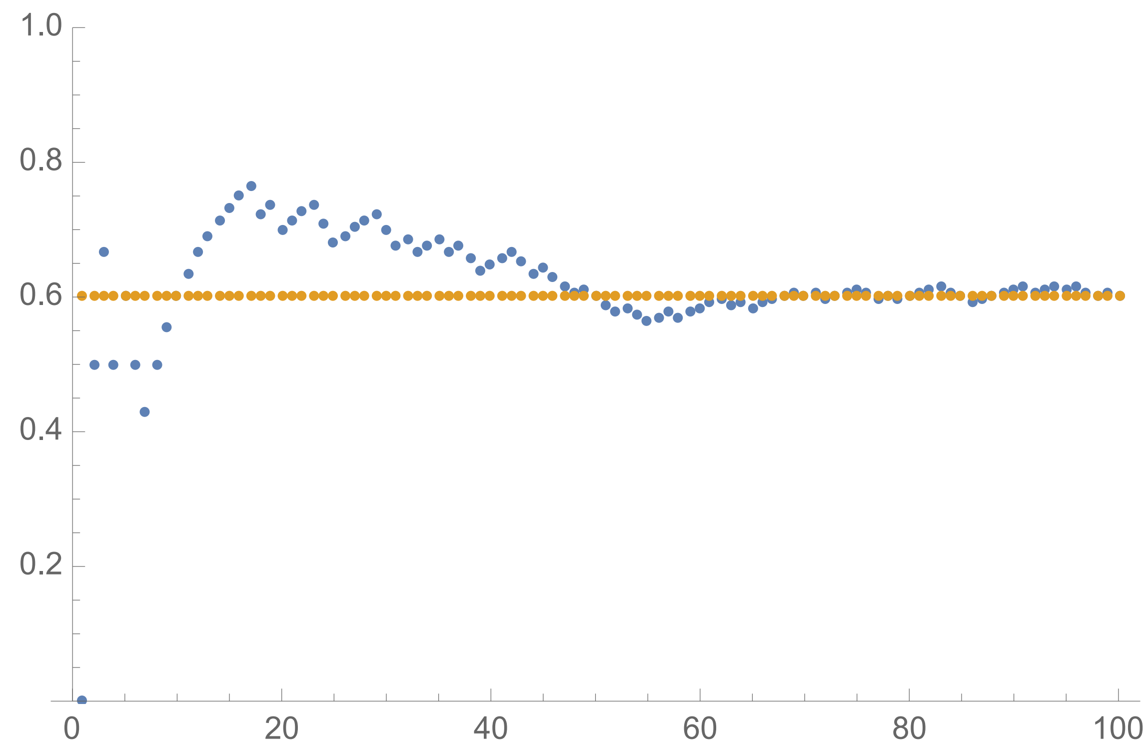


Figure: Simulation of $\left(\frac{S_n}{n} : 1 \leqslant n \leqslant 100\right)$.

The law of large numbers

A bit more formally, for $i \geqslant 1$ set $X_i = 1$ if the i-th throw is heads (happens with probability p) and 0 otherwise (happens with probability $1 - p$). Set $S_n = X_1 + \cdots + X_n$.

↝ What is the behavior of $\frac{S_n}{n}$ as $n \to \infty$?

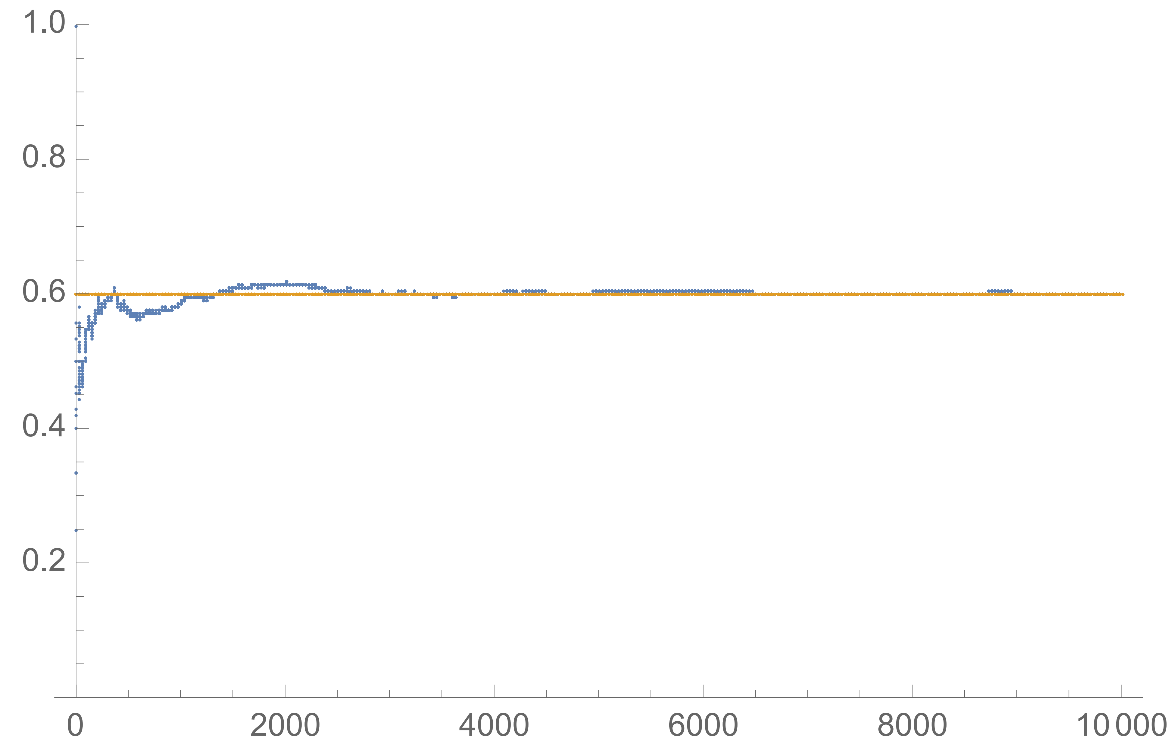


Figure: Simulation of $\left(\frac{S_n}{n} : 1 \leqslant n \leqslant 10000\right)$ for $p = 0.6$.

The law of large numbers

A bit more formally, for $i \geqslant 1$ set $X_i = 1$ if the i-th throw is heads (happens with probability p) and 0 otherwise (happens with probability $1-p$). Set $S_n = X_1 + \cdots + X_n$.

- What is the behavior of $\frac{S_n}{n}$ as $n \to \infty$?

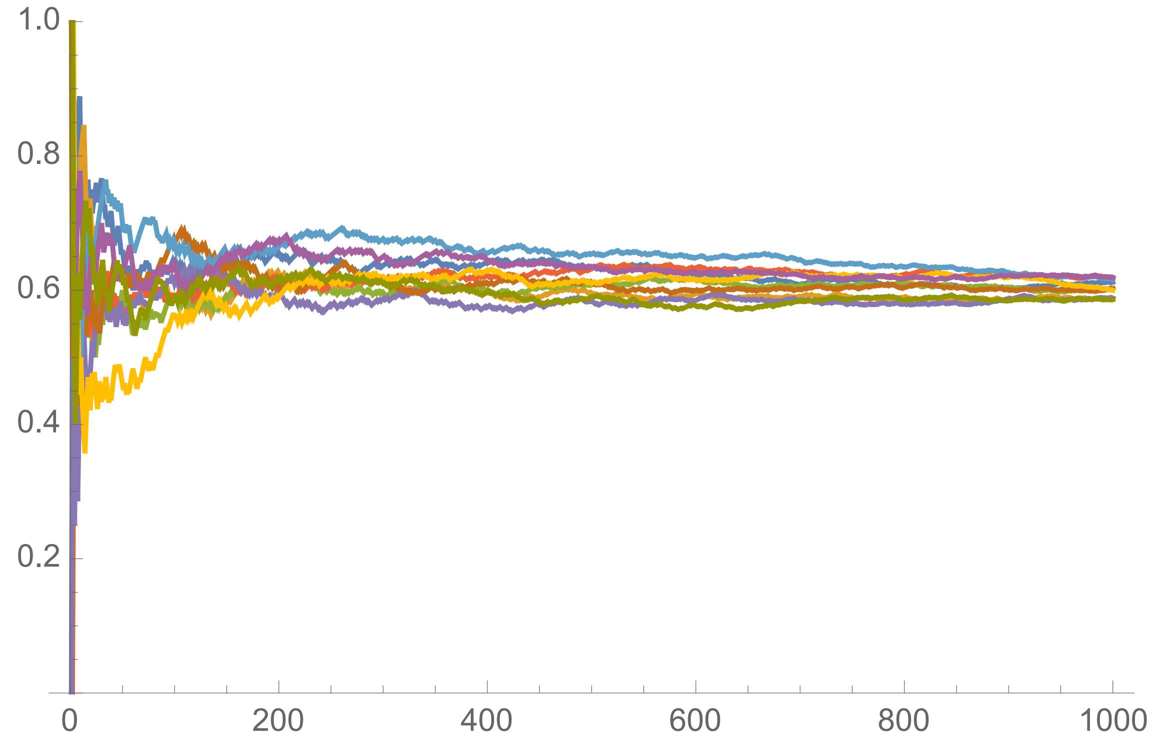


Figure: 10 simulations of $\left(\frac{S_n}{n} : 1 \leqslant n \leqslant 1000\right)$ for $p = 0.6$.

The law of large numbers

A bit more formally, for $i \geqslant 1$ set $X_i = 1$ if the i-th throw is heads (happens with probability p) and 0 otherwise (happens with probability $1-p$). Set $S_n = X_1 + \cdots + X_n$.

- What is the behavior of $\frac{S_n}{n}$ as $n \to \infty$?

