# SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

## Convergence of Random Series

av

**Matilda Colarieti Tosti**

2025 - No K2

# Convergence of Random Series

Matilda Colarieti Tosti

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Alan Sola

2025

**Abstract**

This thesis studies the convergence of random series, where the terms are given by random variables rather than fixed numbers. The main focus is the development and application of martingale methods, particularly in $\mathcal{L}^2$, to reduce convergence questions to tractable conditions such as boundedness of the underlying random variances. After reviewing foundational concepts from real analysis and probability, we present key martingale properties and demonstrate how they yield general criteria for almost sure convergence of random series.

**Sammanfattning**

Denna avhandling behandlar konvergens hos slumpserier, där termerna ges av stokastiska variabler. Huvudfokus ligger på utveckling och tillämpning av martingalmetoder, särskilt i $\mathcal{L}^2$, för att reducera konvergens frågor till hanterbara villkor såsom begränsning av de underliggande slumpvariablernas varianser. Efter en genomgång av grundläggande begrepp från reell analys och sannolikhetsteori presenterar vi centrala egenskaper hos martingaler och visar hur dessa leder till generella kriterier för nästan säker konvergens av slumpmserier.

## Use of AI

I have used deepL to translate words and sentences from Swedish to English when I was unsure about a word, phrasing or grammar.
I have used ChatGPT for help with LaTeX, such as formatting and symbols, as well as for checking grammar when felt needed.

**Acknowledgements**

I want to thank my supervisor, Alan Sola, for his patience, guidance and support throughout this project. His advice and encouragement has been of great help to me.

# Contents

# 1 Introduction

Infinite series are a central object of study in analysis, and many classical convergence questions concern deterministic sequences of real numbers. In probability theory, however, one is often led to study *random series*, where the terms themselves are random variables. In this setting, the convergence is no longer a purely analytic question but depends on probabilistic structure as well. The main focus of this thesis is to understand when such random series converge.

As a guiding example, consider the series

$$\sum_{k=1}^{\infty} \frac{X_k}{k},$$

where $(X_k)_{k\geq 1}$ is a sequence of independent random variables taking the values $+1$ and $-1$ with equal probability.

At first glance, it is not obvious whether this series should converge or diverge. Before proceeding the reader is invited to pause and consider this question.

My own initial thoughts were guided by several competing intuitions. Since the terms $\frac{1}{k}$ decrease only slowly, any hope of convergence must come from cancellation between positive and negative contributions. Without such cancellation, the partial sums would necessarily grow without bound. Thus, the behavior of the series is governed not by the size of individual terms alone, but by how their signs interact over time.

On the one hand, if the signs were arranged to alternate perfectly, positive and negative terms would largely cancel each other out, suggesting convergence. This situation is of course highly idealised, but provides a useful point of reference. Even small deviations from this perfect alternation might still, at least intuitively, allow the positive and the negative terms to balance each other in the long run.

On the other hand, randomness allows for much more irregular behavior. Consecutive terms may share the same sign for extended periods, forming clusters of varying lengths and such clusters can occur repeatedly throughout the series. If sufficiently long runs of positive or negative terms occur infinitely often, the partial sums may be consistently nudged in one direction over time, producing a persistent "drift" that prevents the series from settling. From this perspective, it is conceivable that the series could fail to converge, as these random patterns of aligned signs keep pushing the sum outward. This suggests divergence cannot be ruled out.

Another possible intuition is that the series might converge for some realizations of $(X_k)$ and diverge for others. Depending on how the random signs arrange themselves, the partial sums might eventually settle, or they might be repeatedly be pushed outward by clusters of aligned signs. In this way, one could think of convergence or divergence as probabilistic possibilities and in this case we imagine that both outcomes—convergence and divergence—occur with positive probability.

We will return to this example throughout the thesis and ultimately provide a rigorous answer using the tools provided here.

More generally, the aim of this thesis is to develop general methods for studying convergence of random series. A central role is played by *martingales*, which arise naturally when considering partial sums of random variables. Martingale methods allow convergence questions to be reduced to more tractable conditions, such as boundedness in $\mathcal{L}^2$.

The reader is assumed to have a basic background in probability theory, including probability spaces, random variables, expectation, and independence. Familiarity with elementary real analysis, such as sequences and series of real numbers, is also assumed. For completeness Chapter 2 reviews essential concepts from analysis and probability, including modes of convergence, tail events and the Borel-Cantelli lemmas.

In Chapter 3, we introduce martingales and develop their fundamental properties, with particular emphasis on $\mathcal{L}^2$ methods, orthogonality of the increments and stopping times.

Finally, in Chapter 4, we apply martingale theory and related convergence theorems to establish conditions for the almost sure convergence of random series. We conclude by resolving the motivating example $\sum X_k/k$.

The material and perspective of this thesis are primarily based on the textbooks *Probability with Martingales* by David Williams [Wil91] as well as *Stokastik* by Sven Erick Alm and Tom Britton [AB08], which together form the theoretical foundation for the results presented here.

# 2 Preliminaries

This chapter is intended as a reference for the reader and is divided into two parts: the first recalls classical results on infinite series, and the second introduces the probabilistic tools required to study random variables.

Readers already familiar with these preliminaries may safely proceed directly to Chapter 3 without loss of continuity and return here as needed.

## 2.1 Series

**Definition 2.1.** Let $(a_k)_{k\geq 0}$ be a sequence of real numbers. The corresponding *series* is the expression

$$\sum_{k=0}^{\infty} a_k.$$

Its $n$th *partial sum* is defined by

$$s_n = \sum_{k=0}^{n} a_k.$$

If the limit $s = \lim_{n\to\infty} s_n$ exists and is finite, we say that the series *converges* to $s$. Otherwise, we say that the series *diverges*.

Unless otherwise stated, we write

$$\sum a_k \equiv \sum_{k=1}^{\infty} a_k.$$

**Definition 2.2.** Let $\sum_{k=0}^{\infty} a_k$ be an infinite series. For any integer $m > 0$, the *m-th tail* of the series is the series

$$\sum_{k=m}^{\infty} a_k.$$

The $m$-th tail is simply the original series with its first $m-1$ terms removed.

*Remark* 2.3. A series converges precisely when all of its tails converge; if any tail diverges, then the whole series diverges.

*Example* 2.4. The series $\displaystyle\sum_{k=0}^{\infty} 1$ is divergent.

To see this we examine the sequence of partial sums:

$$s_n = \sum_{k=0}^{n} 1.$$

Computing the first few partial sums, we obtain

$$\begin{aligned} s_0 =& 1, \\ s_1 =& 1 + 1 = 2, \\ s_2 =& 1 + 1 + 1 = 3, \\ &\vdots \\ s_n =& 1 + n. \end{aligned}$$

As $n \to \infty$, we have $s_n \to \infty$. Thus, the series diverges.

This can also be seen by considering the $m$-th tail of the series. For any integer $m > 0$, the $m$-th tail is

$$\sum_{k=m}^{n} 1 = n - m + 1,$$

which also tend to infinity as $n \to \infty$. Hence, every tail of the series diverges, in agreement with the preceding remark.

Consequently series are not sums in the traditional sense but limits of finite sums.

In many cases, the convergence or divergence of a series is not immediately obvious. To study the behavior of series, it is helpful to have systematic methods and criteria that allow us to determine convergence rigorously. In the following, we consider several important classes of series — geometric series, $p$-series, and alternating series — which provide useful reference points for more general convergence tests.

**Definition 2.5.** A *geometric series* is a series of the form

$$\sum_{k=0}^{\infty} ar^k = a + ar + ar^2 + ar^3 + \dots,$$

where $a, r \in \mathbb{R}$ and $r$ is called the *common ratio.*

**Theorem 2.6.** *Let $a, r \in \mathbb{R}$. The geometric series $\sum_{k=0}^{\infty} ar^k$ converges if and only if $|r| < 1$, in which case*

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1 - r}.$$

*If $|r| \geq 1$, the series diverges.*

*Proof.* We consider different cases of the value of $r$.
If $r = 1$, then $S_n = a + a + ... + a = an$.

For $a \neq 0$,

$$\lim_{n\to\infty} S_n \to \begin{cases} \infty & \text{for } a > 0, \\ -\infty & \text{for } a < 0, \end{cases}$$

so the series is divergent.

If $r = -1$, then $S_n = a - a + a - ... \pm a$ and the partial sums will alternate between 0 and $a$, never settling to a limit. Hence, the series diverges.