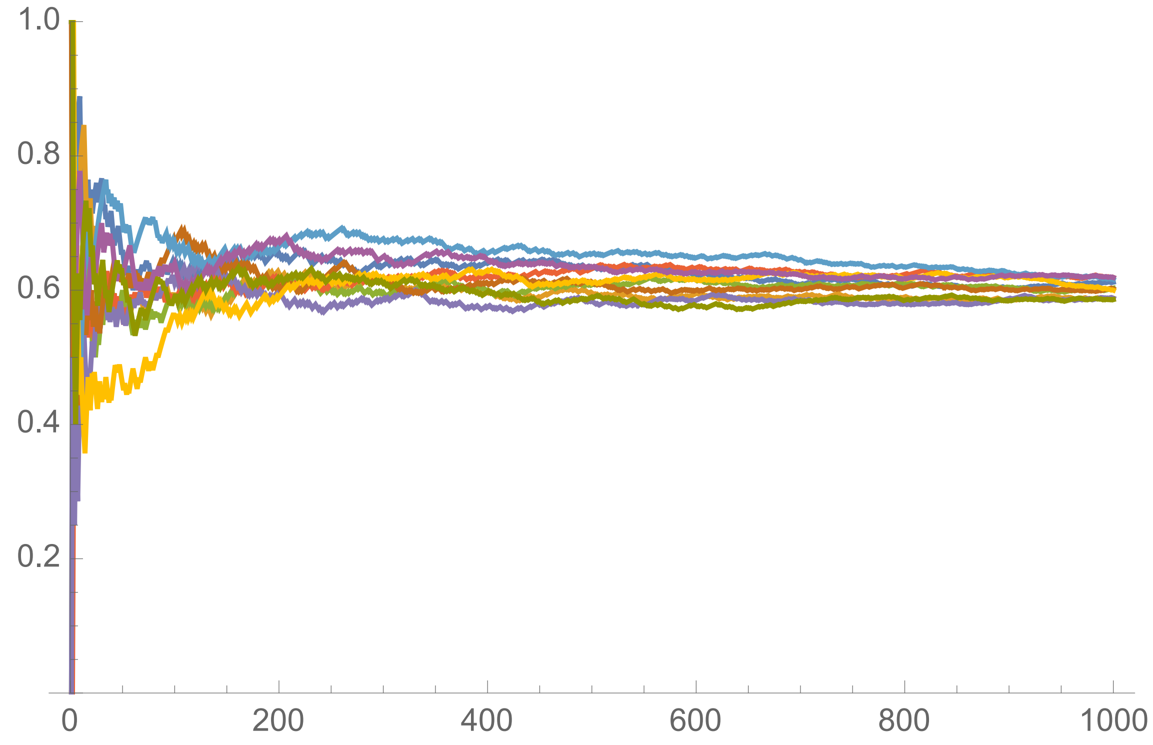


Figure: 10 simulations of $\left(\frac{S_n}{n} : 1 \leqslant n \leqslant 1000\right)$ for $p = 0.6$.

- Law of large numbers: $\frac{S_n}{n}$ converges **almost surely** towards p as $n \to \infty$.

I. Law of large numbers

II. Central limit theorem

The Central Limit Theorem

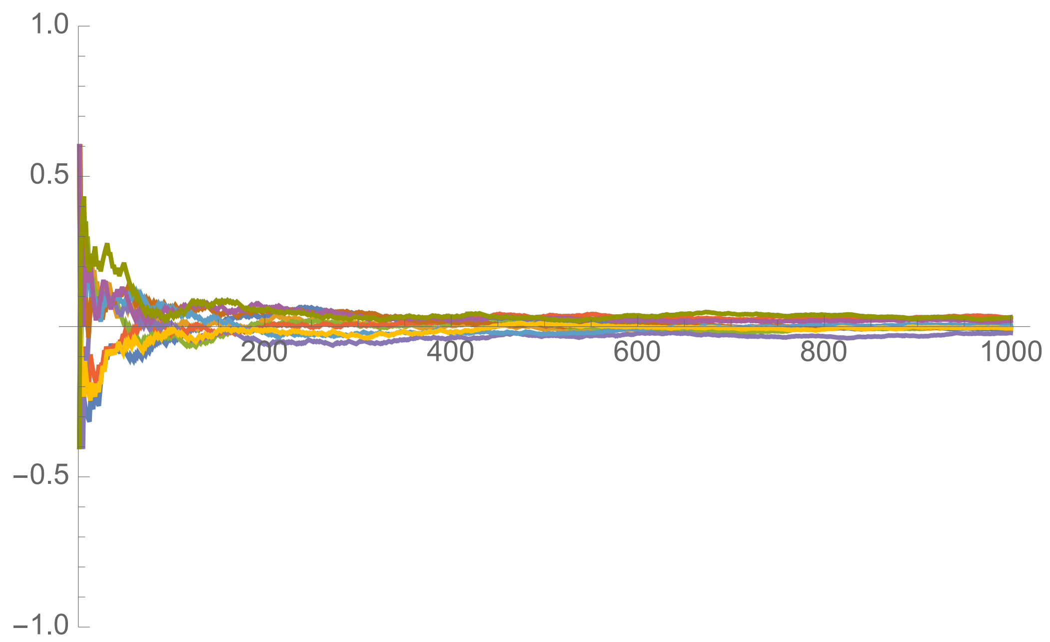


Figure: 10 simulations of $\left(\frac{S_n}{n} - p : 1 \leqslant n \leqslant 1000\right)$ for $p = 0.6$.

The Central Limit Theorem

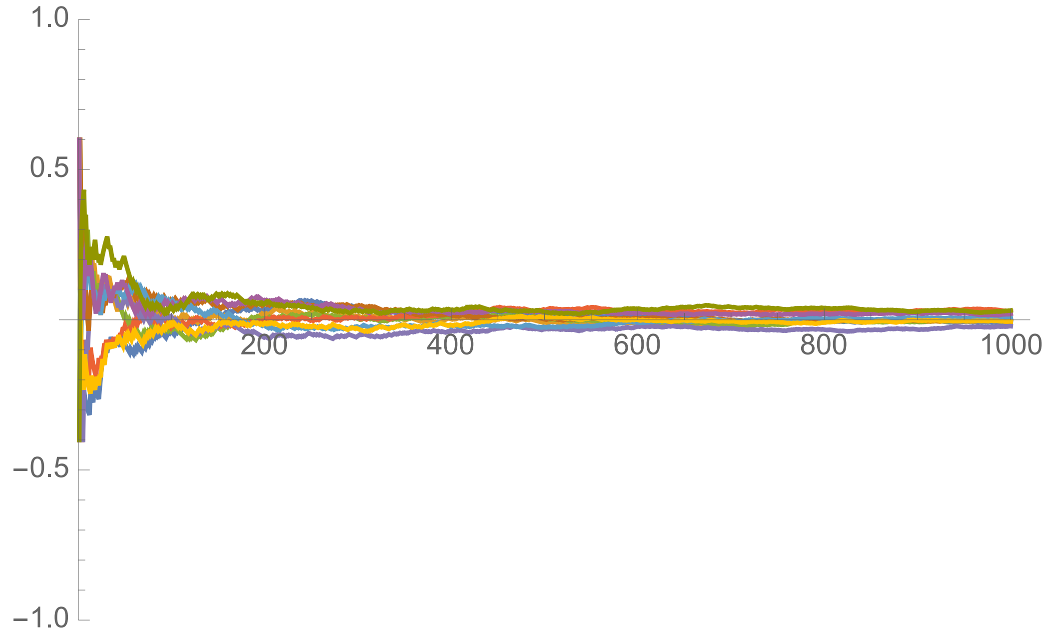


Figure: 10 simulations of $\left(\frac{S_n}{n} - p : 1 \leqslant n \leqslant 1000\right)$ for $p = 0.6$.

Can we "zoom in"?

The Central Limit Theorem

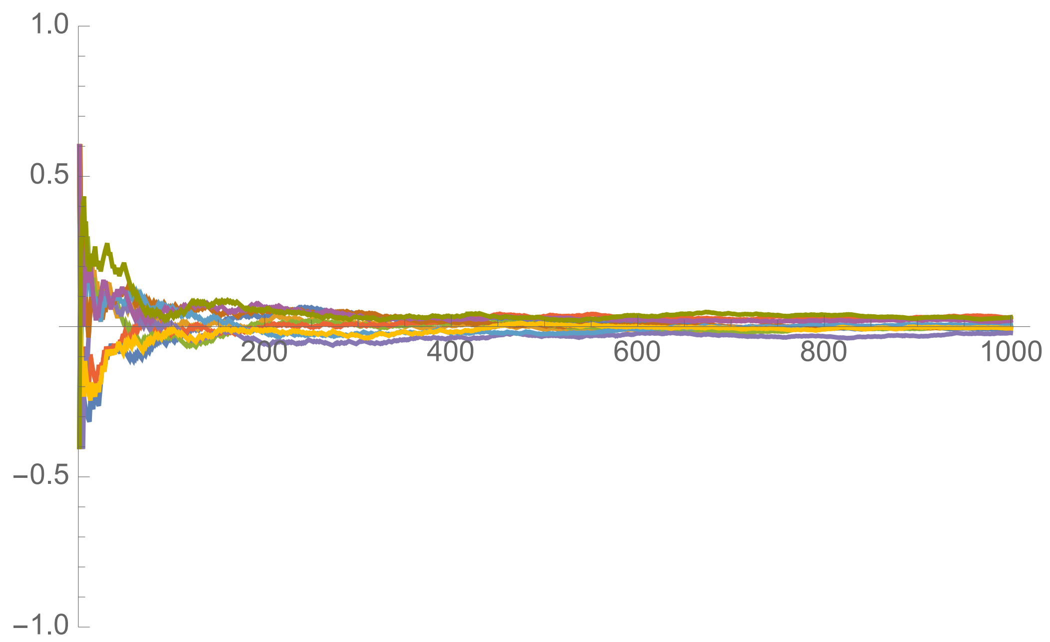


Figure: 10 simulations of $\left(\frac{S_n}{n} - p : 1 \leqslant n \leqslant 1000\right)$ for $p = 0.6$.

- Can we "zoom in"?
- Is there a function $f(n)$ such that $f(n)\left(\frac{S_n}{n} - p\right)$ has a nice behavior for n large?

The Central Limit Theorem

The speed of convergence is $\frac{1}{\sqrt{n}}$. This means that $\sqrt{n}(\frac{S_n}{n} - p)$ has a nondegenerate behavior as $n \to \infty$.

The Central Limit Theorem

The speed of convergence is $\frac{1}{\sqrt{n}}$. This means that $\sqrt{n}(\frac{S_n}{n} - p)$ has a nondegenerate behavior as $n \to \infty$.

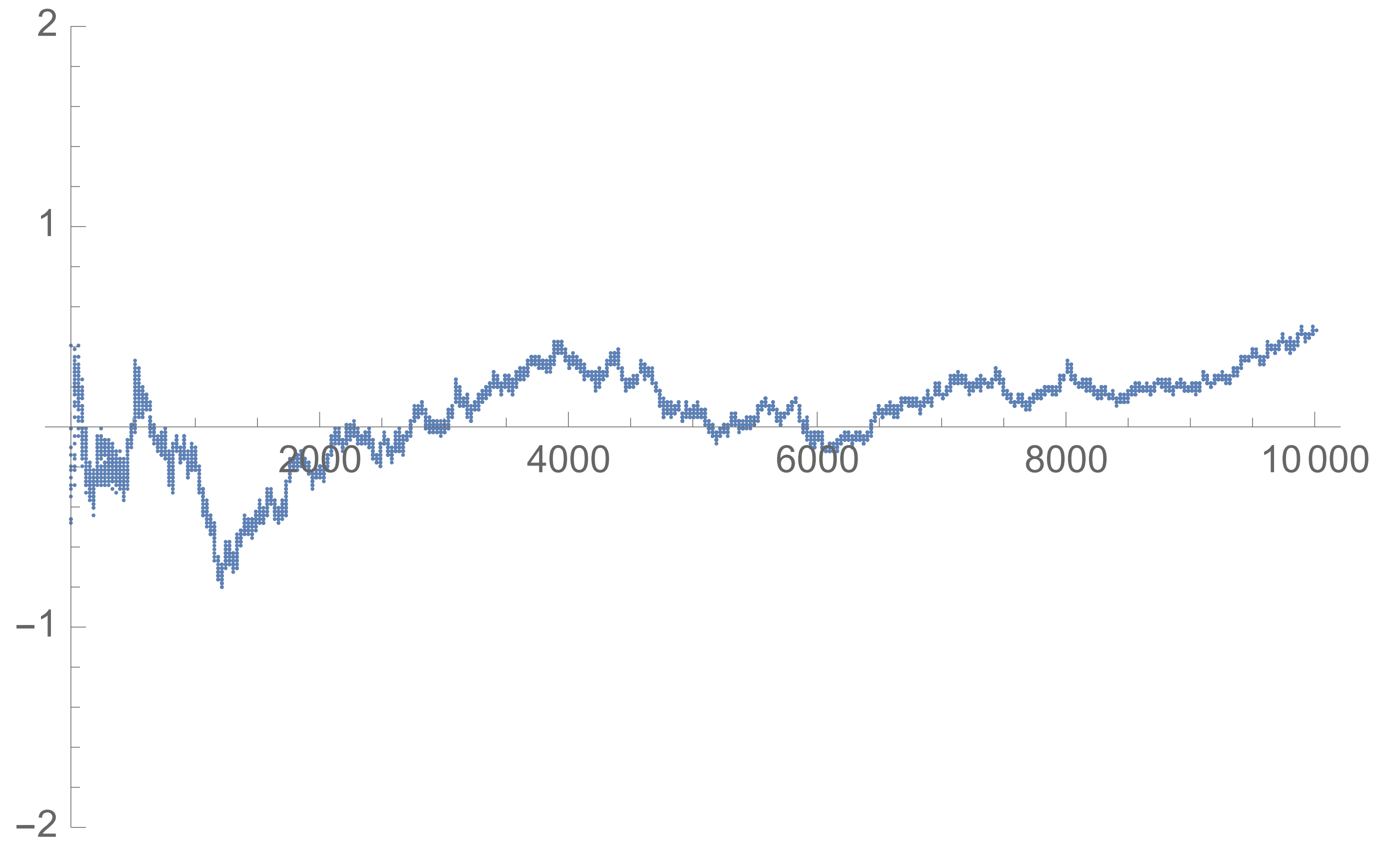


Figure: Simulation of $\left(\sqrt{n} \cdot \left(\frac{S_n}{n} - p\right) : 1 \leqslant n \leqslant 10000\right)$ for $p = 0.6$.

The Central Limit Theorem

The speed of convergence is $\frac{1}{\sqrt{n}}$. This means that $\sqrt{n}(\frac{S_n}{n} - p)$ has a nondegenerate behavior as $n \to \infty$.

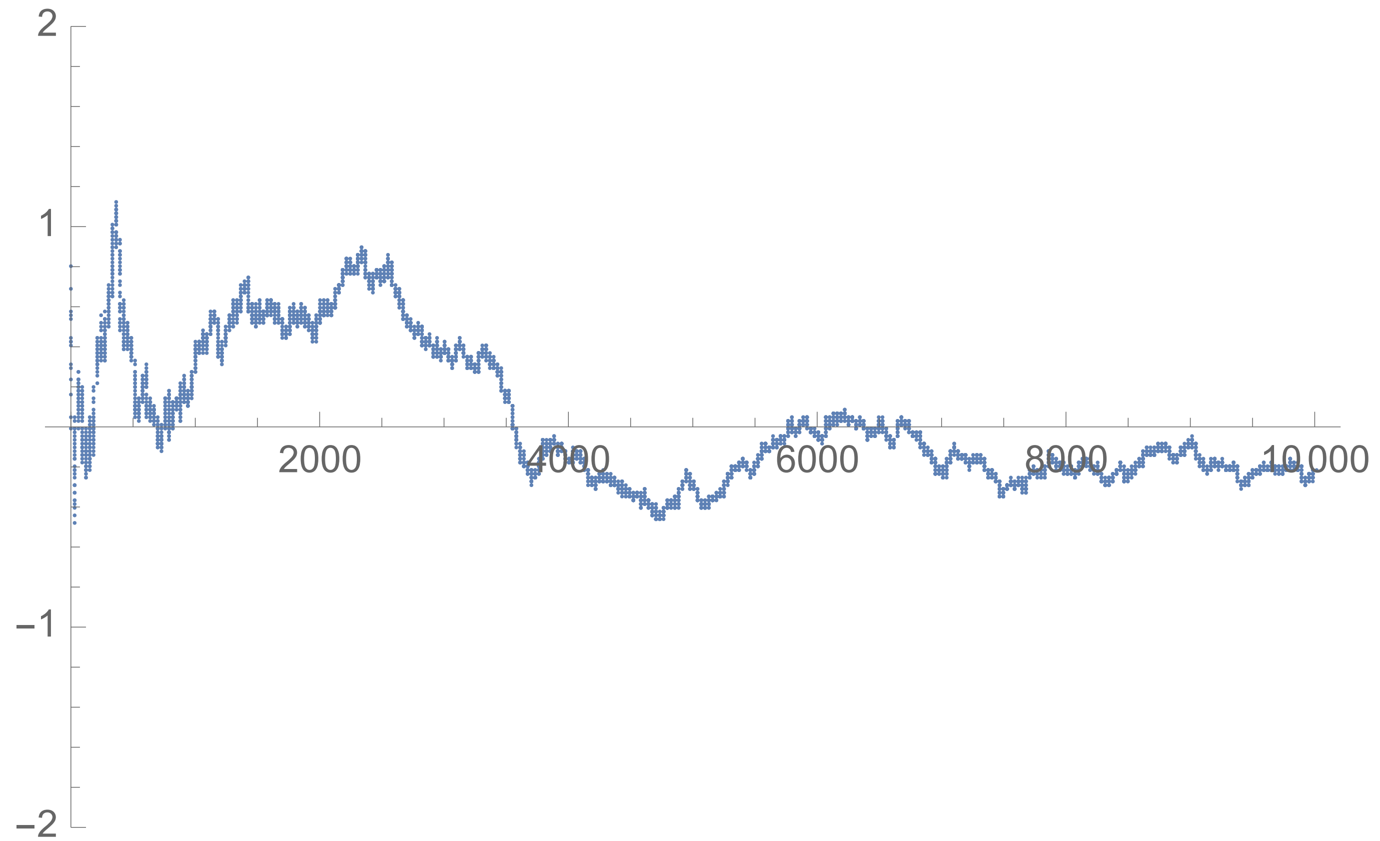


Figure: Another simulation of $\left(\sqrt{n} \cdot \left(\frac{S_n}{n} - p\right) : 1 \leqslant n \leqslant 10000\right)$ for $p = 0.6$.

The Central Limit Theorem

The speed of convergence is $\frac{1}{\sqrt{n}}$. This means that $\sqrt{n}(\frac{S_n}{n} - p)$ has a nondegenerate behavior as $n \to \infty$.

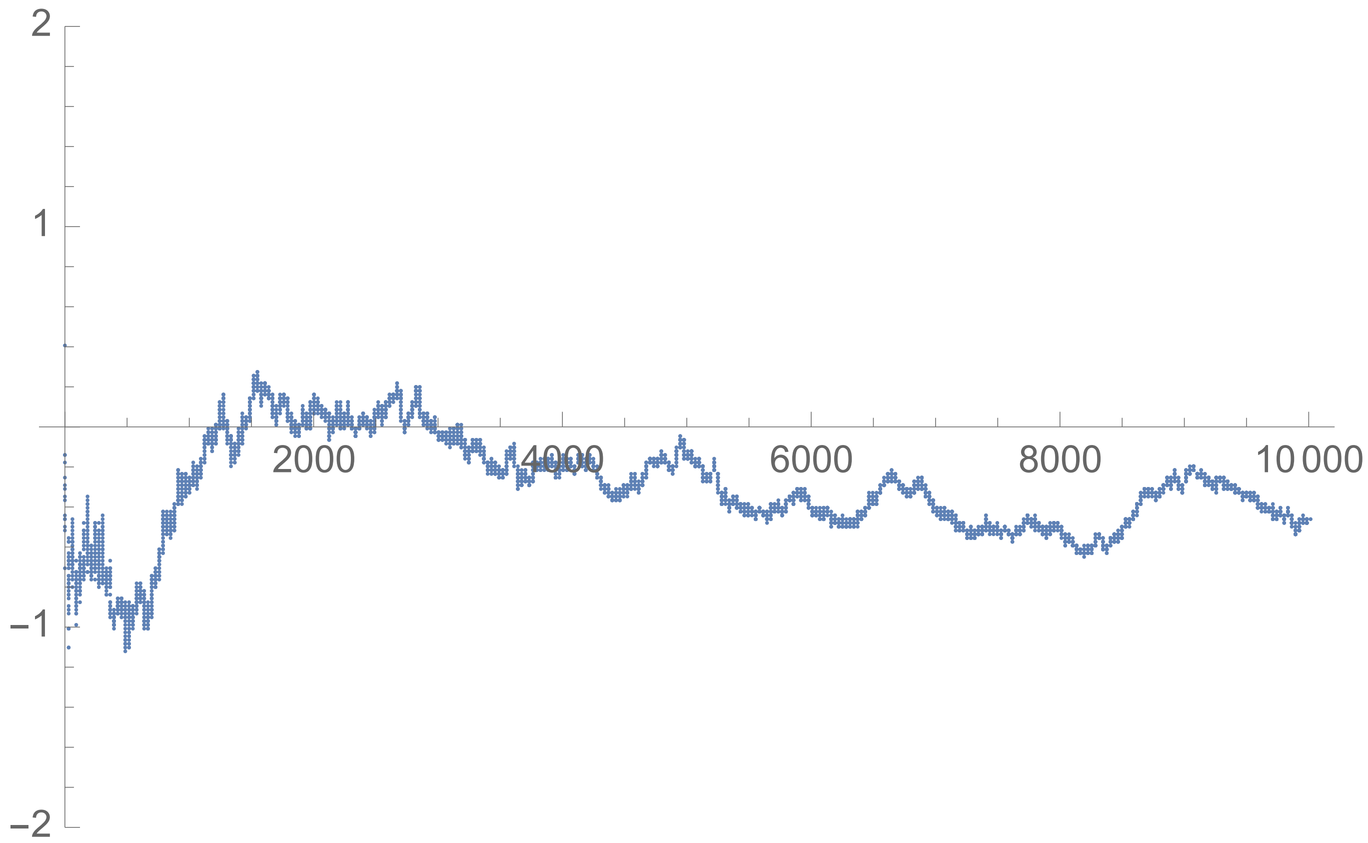


Figure: Another simulation of $\left(\sqrt{n} \cdot \left(\frac{S_n}{n} - p\right) : 1 \leqslant n \leqslant 10000\right)$ for $p = 0.6$.