Now assume $r \neq \pm 1$. The $n$-th partial sum is

$$S_n = a + ar + ar^2 + ... + ar^{n-1}.$$

Multiplying by $r$ and subtracting gives

$$S_n - rS_n = (a + ar + ar^2 + \cdots + ar^{n-a}) - (ar + ar^2 + ... + ar^n) = a - ar^n.$$

Therefore, $S_n(1-r) = a(a - r^n)$ and the $n$-th partial sum is

$$S_n = \frac{a}{1-r}(1 - r^n).$$

If $|r| > 1$, then $|r^n| \to \infty$ so the series diverges.
If $0 < |r| < 1$, then $r^n \to 0$ as $n \to \infty$, which yields

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{a}{1-r}(1 - r^n) = \frac{a}{1-r} \lim_{n\to\infty}(1 - r^n) = \frac{a}{1-r}.$$

So, the series converges and its sum is $\frac{a}{1-r}$. □

*Example* 2.7. The series

$$\sum_{k=0}^{\infty} \frac{1}{2^k}$$

is a geometric series with common ratio $r = \frac{1}{2}$. By Theorem 2.6 the series converges, and its sum is given by

$$\sum_{k=0}^{\infty} \frac{1}{2^k} = \frac{1}{1 - \frac{1}{2}} = 2.$$

Another basic class of series is the *p-series*, defined by

$$\sum \frac{1}{k^p}, \qquad p > 0.$$

**Theorem 2.8.** *The series* $\sum_{k=1}^{\infty} \frac{1}{k^p}$ *converges if and only if* $p > 1$*. If* $0 < p \leq 1$*, the series diverges.*

*Example* 2.9. The series $\sum_{k=1}^{\infty} \frac{1}{k}$, called the *harmonic series*, is the special case of a $p$-series with $p = 1$. By Theorem 2.8 it follows immediately that the harmonic series diverges, since $p = 1 \not> 1$.

*Proof.* We show that the harmonic series diverges by grouping its terms.

$$\sum_{k=1}^{\infty} \frac{1}{k} = 1 + \frac{1}{2} + \underbrace{\frac{1}{3} + \frac{1}{4}}_{>\frac{1}{2}} + \underbrace{\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}}_{>\frac{1}{2}} + \underbrace{\frac{1}{9} + \cdots + \frac{1}{16}}_{>\frac{1}{2}} + \cdots$$

Each group contains twice as many terms as the previous one and has sum greater than $\frac{1}{2}$. Therefore,

$$\sum_{k=1}^{\infty} \frac{1}{k} \geq 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \ldots$$

The series on the right diverges, since its partial sums grow without bound.
By comparison, the harmonic series must also diverge. □

This argument relies on comparing a series whose behavior is unknown with another series whose behavior is already understood.
By showing that one series is always larger than a divergent series (or smaller than a convergent one), we can often determine whether it diverges or converges without computing its sum.

A further class of series is the class of alternating series, that is, series whose terms alternate between positive and negative values. A useful method to check convergence of these series is with the Leibniz Criterion, also called the alternating series test.

**Theorem 2.10** (Leibniz Criterion)**.** *Let* $(a_k)$ *be a sequence of positive real numbers such that*

1. $a_{k+1} \leq a_k$ *for all* $k$ *(the sequence is non-increasing),*
2. $\lim_{k\to\infty} a_k = 0$.

*Then the series* $\sum_{k=1}^{\infty} (-1)^k a_k$ *converges.*

For a proof of this theorem see Tamm [Tam, p.14].

*Example* 2.11. The alternating harmonic series

$$\sum_{k=1}^{\infty} \frac{(-1)^k}{k} = -1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \dots$$

satisfies the conditions of the Leibniz criterion: $\frac{1}{k}$ is positive, decreasing, and tends to 0. Therefore, the series converges.

The alternating harmonic series provides an important example: it converges by the Leibniz criterion, even though the corresponding series with positive terms,

$$\sum_{k=1}^{\infty} \frac{1}{k},$$

diverges.
This shows that the convergence of a series may depend crucially on the signs of its terms. To make this distinction precise we introduce the notions of absolute and conditional convergence.

**Definition 2.12.** A series $\sum_{k=0}^{\infty} a_k$ is said to *converge absolutely* if the series of absolute values $\sum_{k=0}^{\infty} |a_k|$ converges.

**Definition 2.13.** A series $\sum_{k=0}^{\infty} a_k$ is said to *converge conditionally* if $\sum_{k=0}^{\infty} a_k$ converges but $\sum_{k=0}^{\infty} |a_k|$ diverges.

Absolute convergence is strictly stronger than ordinary convergence: every absolutely convergent series converges.

**Theorem 2.14.** *Let $(a_k)_{k\geq 1}$ be a series of real numbers, then*

$$\sum_{k=1}^{\infty} |a_k| \text{ converges} \implies \sum_{k=1}^{\infty} a_k \text{ converges.}$$

For a proof of this theorem see Tamm [Tam, p.11].

*Example* 2.15. Consider the series

$$\sum_{k=1}^{\infty} \frac{(-1)^k}{k^2}.$$

It is not immediately obvious whether this series converges. However, considering the series of absolute values gives

$$\sum_{k=1}^{\infty} \left| \frac{(-1)^k}{k^2} \right| = \sum_{k=1}^{\infty} \frac{1}{k^2}.$$

Since $\sum \frac{1}{k^2}$ is a $p$-series with $p = 2 > 1$, it converges by Theorem 2.8. Therefore, the original series converges absolutely.

*Remark* 2.16. Absolute convergence implies convergence, but the converse is not true: there exist convergent series that fail to converge absolutely.

To see this, let us revisit Example 2.11, where we showed that the series

$$\sum_{k=1}^{\infty} \frac{(-1)^k}{k}$$

converges. However, the series of absolute values

$$\sum_{k=1}^{\infty} \left| \frac{(-1)^k}{k} \right| = \sum_{k=1}^{\infty} \frac{1}{k},$$

is the harmonic series which diverges (as we saw in Example 2.9). So, the alternating harmonic series converges conditionally but not absolutely.

*Remark* 2.17. In general, absolutely convergent series are simpler to analyse than conditionally convergent ones, since there are relatively few standard tests available to determine conditional convergence.

Now, recall the series from the introduction:

$$\sum_{k=1}^{\infty} \frac{X_k}{k},$$

where $(X_k)_{k \geq 1}$ is a sequence of independent random variables taking values $+1$ and $-1$ with equal probability. We asked ourselves whether or not it converges.

The alternating harmonic series provides a useful reference point. In the introduction, we reasoned intuitively that perfect alternation of signs could allow the slowly decreasing terms $1/k$ to cancel just enough to produce convergence. Here, we see that this intuition is correct: the series with perfectly alternating signs does indeed converge. The positive and negative terms balance precisely, turning a divergent series (the harmonic series) into a convergent one. But in the random series, the signs are no longer perfectly alternating. Whether the partial sums settle down or drift to infinity remains an open question. Thus, while the alternating harmonic series confirms our intuition about perfect alternation, it also highlights the challenge of the random case: cancellation can work, but randomness complicates the story.

## 2.2 Probability theory

**Definition 2.18.** Let $\Omega \neq \emptyset$ be a set. Let $\mathcal{P}(\Omega)$ be its power set. Then $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is called a $\sigma$-algebra if it satisfies the following properties:

1. $\mathcal{F}$ contains the sample space: $\Omega \in \mathcal{F}$,
2. $\mathcal{F}$ is closed under complements: If $A \in \mathcal{F}$, then $A^C \in \mathcal{F}$,
3. $\mathcal{F}$ is closed under countable unions: If $A_i \in \mathcal{F}$ for $i = 1, 2, ...$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

*Example* 2.19. Let $\Omega = \{a, b, c\}$. Consider the collection of subsets

$$\mathcal{F} = \Big\{\emptyset, \{a\}, \{b, c\}, \Omega\Big\}.$$

We claim that $\mathcal{F}$ is a $\sigma$-algebra on $\Omega$. To check this, we verify the defining properties:

1. *Contains the empty set and the whole set:* clearly, $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$.
2. *Closed under complement:* each set in $\mathcal{F}$ has its complement in $\mathcal{F}$

$$\emptyset^C = \Omega \in \mathcal{F}, \quad \{a\}^C = \{b, c\} \in \mathcal{F}, \quad \{b, c\}^C = \{a\} \in \mathcal{F}, \quad \Omega^C = \emptyset \in \mathcal{F}.$$

3. *Closed under countable unions:* since $\mathcal{F}$ is finite, we only need to check unions of its elements. In this case the only interesting union to check is

$$\{a\} \cup \{b, c\} = \Omega$$

which is in $\mathcal{F}$.

All the conditions are met, thus, $\mathcal{F}$ is a $\sigma$-algebra on $\Omega$.

It's crucial to note that the choice of $\mathcal{F}$ is often not unique. Depending on the context, different $\sigma$-algebras may be chosen to capture varying levels of detail about the sample space $\Omega$. For instance, consider $\mathcal{F} = \{\emptyset, \{a, c\}, \{b\}, \Omega\}$. This is also a $\sigma$-algebra on $\Omega$.

In fact, the power set $\mathcal{P}(\Omega)$ is always a $\sigma$-algebra on $\Omega$. However, the power set is often much larger than necessary and contains many subsets that may not be relevant for the particular events we want to consider. Choosing a smaller $\sigma$-algebra allows us to focus only on the sets of interest, without needing to consider every subset of $\Omega$ individually.

**Definition 2.20.** A pair $(\Omega, \mathcal{F})$, where $\Omega$ is a set and $\mathcal{F}$ is a $\sigma$-algebra on $\Omega$, is called a *measurable space*. An element of $\mathcal{F}$ is called a $\mathcal{F}$-measurable subset of $\Omega$.