The Central Limit Theorem

The speed of convergence is $\frac{1}{\sqrt{n}}$. This means that $\sqrt{n}(\frac{S_n}{n} - p)$ has a nondegenerate behavior as $n \to \infty$.

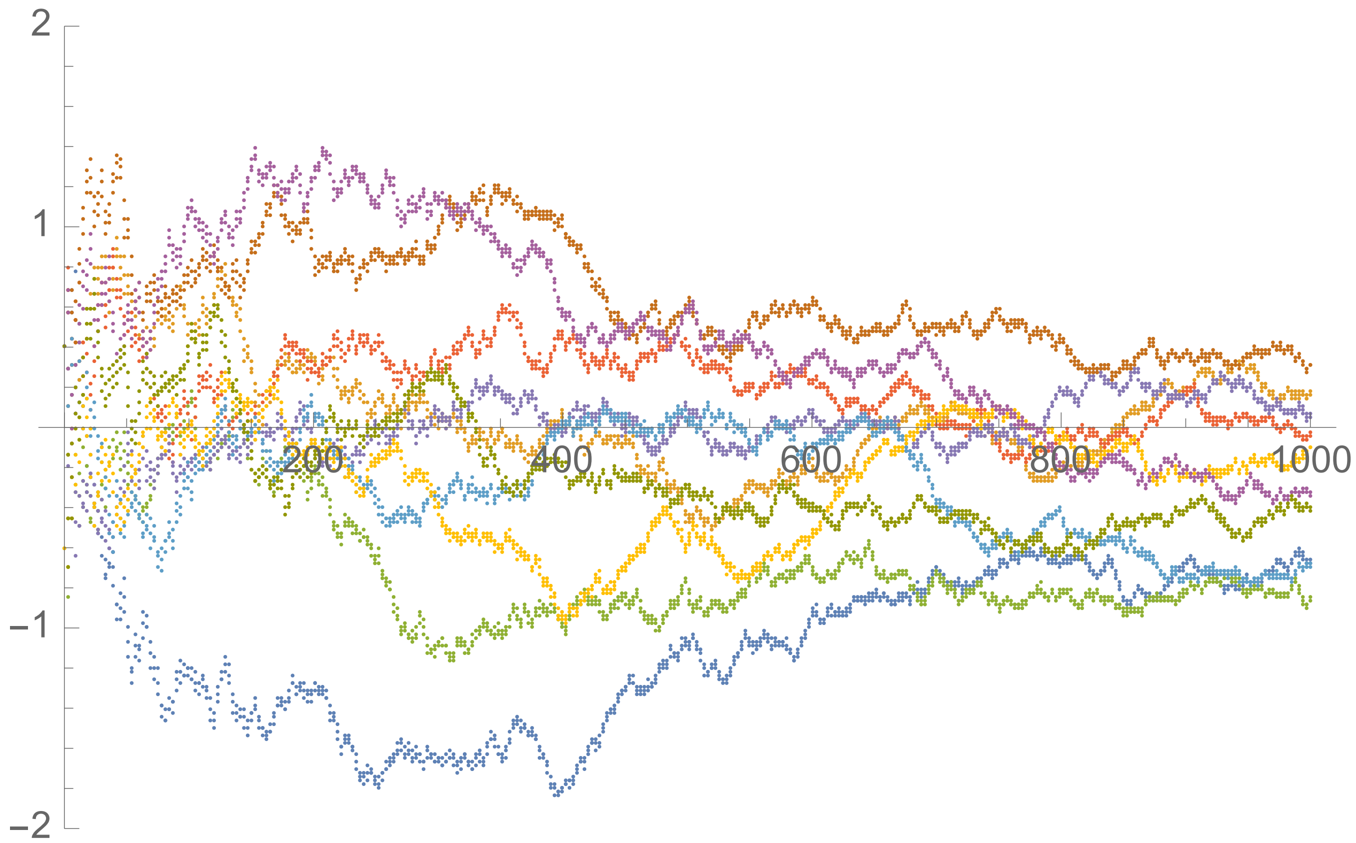


Figure: 10 simulations of $\left(\sqrt{n} \cdot \left(\frac{S_n}{n} - p\right) : 1 \leqslant n \leqslant 1000\right)$ for $p = 0.6$.

The Central Limit Theorem

The speed of convergence is $\frac{1}{\sqrt{n}}$. This means that $\sqrt{n}(\frac{S_n}{n} - p)$ has a nondegenerate behavior as $n \to \infty$.

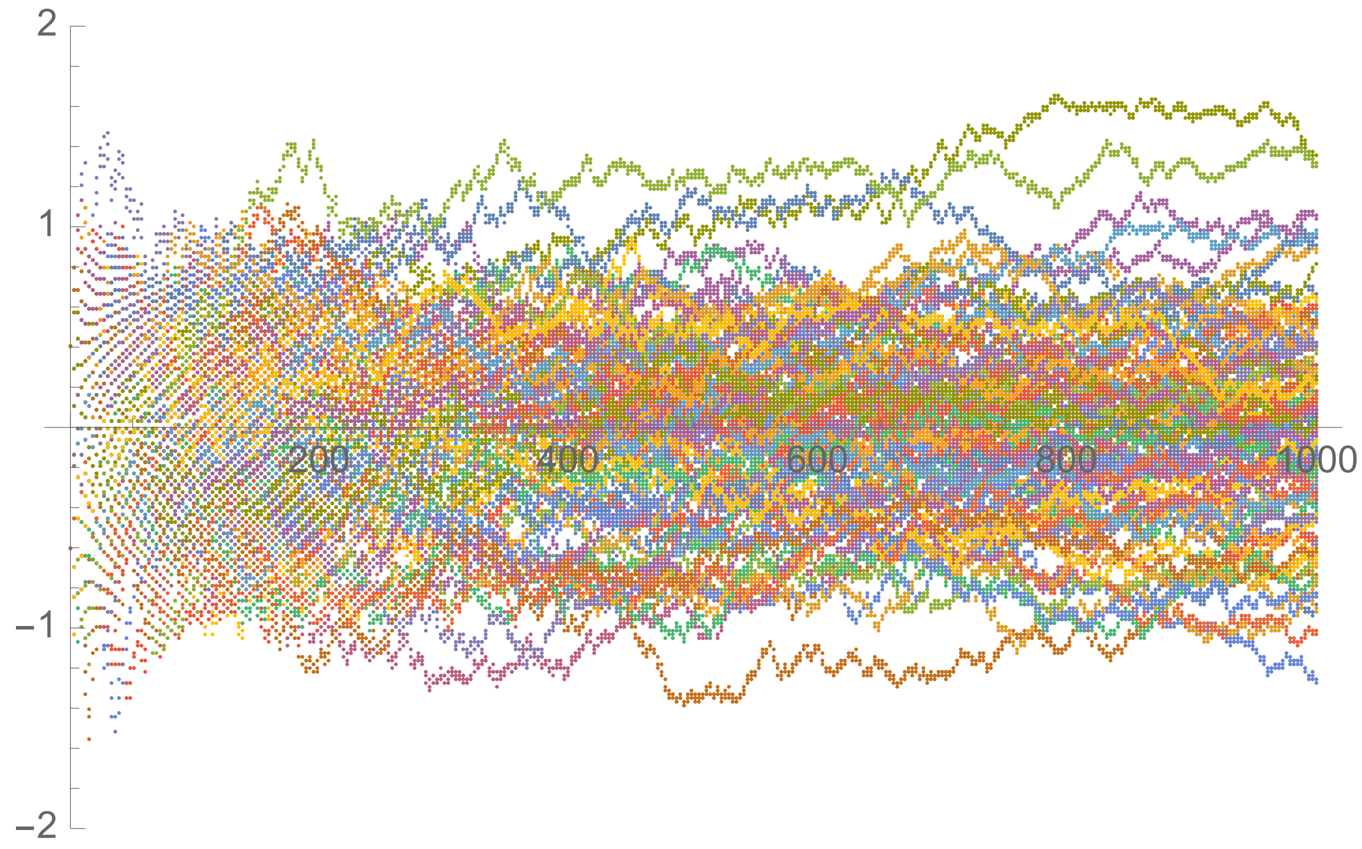


Figure: 100 simulations of $\left(\sqrt{n} \cdot \left(\frac{S_n}{n} - p\right) : 1 \leqslant n \leqslant 1000\right)$ for $p = 0.6$.

Structure in randomness

There is structure in this randomness!

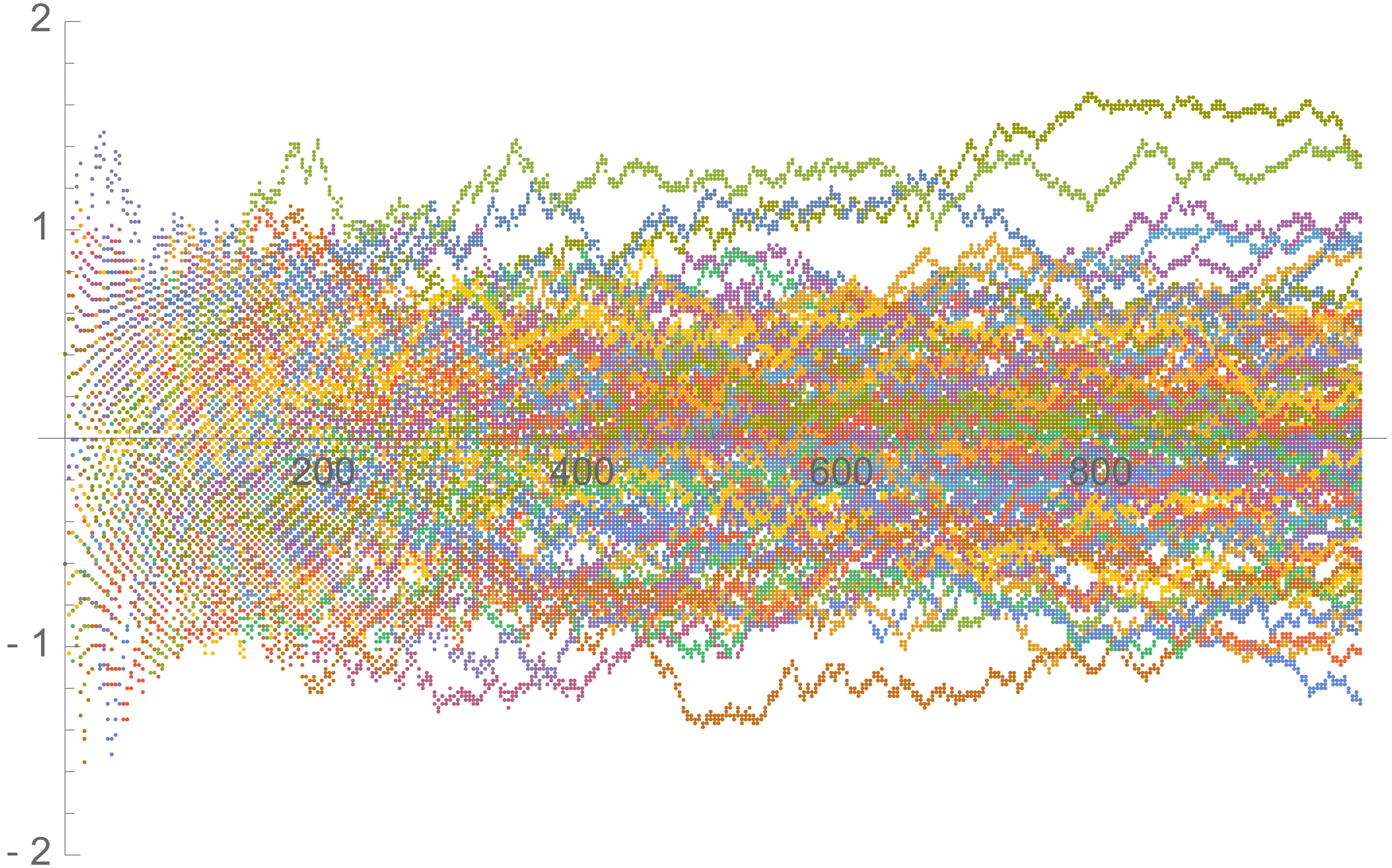


Figure: 100 simulations of $\left(\sqrt{n} \cdot \left(\frac{S_n}{n} - p\right) : 1 \leqslant n \leqslant 1000\right)$ for $p = 0.6$.

Structure in randomness

There is structure in this randomness! Look at the "endpoints" $\sqrt{n}(\frac{S_n}{n} - p)$

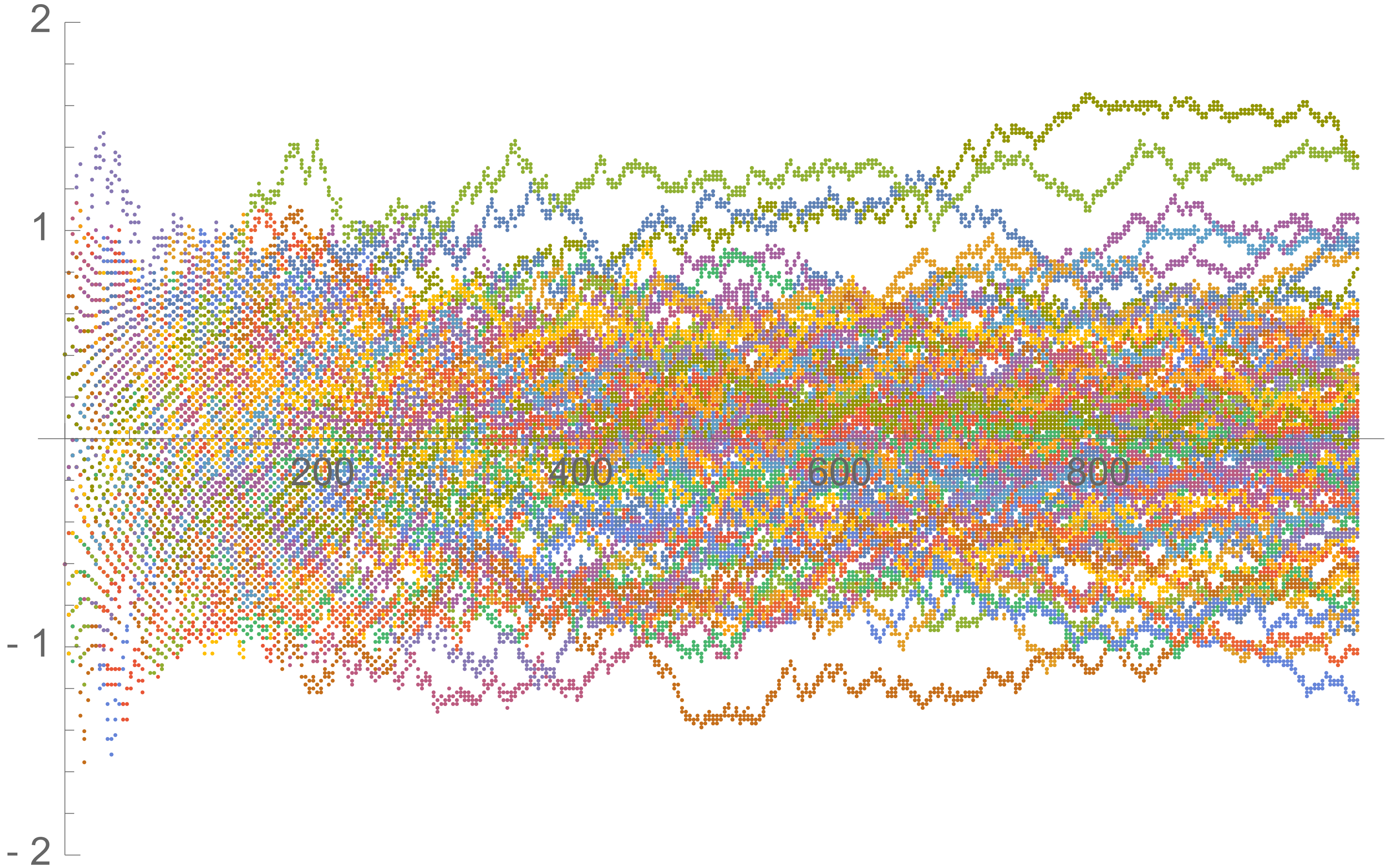


Figure: 100 simulations of $\left(\sqrt{n} \cdot \left(\frac{S_n}{n} - p\right) : 1 \leqslant n \leqslant 1000\right)$ for $p = 0.6$.

Structure in randomness

There is structure in this randomness! Look at the "endpoints" $\sqrt{n}(\frac{S_n}{n} - p)$ and draw the empirical histogram.

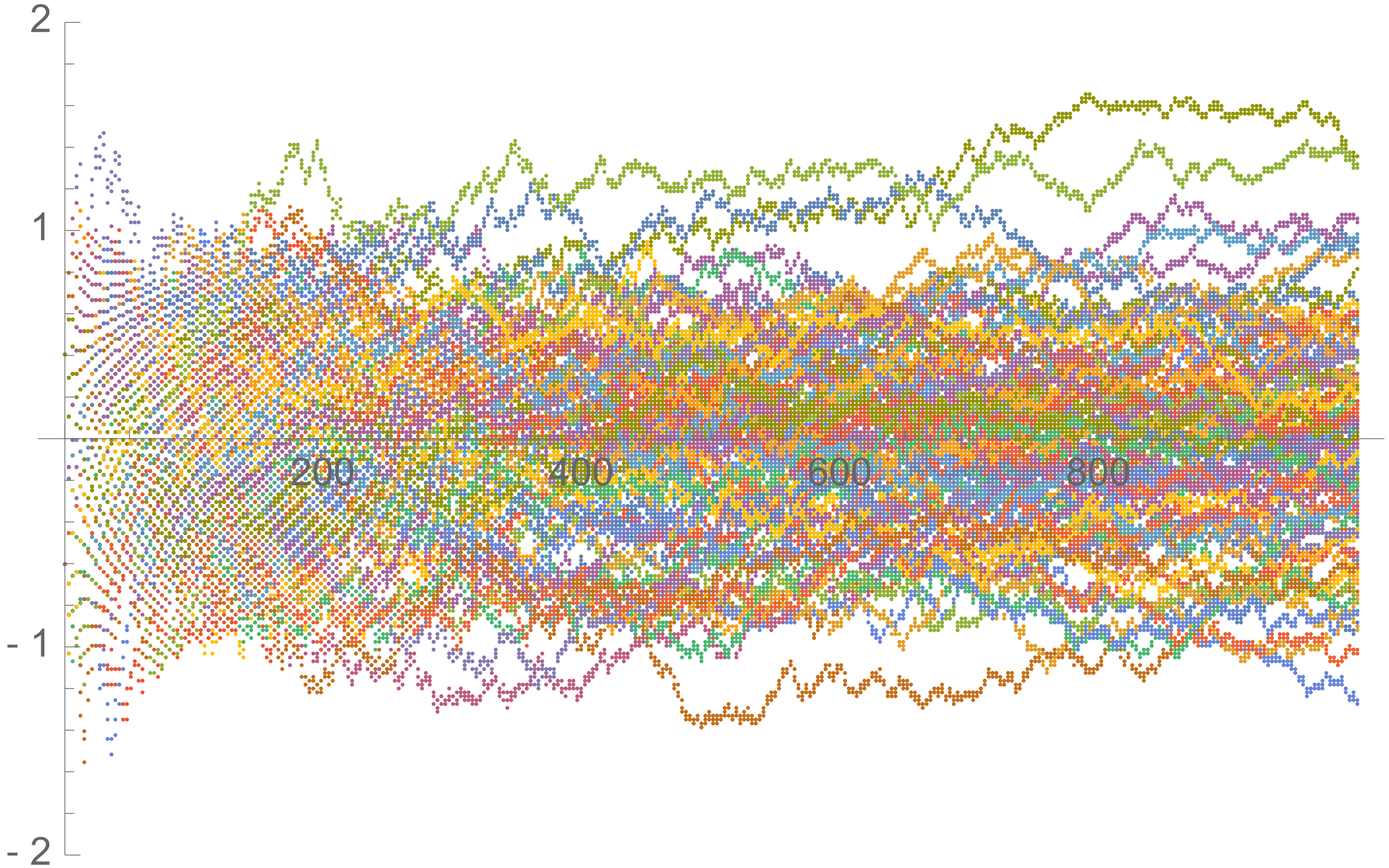


Figure: 100 simulations of $\left(\sqrt{n} \cdot \left(\frac{S_n}{n} - p\right) : 1 \leqslant n \leqslant 1000\right)$ for $p = 0.6$.