**Definition 2.21.** Let $(\Omega, \mathcal{F})$ be a measurable space. A function $P : \mathcal{F} \to [0, 1]$ is called a *probability measure* if it satisfies the following properties:

1. Non-negativity: For all $A \in \mathcal{F}$, $P(A) \geq 0$.

2. Normalization: $P(\Omega) = 1$.

3. Countable additivity: For any countable collection of pairwise disjoint sets $(A_i)_{i=1}^{\infty} \subseteq \mathcal{F}$,
$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i).$$

Note: If the normalization condition $P(\Omega) = 1$ is replaced by allowing $\mu(\Omega)$ to take any value in $[0, \infty]$, the function $\mu$ is called a *measure*, and $(\Omega, \mathcal{F}, \mu)$ is called a *measure space*.

**Definition 2.22.** A probability space is a triple $(\Omega, \mathcal{F}, P)$ consisting of three components:

- The sample space $\Omega$, which is the non-empty set of all possible outcomes.

- The event space $\mathcal{F}$, which is a $\sigma$-algebra on $\Omega$, contains all possible sets of things (events) that can happen.

- The probability measure $P$ which assigns each event in $\mathcal{F}$ a probability $[0, 1]$.

Note that probability spaces are by definition always measurable spaces.

Let us look at an easy example of a probability space.

*Example* 2.23. Consider the experiment of flipping a fair coin once. The outcome is either heads (H) or tails (T). Hence the sample space is

$$\Omega = \{H, T\}.$$

The $\sigma$-algebra of events is the power set of $\Omega$:

$$\mathcal{F} = \Big\{\emptyset, \{H\}, \{T\}, \Omega\Big\}.$$

A probability measure $P$ is defined by

$$P(\{H\}) = P(\{T\}) = \frac{1}{2}, \quad P(\emptyset) = 0, \quad P(\Omega) = 1.$$

Thus, $(\Omega, \mathcal{F}, P)$ is a probability space describing a single fair coin flip.

Suppose instead that our experiment is to pick a number randomly from the interval $[0, 1]$. Then $\Omega = [0, 1]$. One might consider defining the event space as the full power set $\mathcal{P}([0, 1])$, but this set is very large, in fact too large to define a consistent probability measure. More generally, if we were to try to create an event space "from scratch", it would be practically impossible to verify that all the $\sigma$-algebra conditions hold for every set individually (like we did in Example 2.19). This motivates the introduction of a Borel-$\sigma$-algebra.

**Definition 2.24.** Let $S$ be a set and let $\mathcal{G} \subseteq \mathcal{P}(S)$ be a collection of subsets of $S$. The *$\sigma$-algebra generated by $\mathcal{G}$*, denoted $\sigma(\mathcal{G})$, is the smallest $\sigma$-algebra on $S$ containing $\mathcal{G}$. Formally,

$$\sigma(\mathcal{G}) := \bigcap\{\mathcal{A} \subseteq \mathcal{P}(S) : \mathcal{A} \text{ is a } \sigma\text{-algebra and } \mathcal{G} \subseteq \mathcal{A}\}.$$

Using this construction, we define the Borel $\sigma$-algebra and Borel sets.

**Definition 2.25.** Let $S$ be a topological space and let $\mathcal{G}$ be a family of open subsets of $S$. The *Borel $\sigma$-algebra* on $S$ is defined as

$$\mathcal{B}(S) := \sigma(\mathcal{G}).$$

A subset $A \subseteq S$ is called a *Borel set* if $A \in \mathcal{B}(S)$.

This is the smallest $\sigma$-algebra containing all open sets.

*Remark* 2.26. In the case $S = \mathbb{R}$, the standard choice is $\mathcal{G} = \{(a, b) : a < b\}$, generating the usual Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R})$.

*Remark* 2.27. In the examples of this thesis, the sample space $\Omega$ is typically finite or discrete. In such settings, $\sigma$-algebras are straightforward: every subset of a discrete space behaves well, issues that arise in uncountable spaces (such as $[0, 1]$) simply do not occur.
Thus, while the Borel $\sigma$-algebra is essential for defining probabilities on continuous spaces, it plays no significant role when $\Omega$ is finite or countably discrete, as we will see in the next example.

**Definition 2.28.** Let $(\Omega, \mathcal{F}, P)$ be a probability space. A *random variable* is a function

$$X : \Omega \to \mathbb{R}$$

such that for every Borel set $B \in \mathcal{B}(\mathbb{R})$,

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

In other words, $X$ is measurable with respect to $\mathcal{F}$ and the Borel $\sigma$-algebra on $\mathbb{R}$.

*Example* 2.29. Consider the previous example of flipping a fair coin. We can define a random variable $X : \Omega \to \mathbb{R}$ that models a payoff of $+1$ for heads and $-1$ for tails. For each outcome $\omega \in \Omega$ we set

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = H, \\ -1 & \text{if } \omega = T. \end{cases}$$

Then for any Borel set $B \in \mathbb{R}$ we have $X^{-1}(B) \in \mathcal{F}$, so $X$ is a valid random variable. Note that when $B$ contains values that are not attained by $X$, for example $B = \{2\}$, we have $X^{-1}(\{2\}) = \emptyset$ and $\emptyset \in \mathcal{F}$.

Although we may view $X$ as taking values in $\mathbb{R}$, this is more structure than we need. Because $X$ only assumes the two values $-1$ and 1, a more natural approach in this discrete setting is to regard it as a function $X : \Omega \to \{-1, 1\}$. The corresponding event space generated by $X$ is then

$$\sigma(X) = \{\emptyset, \{1\}, \{-1\}, \Omega\},$$

which is much smaller and easier to work with than the entire Borel $\sigma$-algebra on $\mathbb{R}$.

The $\sigma$-algebra generated by a random variable $X$, we write $\sigma(X)$, represents the total information content of $X$. Informally, it is the collection of all questions you can answer about the outcome of a random experiment if you are only allowed to observe the value of $X$. If you know $X$ then $\sigma(X)$ describes exactly what events (subsets of possible outcomes) you can now say for sure have happened (or not happened).

*Example* 2.30. Now we flip a fair coin twice. For each head we get $+1$ point and for tails $-1$ points. The sample space is

$$\Omega = \{HH, HT, TH, TT\}.$$

Define the random variable $Y : \Omega \to \{-2, 0, 2\}$ by letting $Y(\omega)$ be the sum of points:

$$Y(\omega) = \begin{cases} 2 & \text{if } \omega = HH \\ 0 & \text{if } \omega = HT \text{ or } w = TH \\ -2 & \text{if } \omega = TT. \end{cases}$$

Intuitively, the $\sigma$-algebra $\sigma(Y)$ represents the collection of events that can be described in terms of the value of $Y$, i.e. the information we have if we only observe the total number of heads but not the exact sequence of flips.

- If $Y = -2$, then the outcome must have been $TT$.
- If $Y = 2$, then the outcome must have been $HH$.
- If $Y = 0$, then the outcome could have been either $HT$ or $TH$, and these two outcomes cannot be distinguished given only the value of $Y$.

Thus, the possible "information states" are the sets $\{TT\}, \{HH\}$, and $\{HT, TH\}$. Since a $\sigma$-algebra must also contain complements, unions, and the whole sample space, we obtain

$$\sigma(Y) = \{\emptyset, \{TT\}, \{HH\}, \{HT, TH\}, \{TT, HH\}, \{HH, HT, TH\}, \{TT, HT, TH\}, \Omega\}.$$

This example illustrates how, for a random variable $X$, $\sigma(X)$ provides the minimal $\sigma$-algebra containing all information determined by $X$. It is one way of choosing the event space $\mathcal{F}$, ensuring that it is as small as possible while still capturing all events measurable with respect to $X$.

**Definition 2.31.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X$ be a random variable. The *expected value* of $X$ is defined as the Lebesgue integral of $X$ with respect to the probability measure $P$:

$$E[X] := \int_\Omega X dP = \int_\Omega X(\omega) P(d\omega).$$

We will not go into the technical details of the Lebesgue integral here, as they are not essential for the purposes of this thesis; see, for example [Wil91, Chapter 6], for a rigorous treatment. Intuitively, the expectation represents the "average" value of $X$ when sampled according to the probability distribution $P$.

In the special case where $X$ is a discrete random variable taking values $x_i$ with probabilities $p_i = P(X = x_i)$, the expectation reduces to the sum

$$E[X] = \sum_i x_i p_i.$$

Here the index set $\{i\}$ may be finite or infinite; in the infinite case, the sum is understood as the limit of its partial sums (provided the series converges).

*Example* 2.32. Consider a single flip of a fair coin, and let $X$ be defined as in Example 2.29. The random variable $X$ will take the values $x_1 = 1$ and $x_2 = -1$. The respective probabilities $p_1 = P(X = 1)$ and $p_2 = P(X = -1)$ are $p_1 = p_2 = \frac{1}{2}$. We now compute the expected value

$$E[X] = \sum_{i=1}^{2} x_i p_i = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0.$$

Note that although $E[X] = 0$, the random variable $X$ never takes the value 0. The expectation should therefore not be interpreted as a typical outcome or median, but as a long-run average.

**Theorem 2.33.** *Let $X, Y$ be random variables with finite expectation. Let $a \in \mathbb{R}$.*

*Then the following properties hold:*

$$
\begin{aligned}
&E[aX] = aE[X] \\
&E[c] = c \qquad \textit{for any constant } c \in \mathbb{R} \\
&E[X+Y] = E[X] + E[Y].
\end{aligned}
$$

**Definition 2.34.** Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable. We say that $X$ is *integrable* if its expectation exists in the sense that $E[|X|] < \infty$.

**Definition 2.35.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X : \Omega \to \mathbb{R}$ be an integrable random variable. The *variance* of $X$ is defined as

$$Var(X) := E\Big[(X - E[X])^2\Big]$$

provided the right-hand side is finite.

A more common and often convenient way to compute the variance follows from expanding the square inside the expectation: $Var(X) = E[X^2] - E[X]^2$.

*Example* 2.36. Reconsider Example 2.30 of flipping a fair coin twice, where $Y$ is the sum of points and the sample space is $\Omega = \{HH, HT, TH, TT\}$. Each outcome has probability $\frac{1}{4}$. Using the discrete formula for expectation, we obtain

$$E[Y] = 2 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + (-2) \cdot \frac{1}{4} = 0.$$

To compute the variance, we first evaluate

$$
\begin{aligned}
&E[Y^2] = 2^2 \cdot \frac{1}{4} + 0^2 \cdot \frac{1}{4} + 0^2 \cdot \frac{1}{4} + (-2)^2 \cdot \frac{1}{4} = 2, \\
&E[Y]^2 = 0^2 = 0.
\end{aligned}
$$

Hence,

$$Var(Y) = E[Y^2] - E[Y]^2 = 2.$$

**Definition 2.37.** Let $X_1, ..., X_n$ be random variables. They are *independent* if for every choice of Borel sets $B_1, ..., B_n \subseteq \mathbb{R}$,

$$P(X_1 \in B_1, .., X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i).$$

Intuitively, the variables do not "influence" each other: information about one variable does not help you predict the others.

*Example* 2.38. Let $X_1$ and $X_2$ denote the outcomes of two fair coin tosses, with

$$P(X_i = H) = P(X_i = T) = \tfrac{1}{2}.$$

Knowing that $X_1 = H$ provides no information about the outcome of the second toss: we still have $P(X_2 = H) = P(X_2 = T) = \frac{1}{2}$. Thus, $X_1$ and $X_2$ are independent.

In contrast, let $D$ be the outcome of a fair die roll, and define $Y$ by

$$Y = \begin{cases} 1, & \text{if } D \text{ is even,} \\ 0, & \text{if } D \text{ is odd.} \end{cases}$$

Although $P(Y = 1) = 1/2$, learning the value of $D$ affects our prediction of $Y$. For example, if we are told that $D = 2$, then $Y = 1$ with probability 1. Since knowing $X$ changes the probabilities for $Y$, the random variables $D$ and $Y$ are not independent.

**Theorem 2.39.** *Let $X, Y$ be independent random variables with finite variances. Let $a \in \mathbb{R}$. Then the following properties hold:*

$$\begin{aligned} &Var(aX) = a^2 Var(X) \\ &Var(X + a) = Var(X) \\ &Var(X + Y) = Var(X) + Var(Y). \end{aligned}$$

**Definition 2.40.** Let $(\Omega, \mathcal{F})$ be a measurable space. A collection $\mathcal{G} \subseteq \mathcal{F}$ is called a *sub-$\sigma$-algebra* of $\mathcal{F}$ if $\mathcal{G}$ is itself a $\sigma$-algebra on $\Omega$.

*Example* 2.41. The simplest example of a sub-$\sigma$-algebra is the trivial $\sigma$-algebra

$$\mathcal{G} = \{\emptyset, \Omega\}.$$

It is a sub-$\sigma$-algebra of any $\sigma$-algebra $\mathcal{F}$ on $\Omega$, since it is the smallest $\sigma$-algebra on $\Omega$.

**Definition 2.42.** Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X$ be an integrable random variable. Let $\mathcal{G}$ be a sub-$\sigma$-algebra of $\mathcal{F}$. The *conditional expectation* of $X$

given $\mathcal{G}$, denoted $E[X \mid \mathcal{G}]$, is any $\mathcal{G}$-measurable random variable $Y$ such that

$$\int_A Y \, dP = \int_A X \, dP \qquad \text{for all } A \in \mathcal{G}.$$

Intuitively, $E[X \mid \mathcal{G}]$ represents the "best prediction" of $X$ based on the information contained in $\mathcal{G}$. In the case of a finite probability space, conditioning corresponds to restricting to the relevant outcomes and renormalising the probabilities.

Once more, consider the experiment of flipping a fair coin twice, and let $X$ be the random variable defined in Example 2.30. Suppose we condition on the event that *the first flip is heads*. Under this condition, the outcomes $TT$ and $TH$ have probability zero, while the remaining outcomes $HH$ and $HT$ are equally likely, since their probabilities depend only on the second flip, which is fair. Each therefore has probability $\frac{1}{2}$.

Hence, the conditional expectation is

$$E[X \mid \text{first flip is heads}\,] = 2 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = 1.$$

It is not always straightforward to determine these new probabilities by intuition alone, especially in more complex or continuous settings. In such cases, it is often convenient to use a general result, known as Bayes' Theorem, which provides a systematic way to compute these probabilities, known as conditional probabilities.

**Definition 2.43.** Let $(\Omega, \mathcal{F}, P)$ be a probability space. For events $A, B \in \mathcal{F}$ with $P(B) > 0$ the *conditional probability* of $A$ given $B$ is

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}.$$

Note: From the definition, whenever $P(B) > 0$ one has

$$P(A \cap B) = P(B)\, P(A \mid B).$$

This identity will be used repeatedly.

**Theorem 2.44** (Bayes' theorem)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A, B \in \mathcal{F}$ satisfy $P(A) > 0$ and $P(B) > 0$. Then*

$$P(A \mid B) = \frac{P(A)\, P(B \mid A)}{P(B)}.$$

We can use Bayes' theorem to compute the conditional probabilities needed in the previous example. For instance, consider the probability that both flips are heads, given that the first flip is heads:

$$P(HH \mid \text{ first flip is heads }) = \frac{P(HH) \cdot P(\text{ first flip is heads } \mid HH)}{P(\text{ first flip is heads })}.$$

The probability of the event "first flip is heads" is $P(\text{first flip is heads}) = \frac{1}{2}$, and since $P(HH) = \frac{1}{4}$ and $P(\text{first flip is heads} \mid HH) = 1$, we obtain

$$P(HH \mid \text{ first flip is heads }) = \frac{\frac{1}{4} \cdot 1}{\frac{1}{2}} = \frac{1}{2}.$$

By the same reasoning, one can determine the conditional probabilities for all outcomes in $\Omega$ and then compute the conditional expectation accordingly.

**Theorem 2.45** (Properties of Conditional Expectation)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}, \mathcal{H}$ be sub-$\sigma$-algebras of $\mathcal{F}$, and let $X, Y, Z$ be integrable real-valued random variables. All statements below are understood to hold almost surely.*

*a. If $X$ is $\mathcal{G}$-measurable, then $E[X \mid \mathcal{G}] = X$.*

*b. (Linearity) For $a, b \in \mathbb{R}$, $E[aX + bY \mid \mathcal{G}] = aE[X \mid \mathcal{G}] + bE[Y \mid \mathcal{G}]$.*

*c. (Positivity) If $X \geq 0$, then $E[X \mid \mathcal{G}] \geq 0$.*

*d. (Tower Property) If $\mathcal{H} \subseteq \mathcal{G}$, then $E\left[E[X \mid \mathcal{G}] \mid \mathcal{H}\right] = E[X \mid \mathcal{H}]$.*

*e. (Known factors) If $Z$ is $\mathcal{G}$-measurable and bounded, then $E[ZX \mid \mathcal{G}] = ZE[X \mid \mathcal{G}]$.*

*f. (Role of Independence) If $X$ is independent of $\mathcal{G}$, then $E[X \mid \mathcal{H}] = E[X]$.*

For the proof see Billingsley [Bil95, p.468, p.470].

We now turn our attention to sequences of random variables. Just as with sequences of real numbers, we can ask whether a sequence (or series) of random variables converges. However, because random variables are defined on probability spaces, we often seek convergence in a probabilistic sense — such as convergence *almost surely*, or for short a.s.

**Definition 2.46.** Let $(X_k)_{k\geq 1}$ be random variables on a probability space $(\Omega, \mathcal{F}, P)$, and define the partial sums

$$S_n := \sum_{k=1}^{n} X_k, \qquad n \geq 1.$$

We say the series $\sum_{k=1}^{\infty} X_k$ *converges almost surely* if there exists a random variable $S$ such that

$$P\Big(\Big\{\omega \in \Omega : \lim_{n\to\infty} S_n(\omega) = S(\omega)\Big\}\Big) = 1.$$

Note that, for each fixed outcome $\omega \in \Omega$, the sequence $(X_k(\omega))_{k\geq 1}$ is just a sequence of real numbers. In this sense, the infinite sum $X_1(\omega) + X_2(\omega) + \dots$ behaves like a classical real series. The distinction is that in probability theory, we study the behavior of this sum as $\omega$ varies, and ask whether convergence occurs for "almost all" outcomes — that is, with probability 1.