Structure in randomness

There is structure in this randomness! Look at the "endpoints" $\sqrt{n}(\frac{S_n}{n} - p)$ and draw the empirical histogram.

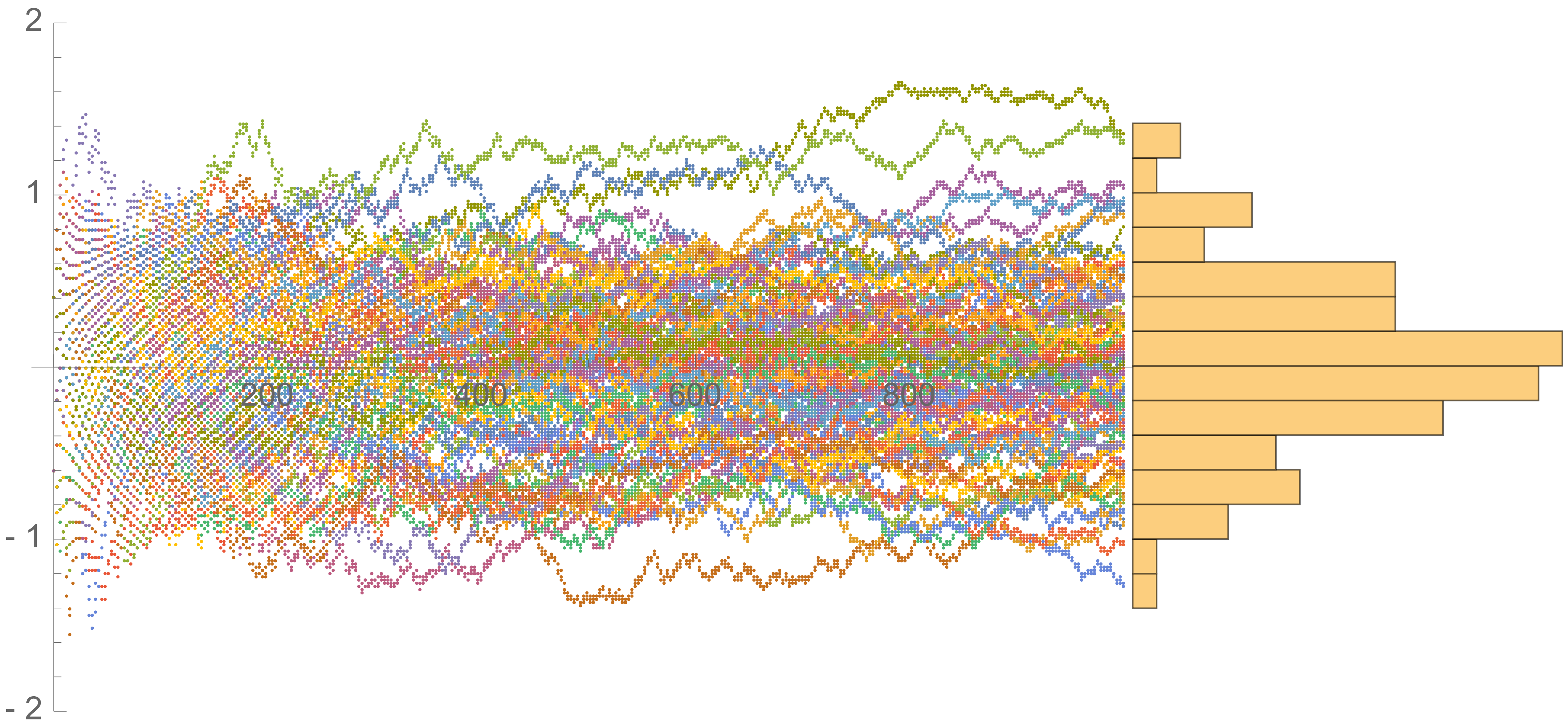


Figure: 100 simulations of $\left(\sqrt{n} \cdot \left(\frac{S_n}{n} - p\right) : 1 \leqslant n \leqslant 1000\right)$ for $p = 0.6$.

Structure in randomness

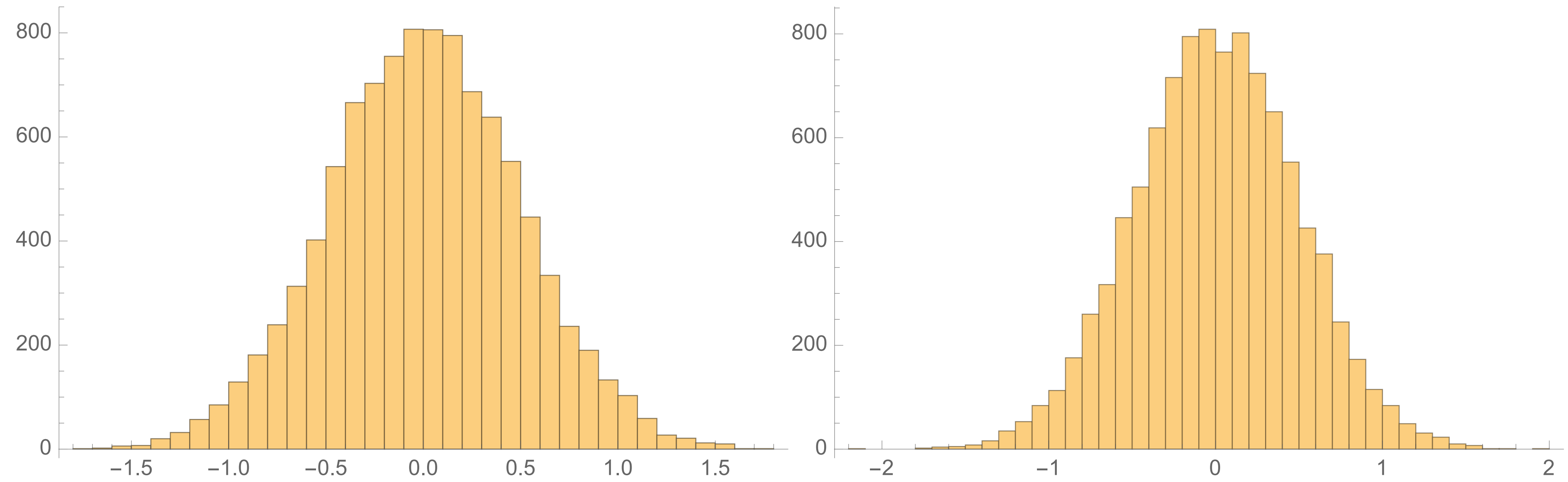


Figure: Empirical histograms of 10000 simulations of $\sqrt{n} \cdot \left(\frac{S_n}{n} - p\right)$ for $n = 10000$.
Left: $p = 0.6$; Right: $p = 0.4$.

Structure in randomness

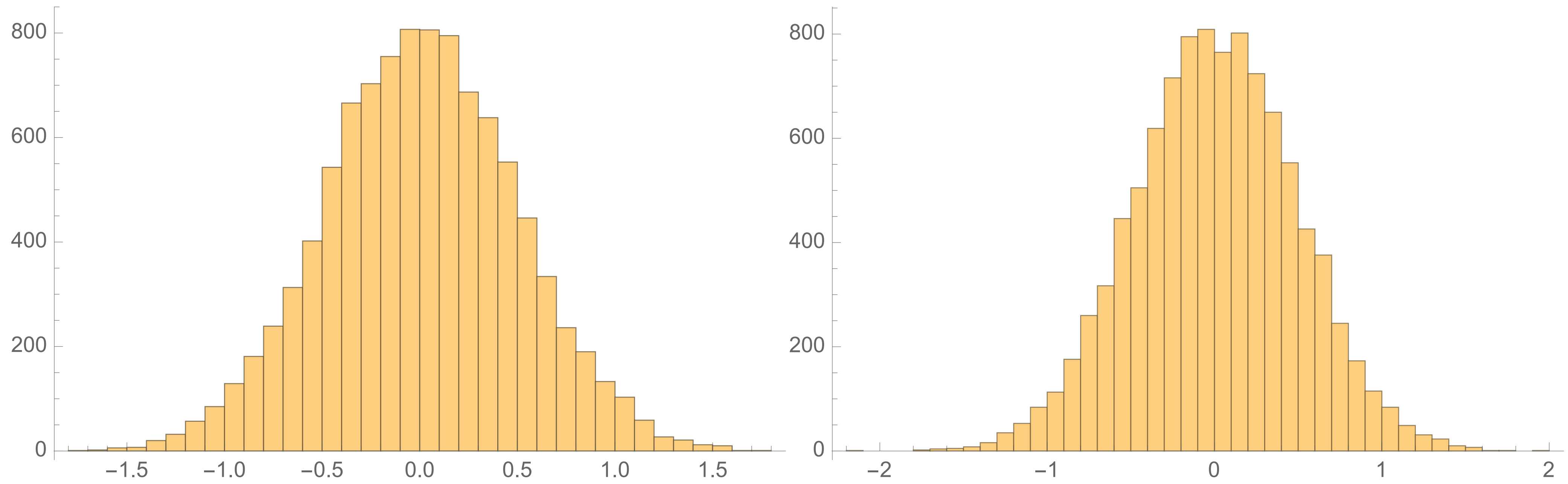


Figure: Empirical histograms of 10000 simulations of $\sqrt{n} \cdot \left(\frac{S_n}{n} - p\right)$ for $n = 10000$. Left: $p = 0.6$; Right: $p = 0.4$.