*Example* 2.47. Let $(X_k)_{k\geq 1}$ be independent random variables defined as

$$X_k = \begin{cases} k, & \text{with probability } 2^{-k}, \\ 0, & \text{with probability } 1 - 2^{-k}. \end{cases}$$

We are interested in the convergence of the random series

$$\sum_{k}^{\infty} X_k.$$

At first glance, it may not be obvious whether this series converges or not. Observe that each $X_k$ is large with small probability and zero otherwise, suggesting that only finitely many terms may be nonzero. To formalise this intuition, we introduce a classical result from probability theory: the first Borel–Cantelli lemma.

But fist, we recall the following definitions.

**Definition 2.48.** Let $(A_n)_{n\geq 1}$ be a sequence of sets. The *limit superior* and the *limit inferior* are defined by

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} A_j.$$

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} A_j.$$

These notions describe the long-term behavior of a sequence of sets. To understand them, imagine moving further and further out along the sequence $A_n$. For each starting point $n$, the set $\cup_{j=n}^{\infty} A_j$ consists of all points that appear at least once among the sets from index $n$ onward. Taking the intersection over all $n$ forces a point to appear in every such union, meaning it must occur in $A_j$ for infinitely many indices $j$. This is exactly the set $\limsup A_n$:

$$x \in \limsup A_n \iff x \in A_n \text{ for infinitely many } n.$$

Similarly, the set $\cap_{j=n}^{\infty} A_j$ contains the points that appear in every set from index $n$ onward. Taking the union over all $n$ means we include any point for which this eventually holds, even if it fails for the first few sets. This gives $\liminf A_n$:

$$x \in \liminf A_n \iff x \in A_n \text{ for all but finitely many } n.$$

These constructions therefore separate behavior that happens infinitely often from behavior that holds eventually always — meaning it fails only finitely many times.

**Lemma 2.49** (First Borel–Cantelli Lemma)**.** *Let* $(A_k)_{k\geq 1}$ *be a sequence of events in a probability space* $(\Omega, \mathcal{F}, P)$*.*

$$\textit{If } \sum_{k}^{\infty} P(A_k) < \infty, \quad \textit{then} \quad P(\limsup_{k\to\infty} A_k) = 0.$$

In words: if the total probability that the events $A_k$ occur is finite, then the probability that infinitely many of them occur is zero.

Returning to Example 2.47. Let $X_k$ be defined as before and define the events

$$A_k := \{X_k \neq 0\} = \{X_k = k\}.$$

Then $P(A_k) = 2^{-k}$. Consider the sum of the probabilities:

$$\sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^{\infty} 2^{-k} = 1 < \infty.$$

By the first Borel-Cantelli lemma, it follows that

$$P(\limsup_{k\to\infty} A_k) = 0.$$

That is, with probability 1 only finitely many of the events $A_k$ occur. Equivalently, almost surely, $X_k \neq 0$ for only finitely many values of $k$. Thus, the random series $\sum X_k$ almost surely consists of only finitely many nonzero terms and therefore converges almost surely.

Almost sure convergence means that the sequence (or series) of random variables converges for "almost every" outcome in the sample space, in the sense of probability. However, this does not imply that the series converges for all outcomes. In fact, there generally exist outcomes $\omega \in \Omega$ for which the series $\sum X_k(\omega)$ fails to converge, in Example 2.47, these outcomes would be the ones in which infinitely many $X_k \neq 0$. What the definition ensures is that the set of such non-convergent outcomes has probability zero — they are negligible from a probabilistic point of view.

This highlights a key feature of almost sure convergence: it permits exceptional outcomes where convergence fails, as long as those outcomes are contained in a set of probability zero.

In contrast, *pointwise convergence*, which is a strictly stronger condition, requires that the series converges for *every* outcome in the sample space, without exception — even those with extremely low (or zero) probability.

*Example* 2.50. Let $X_k$ be independent coin tosses taking values $\pm 1$ with equal probability. Consider

$$\sum_{k=1}^{\infty} \frac{X_k}{2^k}.$$

For each outcome $\omega$, this is an ordinary real series of the form

$$\pm \frac{1}{2} \pm \frac{1}{4} \pm \frac{1}{8} \pm \dots$$

To study its convergence we examine the series of absolute values:

$$\sum_{k=1}^{\infty} \left| \frac{X_k}{2^k} \right| = \sum_{k=1}^{\infty} \frac{1}{2^k}.$$

This is the geometric series with ratio $\frac{1}{2}$ that converges (see Chapter 2.1). Hence the original series converges absolutely for every $\omega$, and therefore converges pointwise.

This series is structurally similar to the series $\sum \frac{X_k}{k}$ introduced in the introduction. However, the argument used above relies crucially on the summability of the deterministic sequence $(2^{-k})$. Since the harmonic series $\sum \frac{1}{k}$ diverges, this approach cannot be applied to the series $\sum \frac{X_k}{k}$. We must first develop additional tools, which will eventually allow us to address this problem.

**Lemma 2.51** (Second Borel-Cantelli Lemma)**.** *If $A_n$ is a sequence of independent events and $\sum P(A_n) = \infty$, then $P(\limsup A_n) = 1$.*

In simpler terms: if a sequence of independent events has probabilities that sum to infinity, then the probability that infinitely many of these events occur is 1.

While the first Borel-Cantelli lemma shows that events with summable probabilities occur only finitely often a.s., the second lemma tells us that for events with divergent probabilities, infinitely many occor a.s. Both lemmas together illustrate that long-term behavior of independent sequences is often deterministic in probabilities.

**Definition 2.52.** Let $X_1, X_2, ...$ be random variables. Define

$$\mathcal{T}_n := \sigma(X_{n+1}, X_{n+2}, ...), \quad \mathcal{T} := \cap_n \mathcal{T}_n.$$

The $\sigma$-algebra $\mathcal{T}$ is called the *tail $\sigma$-algebra* of the sequence $(X_n : n \in \mathbb{N})$.

*Remark* 2.53. The tail $\sigma$-algebra contains all the events whose occurence depends only on the "far future" of the sequence $(X_n)$. Modifying finitely many of the $X_k$ does not change whether these events occur. Important examples include:

$F_1 := (\lim X_k \text{ exists }) := \{\omega : \lim_x X_k(\omega) \text{ exists }\}$,

$F_2 := (\sum X_k \text{ converges })$.

**Theorem 2.54** (Kolmogorov's $0-1$ Law)**.** *Let $(X_n : n \in \mathbb{N})$ be a sequence of independent random variables, and let $\mathcal{T}$ be the tail $\sigma$-algebra of $(X_n : n \in \mathbb{N})$. Then $\mathcal{T}$ is $P$-trivial, that is,*

*(i)* $F \in \mathcal{T} \implies P(F) = 0$ *or* $P(F) = 1$,

*(ii) if $\xi$ is $\mathcal{T}$-measurable random variable, then, $\xi$ is almost deterministic in that for some constant $c$ in $[-\infty, \infty]$,*

$$P(\xi = c) = 1.$$

In other words, any event in the tail $\sigma$-algebra is trivial in probability. This includes the convergence events such as $F_1$, $F_2$ above. Hence, for independent sequences, the event that the series $\sum X_k$ converges either has probability 1 or 0 — there is no intermediate chance.

Once again, we recall our random series from the introduction. With this theorem we can observe that the event

$$\left\{\sum \frac{X_k}{k} \text{ converges }\right\}$$

belongs to the tail $\sigma$-algebra of the sequence $(X_k)$. Indeed, convergence of the series is unaffected by finite modification of the sequence. Consequently, the convergence does not depend on any initial segment of the sequence.

Since the variables $X_k$ are independent, Kolmogorov's $0-1$ law applies. It follows that the probability of convergence of the series must be either 0 or 1.

We will now move on to look at ways to control the size of random variables, in particular through their expected values and moments. This naturally leads us to the concept of $\mathcal{L}^p$ spaces.

**Definition 2.55.** Let $(X, \mathcal{M}, \mu)$ be a measure space. For a real number $p$ such that $1 \leq p < \infty$, we define the space $\mathcal{L}^p(X, \mathcal{M}, \mu)$ (often just written $\mathcal{L}^p(\mu)$) as

$$\mathcal{L}^p(\mu) = \left\{ f : X \to \mathbb{C} \mid f \text{ is } \mathcal{M}\text{-measurable and } \int_X |f|^p d\mu < \infty \right\}.$$

In other words, a function $f$ belongs to $\mathcal{L}^p$ if the integral with respect to $\mu$ of $|f(x)|^p$ is finite.

**Definition 2.56.** A sequence of random variables $(X_n)$ is said to be bounded in $\mathcal{L}^1$, if there exists a finite constant $K$ such that

$$E[|X_n|] \leq K < \infty, \text{ for all } n.$$

**Definition 2.57.** A sequence of random variables $(X_n)$ is said to be bounded in $\mathcal{L}^2$, if there exists a finite constant $K$ such that

$$E[|X_n|^2] \leq K < \infty, \text{ for all } n.$$

**Theorem 2.58.** *Let $(X_n)$ be a sequence of random variables. If $(X_n)$ is bounded in*

*$\mathcal{L}^2$, then it is also bounded in $\mathcal{L}^1$.*

To prove this we will use the Cauchy-Schwarz inequality, that states that for any two random variables $X, Y$ bounded in $\mathcal{L}^2$

$$(E[|XY|])^2 \leq E[X^2] \cdot E[Y^2].$$

*Proof.* We assume $(X_n)$ is bounded in $\mathcal{L}^2$. We apply the Cauchy-Schwarz inequality cleverly choosing $Y$ to be the constant random variable 1 and get

$$(E[1 \cdot |X_n|])^2 \leq E[1^2] \cdot E([X_n^2].$$

Simplifying gives

$$(E[|X_n|])^2 \leq E[X_n^2].$$

Since, we assumed $X_n$ bounded in $\mathcal{L}^2$ we have

$$(E[|X_n|])^2 \leq E[X_n^2] \leq K.$$

We take the square root and get

$$(E[|X_n|]) \leq \sqrt{E[X_n^2]} \leq \sqrt{K}.$$

The above inequality holds for every $n$ in the sequence. Therefore we have found a constant $M = \sqrt{K} < \infty$ such that:

$$E[|X_n|] \leq M < \infty, \text{ for all } n.$$

This is precisely the definition of the sequence $(X_n)$ being bounded in $\mathcal{L}^1$. □

# 3 Martingales

Martingales are a fundamental class of stochastic processes that formalise the concept of a "fair game." We will find them to be very helpful for studying the convergence of random series.

Before defining martingales, we introduce some basic concepts needed for their precise formulation. In particular, we review stochastic processes and the mathematical structures used to describe the evolution of information over time.

## 3.1 Stochastic processes

**Definition 3.1.** A *stochastic process* is a family of random variables

$$\{X_t : t \in T\},$$

indexed by a set $T$, called the *index set*.

By convention, we will use $t$ for continuous-time processes ($T \subset \mathbb{R}$) and $n$ for discrete-time processes ($T \subset \mathbb{N}$).

A stochastic process can be thought of as a random system evolving over time. Each random variable $X_t$ represents the state of the system at a particular time $t$, and the entire collection $(X_t)_{t \in T}$ captures all possible ways the system can evolve.

*Example* 3.2. Let $(X_n)_{n \geq 1}$ be a process defined by

$$X_n = \begin{cases} n & \text{if the } n\text{-th coin toss is heads,} \\ -n & \text{if the } n\text{-th coin toss is tails.} \end{cases}$$

Here, $n$ indexes the toss number. Each $X_n$ is a random variable taking values $\pm n$, depending on the outcome of the corresponding coin toss. A single outcome of the process is a sequence $(X_1, X_2, X_3 \dots)$, representing the results of all tosses.

For a stochastic process, we can define the expectation and variance of each $X_t$ just as for any random variable, and all the usual properties hold. Conditional expectation and variance also carry over naturally, where the conditioning is usually based on information accumulated up to the current index.

*Example* 3.3. Let the process $X_n$ be defined as before. The expected value is

$$E[X_n] = n \cdot \frac{1}{2} + (-n) \cdot \frac{1}{2} = 0,$$

and the variance is

$$Var(X_n) = E[X_n^2] - E[X_n]^2 = n^2.$$

*Example* 3.4. Let $X_n$ be the coin-toss process defined previously, and define a new process

$$Y_n = X_1 + X_2 + \cdots + X_n,$$

so that $Y_n$ represents the cumulative sum of the first $n$ tosses.

Suppose we know the first toss was heads, $X_1 = +1$, so $Y_1 = 1$. Then the conditional expectation of $Y_2$ given this information is

$$\begin{aligned} E[Y_2 \mid Y_1 = 1] &= E[X_1 + X_2 \mid X_1 = 1] \\ &= E[X_1 \mid X_1 = 1] + E[X_2 \mid X_1 = 1] \\ &= 1 + E[X_2] = 1 + 0 = 1, \end{aligned}$$

since $X_2$ is independent of $X_1$.

## 3.2 Filtered space, Filtration, Adapted process

**Definition 3.5.** A filtered space is a quadruple $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, P)$ where

1. $(\Omega, \mathcal{F}, P)$ is a probability space, and,
2. $\{\mathcal{F}_n : n \geq 0\}$ is a filtration, that is an increasing family of sub-$\sigma$-algebras, of $\mathcal{F}$: $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq ... \subseteq \mathcal{F}$.

A natural way of looking at $\{\mathcal{F}_n\}$ is to think of $n$ as a moment in time and $\mathcal{F}_n$ as the collection of all information available up to that time. As $n$ increases, we accumulate more information, which is why the $\sigma$-algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq ...$ become progressively larger.

**Definition 3.6.** A process $X = (X_n)_{n \geq 0}$ is called adapted (to the filtration $\{\mathcal{F}_n\}$) if for each $n$, $X_n$ is $\mathcal{F}_n$-measurable.

Essentially, a process is adapted if, at time $n$, its value $X_n$ depends only on the information available up to that time, never on future events. For example, let $n$ denote the number of flips: at time $n$, after $n$ number of flips have occurred, the process $X_n$ giving the result of the $n$-th flip is adapted, while the process $Y_n$ giving the result of the $(n+1)$-th flip is not, since it depends on a future outcome.

## 3.3 Martingales

**Definition 3.7.** A process $M = (M_n)_{n\geq 0}$ is called a martingale (relative to $(\{\mathcal{F}_n\}, P)$) if, for all $n$:

(i) $M$ is adapted,

(ii) $M$ is integrable, meaning $E[|M_n|] < \infty$,

(iii) $M$ satisfies the martingale property: $E[M_n \mid \mathcal{F}_{n-1}] = M_{n-1}$.

*Remark* 3.8. An immediate consequence of the martingale property is that the expected value of the process remains constant over time. Indeed, taking expectations on both sides of the identity $E[M_n \mid \mathcal{F}_{n-1}] = M_{n-1}$ and using the tower property of conditional expectation yields

$$E[M_n] = E[M_{n-1}].$$

By iterating this relation, we obtain $E[M_n] = E[M_0]$ for all $n \geq 0$. Thus, a martingale has constant mean.

*Example* 3.9. Imagine a game where a fair coin is flipped repeatedly. Each time the coin lands on heads, you earn 1 point; each time it lands on tails, you lose 1 point. Define $X_n$ as the result of the $n$th flip and let

$$M_n = \sum_{k=1}^{n} X_k$$

denote the score after $n$ flips. We now verify that $(M_n)_{n\geq 1}$ is a martingale with respect to the natural filtration $\mathcal{F}_n = \sigma(X_1, ..., X_n)$.

First, $M_n$ is adapted: at time $n$ we know the outcomes $X_1, .., X_n$, hence their sum $M_n$ is $\mathcal{F}_n$-measurable.

Second, $M_n$ is integrable: since $|M_n(\omega)| \leq n$ for each $\omega$, we have $E[|M_n|] \leq n < \infty$.

Finally, we check the martingale property using $M_n = M_{n-1} + X_n$,

$$\begin{aligned} E[M_n \mid \mathcal{F}_{n-1}] &= E[M_{n-1} + X_n \mid \mathcal{F}_{n-1}] \\ &= E[M_{n-1} \mid \mathcal{F}_{n-1}] + E[X_n \mid \mathcal{F}_{n-1}] \\ &= M_{n-1} + E[X_n] \\ &= M_{n-1}. \end{aligned}$$

Thus, all three conditions are satisfied, and $M$ is a martingale.

This example illustrates how the martingale property captures the idea of a fair game: the expected future value, conditioned on the present, equals the current value. In fact, the sum of zero mean random variables is a classic example of martingales.

We will now look at an example of a process that is not a martingale.

*Example* 3.10. Consider the same coin-tossing game as above, but suppose that each time the coin lands on heads you earn 2 points, while each time it lands on tails you lose 1 point. Let $(X_n)_{n\geq 1}$ be the outcomes of the flips. Define

$$N_n = \sum_{k=1}^{n} X_k,$$

and let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ be the natural filtration.

As before, $N$ is adapted and integrable. However, it does *not* satisfy the martingale property. Indeed, since the coin is fair,

$$E[X_n] = \frac{1}{2}(2) + \frac{1}{2}(-1) = \frac{1}{2}.$$

Using the decomposition $N_n = N_{n-1} + X_n$, we obtain

$$\begin{aligned} E[N_m \mid \mathcal{F}_{n-1}] &= E[N_{n-1} + X_n \mid \mathcal{F}_{n-1}] \\ &= N_{n-1} + E[X_n] \\ &= N_{n-1} + \frac{1}{2} \neq N_{n-1}. \end{aligned}$$

Thus the martingale property fails, and $N$ is not a martingale.