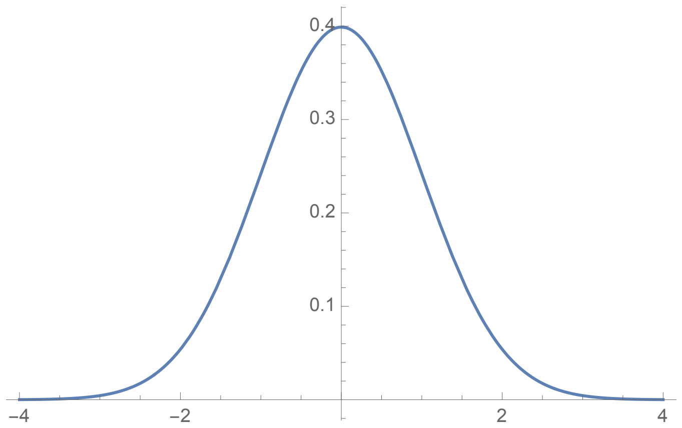


Figure: Plot of the function $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

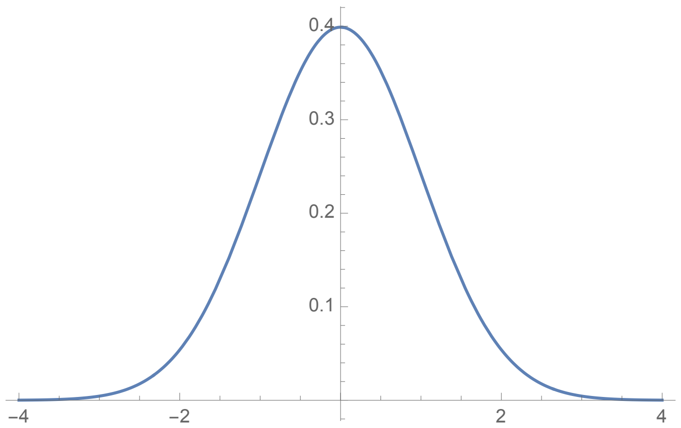


Figure: Plot of the function $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

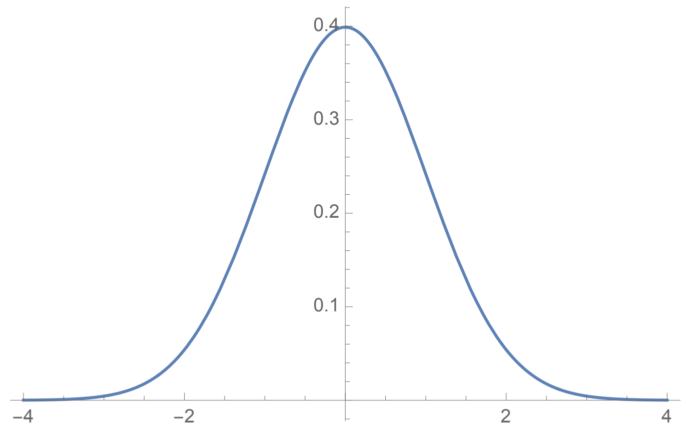


Figure: Plot of the function $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Theorem (Central limit theorem – De Moivre Laplace theorem)**.**

Let S_n be the sum of n independent Bernoulli random variables of parameter $p \in (0,1)$. Then, for every $a < b$:

$$\mathbb{P}\left(a \leqslant \frac{\sqrt{n}}{\sqrt{p(1-p)}} \left(\frac{S_n}{n} - p \right) \leqslant b \right) \underset{n \to \infty}{\longrightarrow} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

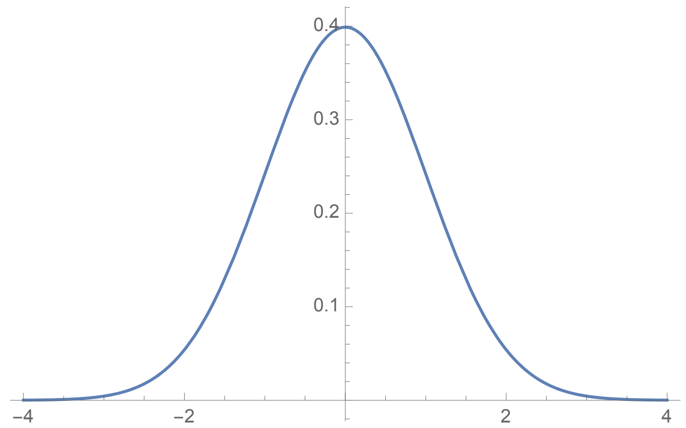


Figure: Plot of the function $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

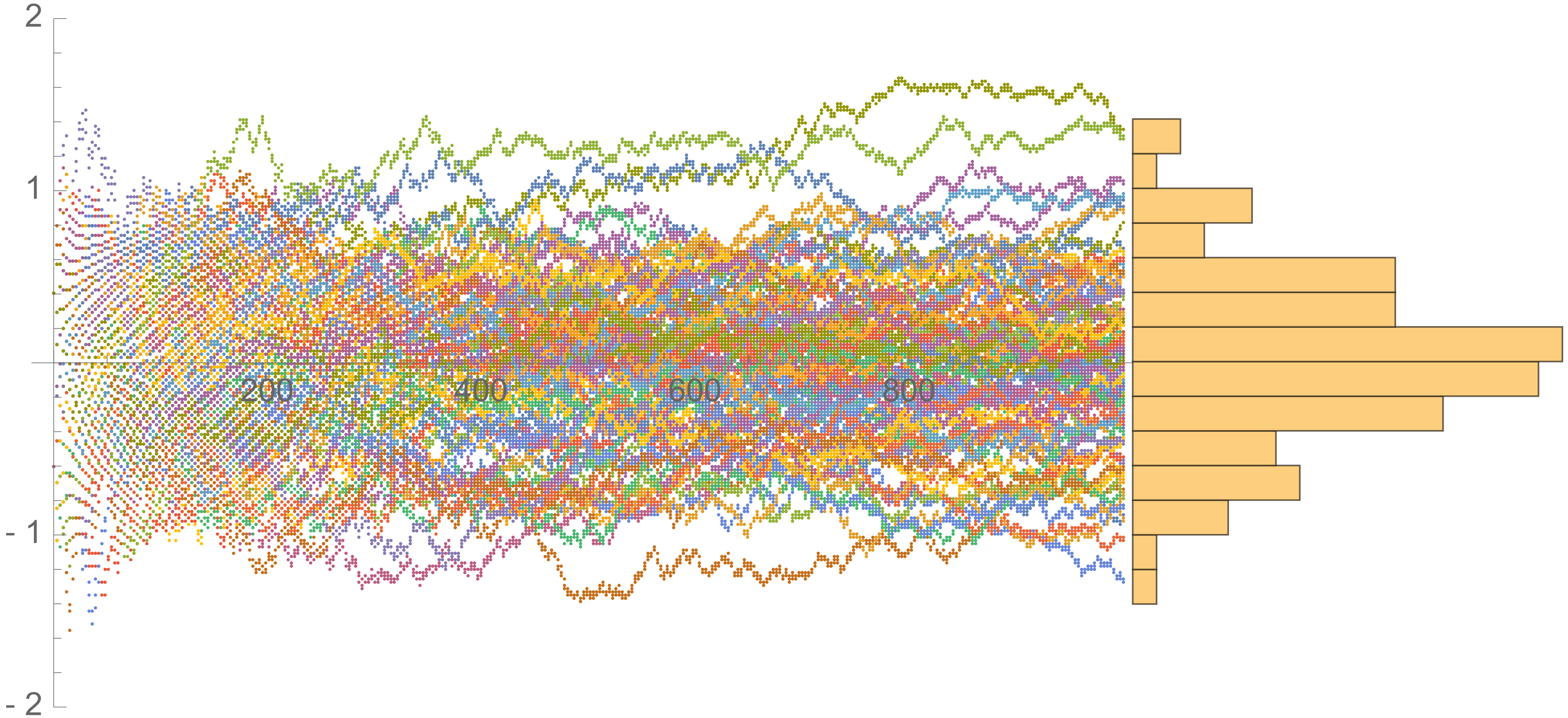
Theorem (Central limit theorem – De Moivre Laplace theorem)**.**

Let S_n be the sum of n independent Bernoulli random variables of parameter $p \in (0,1)$. Then, for every $a < b$:

$$\mathbb{P}\left(a \leqslant \frac{\sqrt{n}}{\sqrt{p(1-p)}}\left(\frac{S_n}{n} - p\right) \leqslant b\right) \quad \underset{n\to\infty}{\longrightarrow} \quad \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

We say that $\frac{\sqrt{n}}{\sqrt{p(1-p)}}\left(\frac{S_n}{n} - p\right)$ converges *in distribution* to a Gaussian random variable.

The central limit theorem



We do not know where the “trajectory" will arrive, but we know an esimate of the probability that it arrives in a certain region thanks to the central limit theorem.