## 3.4 Orthogonality, Pythagorean identity for martingales

**Definition 3.11.** Two random variables $X, Y \in \mathcal{L}^2$ are said to be *orthogonal* if

$$E[XY] = 0.$$

A sequence $(X_k)$ is said to be pairwise orthogonal in $\mathcal{L}^2$ if $E[X_k X_j] = 0$ for all $k \neq j$.

Boundedness in $\mathcal{L}^2$ for martingales is closely tied to the fact that their increments are orthogonal. This leads to a Pythagorean identity, which makes $\mathcal{L}^2$-boundedness particularly easy to check.

**Theorem 3.12** (Pythagorean identity for martingales)**.** *Let $(M_n)_{n\geq 0}$ be a martingale such that $M_n \in \mathcal{L}^2$ for all $n$. Then*

$$E[M_n^2] \;=\; E[M_0^2] + \sum_{k=1}^{n} E[(M_k - M_{k-1})^2].$$

*Proof.* Define the increments

$$d_k := M_k - M_{k-1}, \qquad k \geq 1.$$

We first show that the increments $(d_k)$ are pairwise orthogonal in $\mathcal{L}^2$. Let $i < j$. Then

$$E[d_i d_j] \;=\; E\Big[E[d_i d_j \mid \mathcal{F}_i]\Big].$$

Since $d_i$ is $\mathcal{F}_i$-measurable, this becomes

$$E[d_i d_j] = E\Big[d_i \cdot E[d_j \mid \mathcal{F}_i]\Big].$$

Now, by the tower property of conditional expectation,

$$E[d_j \mid \mathcal{F}_i] = E\Big[E[d_j \mid \mathcal{F}_{j-1}] \mid \mathcal{F}_i\Big].$$

Using the martingale property,

$$E[d_j \mid \mathcal{F}_{j-1}] = E[M_j - M_{j-1} \mid \mathcal{F}_{j-1}] = 0.$$

Hence,

$$E[d_j \mid \mathcal{F}_i] = 0,$$

and therefore

$$E[d_i d_j] = E[d_i \cdot 0] = 0.$$

Thus, the increments $(d_k)$ are orthogonal in $\mathcal{L}^2$.

Now expand $M_n$ in terms of its increments:

$$M_n = M_0 + \sum_{k=1}^{n} d_k.$$

Squaring gives

$$M_n^2 = M_0^2 + \sum_{k=1}^{n} d_k^2 + 2\sum_{i<j} d_i d_j.$$

Taking expectations and using orthogonality of the increments, the cross terms vanish:

$$E[M_n^2] = E[M_0^2] + \sum_{k=1}^{n} E[d_k^2].$$

This is precisely the claimed identity. □

*Remark* 3.13. The identity can be interpreted as a version of the Pythagorean theorem in the Hilbert space $\mathcal{L}^2$: the decomposition

$$M_n = M_0 + d_1 + \cdots + d_n$$

is an orthogonal sum, so the squared norm satisfies

$$\|M_n\|_2^2 = \|M_0\|_2^2 + \sum_{k=1}^{n} \|d_k\|_2^2,$$

where $\|X\|_2^2 = E[|X|^2]$.

## 3.5 Stopping times

**Definition 3.14.** A map $T : \Omega \to \{0, 1, 2, ...; \infty\}$ is called a *stopping time* with respect to $\mathcal{F}_n$ if

(a) $\{T \leq n\} = \{\omega : T(\omega) \leq n\} \in \mathcal{F}_n, \quad \forall n \leq \infty,$

equivalently,

(b) $\{T = n\} = \{\omega : T(\omega) = n\} \in \mathcal{F}_n, \quad \forall n \leq \infty.$

Note that $T$ can assume the value $\infty$.

In words, $T$ is a stopping time if, at any time $n$, we can determine, using only the information in $\mathcal{F}_n$, whether the time $T$ has already occurred.

*Example* 3.15. Let $(X_n)_{n\geq 0}$ be a stochastic process adapted to a filtration $(\mathcal{F}_n)_{n\geq 0}$ and let $a \in \mathbb{R}$ be fixed. A classic example of a stopping time is the first hitting time

$$T := \inf\{n \geq 0 : X_n = a\}.$$

This time records the first moment at which the process reaches the value $a$. It is a stopping time because at time $n$ we can decide, using only the observations $X_0, X_1, ..., X_n$, whether this is indeed the first time the process has taken the value $a$.

In contrast, the last time the process hits $a$,

$$L := \sup\{n \geq 0 : X_n = a\},$$

is not a stopping time. Deciding whether $L = n$ requires checking that the process does not return to $a$ at any time after $n$, which depends on information strictly beyond time $n$. Therefore, such an event cannot be determined from $\mathcal{F}_n$, and $L$ fails to be a stopping time.

**Definition 3.16.** Given a process $X_n$ and a stopping time $T$, the *stopped process* $X_{T\wedge n}$ is defined by

$$X_{T\wedge n} := X_{\min(T,n)}.$$

For a martingale $M_n$ and a stopping time $T$, we introduce the notation

$$M_n^T := M_{T\wedge n},$$

so that $M_n^T$ denotes the martingale stopped at time T.

*Example* 3.17. Continuing the coin-flipping game, recall that $X_n$ is the gain on the $n$-th flip and

$$M_n = \sum_{k=1}^{n} X_k$$

is the total score after $n$ flips.

Fix a threshold $b \in \mathbb{R}$, and define the stopping time

$$T := \inf\{n \geq 0 : M_n = b\},$$

the first time the cumulative score reaches $b$.

The corresponding stopped process is

$$M^T := M_{T\wedge n} = \begin{cases} M_n, & n < T \\ M_T, & n \geq T. \end{cases}$$

Thus the process evolves normally until the score first hits the level $b$, and from that

moment on it remains constant at the value $M_T = b$. In other words, $M^T$ represents the original game, but frozen at the time you reach the target score.
To visualise the the effect of stopping, fix an outcome $\omega \in \Omega$. Along this outcome, the martingale $(M_n(\omega))$ evolves by steps of $\pm 1$ until it first reaches the first level $b$. For example, a trajectory may look like

$$M_0(\omega), M_1(\omega), M_2(\omega), \dots = 0, -1, 0, \dots, b, b, b, \dots$$

Once time $T(\omega)$ is reached, the stopped process remains constant.

**Theorem 3.18.** *If $M$ is a martingale and $T$ is a stopping time, then $M^T$ is a martingale, so that in particular,*

$$E[M_{T \wedge n}] = E[X_0], \quad \forall n.$$

For the proof see Billingsley [Bil95, p.100]

This ensures that the "fairness" of a martingale is preserved even if we stop at a random (adapted) time.

# 4 Convergence of random series

The goal of this section is to understand when a random series converges almost surely, and in particular to resolve the behavior of the series

$$\sum_{k=1}^{\infty} \frac{X_k}{k},$$

where $X_k$ are independent random variables taking the values $\pm 1$.

**Theorem 4.1** (Doob's Theorem)**.** *Let $(M_n)_{n\geq 0}$ be a martingale for which $M_n \in \mathcal{L}^2$, for all $n$. Then*

$$M \text{ is bounded in } \mathcal{L}^2 \iff \sum E[(M_k - M_{k-1})^2] < \infty; \tag{1}$$

*and when this holds,*

$$M_n \to M_\infty \text{ almost surely and in } \mathcal{L}^2.$$

*Proof of* (1). The equivalence (1) follows from the orthogonality of martingale increments. Indeed, since

$$M_n = M_0 + \sum_{k=1}^{n} (M_k - M_{k-1})$$

and the increments $(M_k - M_{k-1})$ are orthogonal in $\mathcal{L}^2$, the Pythagorean identity for martingales (Theorem 3.12) yields

$$E[M_n^2] = E[M_0^2] + \sum_{k=1}^{n} E[(M_k - M_{k-1})^2].$$

Thus, $(M_n)$ is bounded in $\mathcal{L}^2$ if and only if the series of squared increments converges. □

For the proof of the convergence of $(M_n)$ to $M_\infty$, see Williams [Wil91, p.111].

**Theorem 4.2** (Kolmogorov's Three-Series Theorem)**.** *Let $(X_n)$ be a sequence of independent random variables. Then $\sum X_n$ converges almost surely* if and only if *for some (then for every) $K > 0$, the following three properties hold:*

*(i)* $\sum_n P(|X_n| > K) < \infty,$

*(ii)* $\sum_n E(X_n^K)$ *converges,*

*(iii)* $\sum_n Var(X_n^K) < \infty$,

*where*

$$X_n^K(\omega) := \begin{cases} X_n(\omega) & \text{if } |X_n(\omega)| \leq K, \\ 0 & \text{if } |X_n(\omega)| > K. \end{cases}$$

Kolmogorov's Three-Series Theorem provides a general criterion for almost sure convergence of independent random series; the variance-based results that follow can be seen as concrete and more accessible consequences in the zero-mean setting.

The following theorem shows that, for independent zero-mean variables, almost sure convergence is completely governed by the sum of variances.

**Theorem 4.3.** *Suppose that $(X_k : k \in \mathbb{N})$ is a sequence of independent random variables such that, for every $k$,*

$$E[X_k] = 0, \qquad \sigma_k^2 := \mathrm{Var}(X_k) < \infty.$$

*(a) Then*

$$\sum \sigma_k^2 < \infty \quad \textit{implies that} \quad \sum X_k \text{ converges a.s..}$$

*(b) If the variables $(X_k)$ are bounded by some constant $K \in [0, \infty)$ in the sense that $|X_k(\omega)| \leq K$, for all $k$ and for all $\omega$, then*

$$\sum X_k \text{ converges a.s.} \quad \textit{implies that} \quad \sum \sigma_k^2 < \infty.$$

*Proof.* **(a)**
We define

$$M_n := X_1 + X_2 + \cdots + X_n \quad \text{with} \quad M_0 := 0$$

and let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$.
We can convince ourselves that $M$ is a martingale with respect to $\mathcal{F}_n$. I will skip this step but it is essentially the same as Example 3.9.

Since the random variables $X_k$ are independent and have zero mean, the increments are orthogonal in $\mathcal{L}^2$. By Theorem 3.12 (the Pythagorean identity), it follows that

$$E[M_n^2] = E[M_0^2] + \sum_{k=1}^{n} E[(M_k - M_{k-1})^2]$$
$$= \sum_{k=1}^{n} E[X_k^2].$$

Because $E[X_k]^2 = 0$ and we have

$$E[X_k^2] = E[X_k^2] - E[X_k]^2 = Var(X_k) = \sigma_k^2,$$

and therefore

$$E[M_n^2] = \sum_{k=1}^{n} \sigma_k^2.$$

Recall that for any increasing sequence $(S_n)$ we have $\sup S_n = \lim S_n$. Now, under the assumption $\sum \sigma_k^2 < \infty$, we get

$$\sup E[M_n^2] = \sup \sum_{k=1}^{n} \sigma_k^2 = \sum_{k=1}^{\infty} \sigma_k^2 < \infty.$$

So, by Theorem 4.1 $M$ is bounded in $\mathcal{L}^2$ and by the same theorem, any martingale that is bounded in $\mathcal{L}^2$ converges almost surely. Since $M_n = X_1 + X_2 + \cdots + X_n$, converges a.s. as $n \to \infty$, the series $\sum X_k$ converges almost surely.

**(b)** Define

$$A_n := E[M_n^2] = \sum_{k=1}^{n} \sigma_k^2 \quad \text{and} \quad N := M_n^2 - A_n.$$

Then

$$N_{n+1} = M_{n+1}^2 - A_{n+1} = M_n^2 + 2M_n X_{n+1} + X_{n+1}^2 - (A_n + \sigma_{n+1}^2),$$

and taking conditional expectation with respect to $\mathcal{F}_n$ gives

$$E[N_{n+1} \mid \mathcal{F}_n] = M_n^2 - A_n = N_n,$$

hence, $N$ is a martingale.

For $c \in (0, \infty)$ define the stopping time

$$T := \inf\{r : |M_r| > c\}.$$

This is the first time the partial sums exceed the level $c$. Since $N$ is a martingale, the stopped process $N^T$ is also a martingale.

We now show that $(M_n^T)^2$ is uniformly bounded.

- If $T > n$, then $M_n^T = M_n$ and by definition of $T$,

$$|M_n| \leq c.$$

- If $T \leq n$, then by the triangle inequality

$$|M_n^T| = |M_T| = |M_{T-1} + X_T| \leq |M_{T-1}| + |X_T| \leq c + K,$$

  since the increments are bounded by $K$.

Thus, in all cases

$$|M_n^T| \leq c + K, \qquad \text{for all } n.$$

Because $(N_n)$ is a martingale, we have $E[N_n^T] = 0$, so

$$N_n^T = (M_n^T)^2 - A_{T\wedge n} \quad \Longrightarrow \quad E[A_{T\wedge n}] = E[(M_n^T)^2].$$

Combining with the bound on $|M_n^T|$ gives

$$E[A_{T\wedge n}] \leq (K + c)^2, \qquad \text{for all } n.$$

Finally, since $\sum X_n$ converges almost surely, the partial sums are a.s. bounded. Thus there exists $c > 0$ such that $P(T = \infty) > 0$. On the event $T = \infty$ we have $T \wedge n = n$ for all $n$, so $A_n = A_{T\wedge n} \leq (c + K)^2$ for all $n$. Being a non-decreasing sequence, $A_n$ converges, and we conclude

$$A_\infty := \sum \sigma_k^2 < \infty.$$

□

**Corollary 4.4.** *Let $a_k$ be a sequence of real numbers, and let $(\epsilon_k)$ be a sequence of IID random variables with*

$$P(\epsilon_k = \pm 1) = \frac{1}{2}.$$

*Then the series $\sum \epsilon_k a_k$ converges almost surely if and only if*

$$\sum a_k^2 < \infty.$$

*In particular, if $\sum a_k^2 = \infty$, then the series $\sum \epsilon_k a_k$ diverges almost surely.*

*Proof of Corollary 4.4.* Define $Y_k := \epsilon_k a_k$. Since $E[\epsilon_k] = 0$ and $\epsilon_k^2 = 1$, we obtain

$$E[Y_k] = a_k E[\epsilon_k] = 0, \quad \text{and} \quad Var(Y_k) = E[Y_k^2] = a_k^2 E[\epsilon_k^2] = a_k^2.$$

Thus, in the notation of Theorem 4.3, we have $\sigma_k^2 = a_k^2$.
By part (a) of Theorem 4.3, if $\sum a_k^2 < \infty$, then $\sum Y_k = \sum \epsilon_k a_k$ converges almost surely.
Conversely, by part (b), if $\sum \epsilon_k a_k$ converges a.s., then $\sum \sigma_k^2 = \sum a_k^2 < \infty$, since $|Y_k| = |\epsilon_k a_k| = |a_k|$ is uniformly bounded.

To remove the boundedness assumption in the converse, we can appeal to Kolmogorov's Three-Series Theorem. Since each $Y_k$ satisfies $E[Y_k] = 0$ and $\text{Var}(Y_k) = a_k^2$, and truncating at any level $K > 0$ does not change the variances significantly, the three conditions reduce to the convergence of $\sum a_k^2$.

Hence, $\sum \epsilon_k a_k$ converges a.s. if and only if $\sum a_k^2 < \infty$. □

We now, for the last time, return to the random harmonic series introduced at the beginning of this thesis

$$\sum_{k=1}^{\infty} \frac{X_k}{k}, \qquad P(X_k = \pm 1) = \frac{1}{2}.$$

Writing $X_k = \epsilon_k$ and $a_k = 1/k$, this series is of the form $\sum \epsilon_k a_k$, exactly as in Corollary 4.4.

Since

$$\sum_{k=1}^{\infty} a_k^2 = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty,$$

Corollary 4.4 implies that

$$\sum \frac{X_k}{k} \quad \text{converges almost surely.}$$

We have thus arrived at a definitive answer to the question posed in the introduction: the series

$$\sum_{k=1}^{\infty} \frac{X_k}{k}$$

converges almost surely when the signs $X_k$ are independent and symmetric. So, despite the slow decay of $1/k$ and the unpredictable behavior of the random signs, enough cancellation to prevent the partial sums from diverging.

It is worth emphasising that almost sure convergence is, in a natural sense, the strongest form of convergence one can reasonably expect in this setting. This convergence cannot be strengthened to pointwise convergence. Indeed, consider the realisation for which $X_k = +1$ for all $k$. For this outcome, the series reduces to the classical harmonic series, which diverges. The same goes for realisations in which all but the first $n$ terms are positive, for any $n \in \mathbb{N}$. Convergence of an infinite series is a tail property, and in these cases the tail of the series coincides with the harmonic series and therefore diverges. More generally, the same conclusion holds for any realization in which all but finitely many terms are positive, since finite modifications do not affect convergence.

Since such realisations are possible—even though they form a set of probability zero—the random series fails to converge pointwise.

A natural direction for further investigation is to consider what happens when the symmetry assumption is relaxed. If the random variables $X_k$ satisfy

$$P(X_k = 1) = p, \qquad P(X_k = -1) = 1 - p,$$

with $p \neq \frac{1}{2}$, the series acquires a nonzero drift. In this case, the cancellation mechanism that drives almost sure convergence in the symmetric setting is weakened, and different behavior may emerge. How much further from the $50 - 50$ balance can one move before convergence is lost, and is it possible to introduce a bias while still preserving almost sure convergence? This is a richer and more delicate problem which unfortunately lies beyond the scope of this thesis.

# References

[AB08] Sven Alm and Tom Britton. *Stokastik*. Studentlitteratur, Lund, 2008.

[Bil95] Patrick Billingsley. *Probability and Measure*. Wiley, University of Chicago, 2nd edition, 1995.

[Tam] Martin Tamm. Serier och generaliserade integraler. Compendium.

[Wil91] David Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.