Discrete mathematics

MAA 103

Recap

First part: foundations

- Sets
- Mathematical assertions (quantifiers)
- Functions

Recap
Second part: combinatorics

- Binomial coefficients
- Permutations
- Graphs

Recap

Third part: Probability

- Events, probabilities
- Independence, conditional probabilities

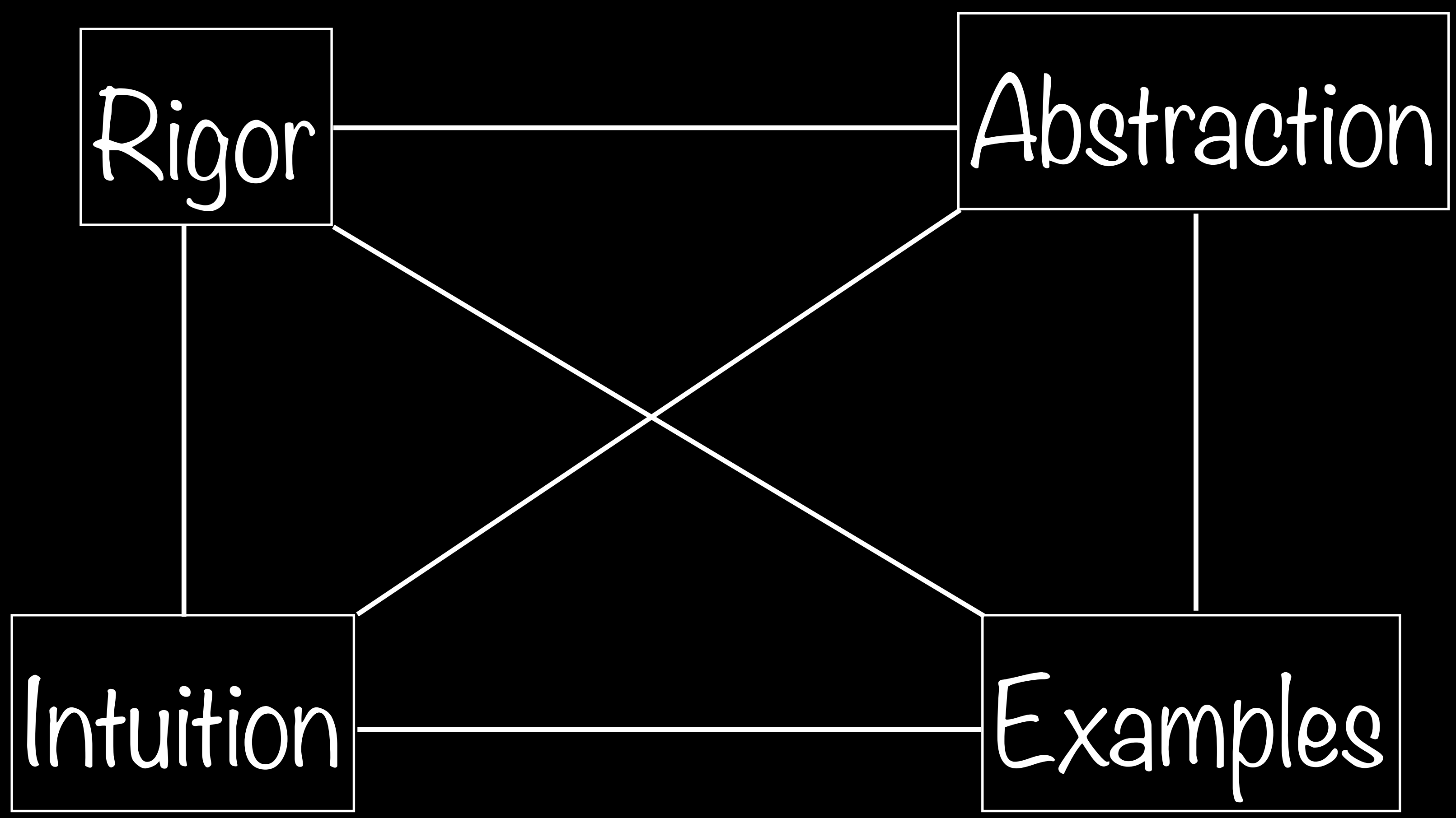
Rigor

Abstraction

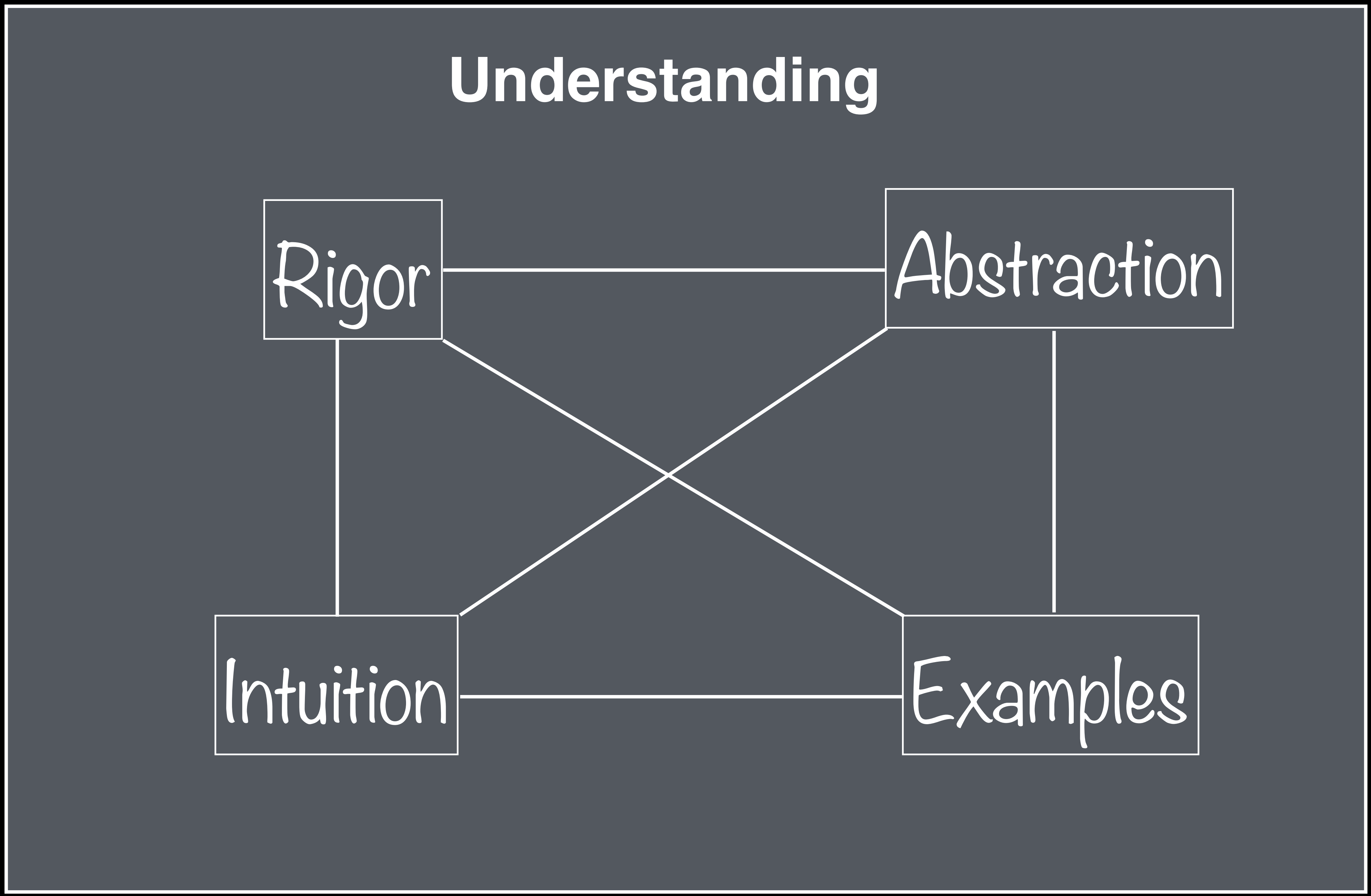
Intuition

Examples

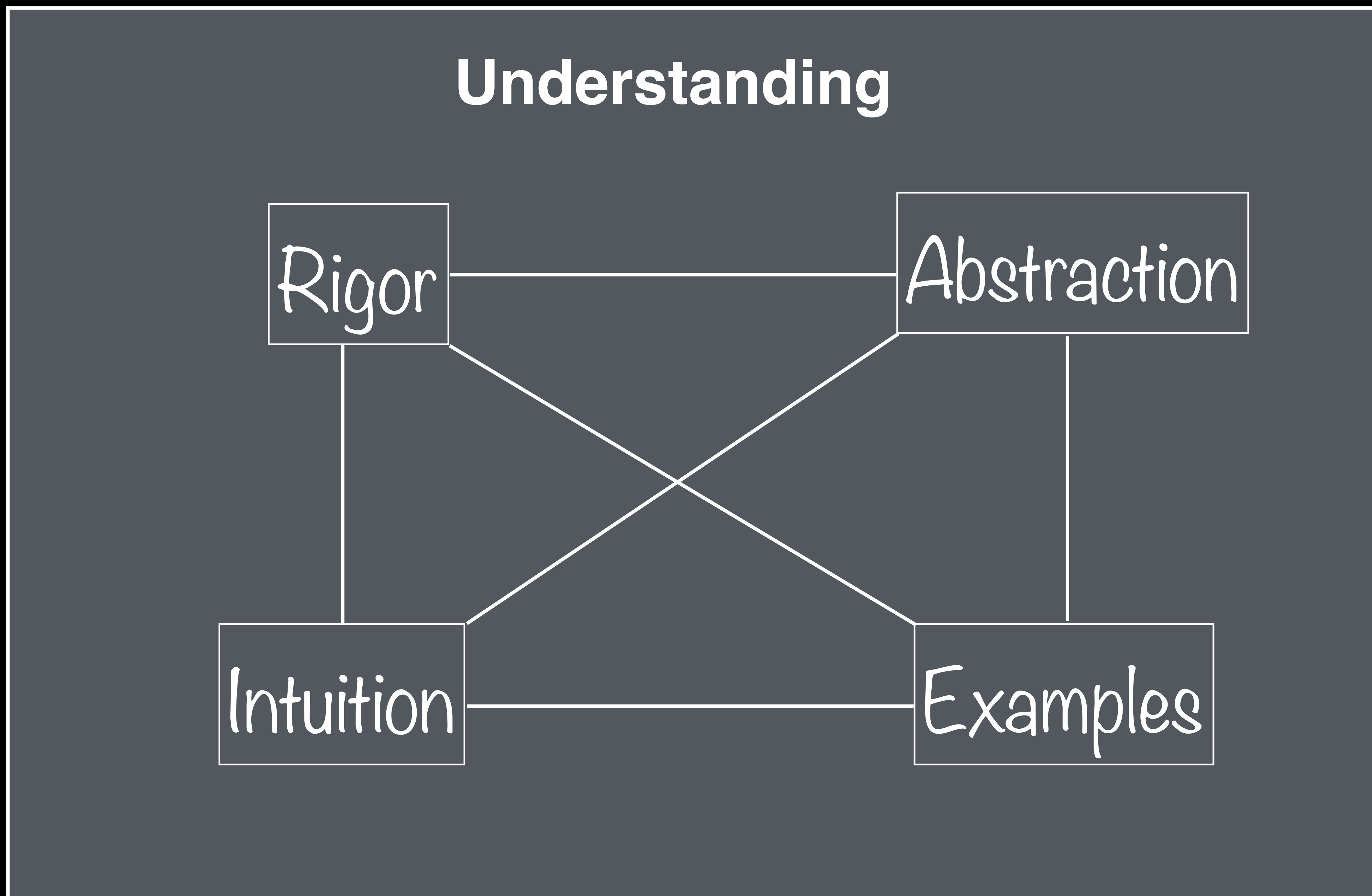
A graph



A graph



A graph



Warning when applying Mathematics in the real world!