BACHELOR

A brief exploration into divergent series in probability theory

van Wijk, Willem J.

Award date:
2020

Department of Mathematics and Computer Science
2WH40 - Bachelor final project

A brief exploration into divergent series in probability theory

W. J. van Wijk

Supervisor:
dr. Jaron Sanders

final version

Eindhoven, July 8, 2020

Abstract

Within the field of analysis, we are familiar with the convergence and divergence of series. For divergent series, there exist summation techniques that assign values to them. In probability theory, we are familiar with convergence theorems, such as the law of large numbers and the central limit theorem. In this report, we will summarize the most important theorems regarding convergence in probability theory. Furthermore, we will deal with the summation techniques for divergent series and we will investigate where and how we could use them in probability theory. We will see a reformulation of the law of large numbers using the Cesàro means method and we investigate some series of random variables that have a divergent expectation. Besides that, we see from a literature research there already exist some theorems where summation techniques are used within probability theory.

Contents

Preliminaries

In this report, we will deal with concepts from analysis and probability theory. We first summarize the most important concepts that will be considered prerequisite knowledge.

Probability theory

Definition 1 ([Grimmett and Stirzaker, 2001, p. 1]). The set of all possible outcomes of an experiment is called the *sample space* and is denoted by Ω.

Definition 2 ([Grimmett and Stirzaker, 2001, p. 3]). A collection $\mathcal{F}$ of subsets of Ω is called a σ-field if it satisfies the following conditions

- $\emptyset \in \mathcal{F}$;
- if $A_1, A_2, \dots \in \mathcal{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$;
- if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.

Definition 3 ([Grimmett and Stirzaker, 2001, p. 5]). A *probability measure* $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfying

- $\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\Omega) = 1$;
- if $A_1, A_2, \cdots$ is a collection of disjoint members of $\mathcal{F}$, in that $A_i \cap A_j = \emptyset$ for all pairs i, j satisfying $i \neq j$, then

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$, comprising a set Ω, a σ-field $\mathcal{F}$ of subsets of Ω and a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$, is called a *probability space*.

Definition 4 ([Grimmett and Stirzaker, 2001, p. 27]). A random variable is a function $X : \Omega \to \mathbb{R}$ with the property that $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$. Such a function is said to be *$\mathcal{F}$-measurable*.

All random variables have a *distribution function* that is either continuous or discrete.

Definition 5 ([Grimmett and Stirzaker, 2001, p. 27]). The *distribution function* of a random variable X is the function $F : \mathbb{R} \to [0, 1]$ given by $F(x) = \mathbb{P}(X \leq x)$.

Definition 6 ([Grimmett and Stirzaker, 2001, p. 33]). A random variable X is called *continuous* if its distribution function $F(x)$ can be expressed as

$$F(x) = \int_{-\infty}^{x} f(u)\mathrm{d}u \quad x \in \mathbb{R}$$

for some integrable function $f : \mathbb{R} \to [0, \infty)$ called the *probability density function*.

Definition 7 ([Grimmett and Stirzaker, 2001, p. 33]). A random variable X is called *discrete* if it takes values in some countable subset of $\mathbb{R}$ only. A discrete random variable X has a *probability mass function* $f : \mathbb{R} \to [0, 1]$ given by $f(x) = \mathbb{P}(X = x)$.

An important property of a random variable X is the *expectation*.

Definition 8 ([Grimmett and Stirzaker, 2001, p. 50]). The *mean value*, or *expected value*, or *expectation* of a discrete random variable X is given by

$$\mathbb{E}[X] = \sum_{i} \mathbb{P}(X = i)i,$$

where i enumerates all possible outcomes in the sample space.

Definition 9 ([Grimmett and Stirzaker, 2001, p. 93])**.** For a continuous random variable X, the *expectation* is given by

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx,$$

whenever this integral exists.

An important property of the expectation is the so-called *law of the unconscious statistician*:

Theorem 10 ([Grimmett and Stirzaker, 2001, p. 51, 93])**.** If X has density function f and $g : \mathbb{R} \to \mathbb{R}$, then in the discrete case

$$\mathbb{E}[g(X)] = \sum_x g(x) f(x),$$

and in the continuous case

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

Theorem 10 is used in the calculation of another concept related to the expectation, the *variance.*

Definition 11 ([Grimmett and Stirzaker, 2001, p. 51])**.** The *variance* of a random variable X is given by

$$\sigma^2 = \mathrm{Var}(X) = \mathbb{E}[X^2] - E[X]^2.$$

The positive square root $\sigma = \sqrt{\mathrm{Var}(X)}$ is called the *standard deviation.*

The *variance* measures the amount by which X tends to deviate from the *expected value.*

Distributions

In this report we will deal with a few distributions. We will shortly discuss their most important properties here.

Definition 12 ([Grimmett and Stirzaker, 2001, p. 29])**.** The distribution function of a *Bernoulli distributed* random variable X with parameter p is given by

$$F(x) = \begin{cases} 0 & x < 0, \\ 1 - p & 0 \leq x < 1, \\ 1 & x \geq 1. \end{cases} \tag{1}$$

Consequently $\mathbb{E}[X] = p$ and $\mathrm{Var}(X) = p(1-p)$. We denote a random variable following this distribution by $X \sim \mathrm{Ber}(p)$.

Definition 13 ([Grimmett and Stirzaker, 2001, p. 95])**.** A *normal distributed* random variable X with parameters μ and σ^2 has probability density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty. \tag{2}$$

Consequently $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$. We denote a random variable following this distribution by $X \sim \mathcal{N}(\mu, \sigma^2)$. If $\mu = 0$ and $\sigma^2 = 1$ then we speak of the *standard normal distribution.*

Definition 14 ([Tijms, 2012, p. 332])**.** A *Pareto distributed* random variable X with parameters $\alpha > 0$ and $x_m > 0$ has probability density function

$$f(x) = \begin{cases} \dfrac{\alpha x_m^\alpha}{x^{\alpha+1}} & \text{for } x \geq x_m, \\ 0 & \text{for } x < x_m. \end{cases} \tag{3}$$

Consequently $\mathbb{E}[X] = \alpha x_m/(\alpha - 1)$ for $\alpha > 1$ and infinity otherwise. Furthermore, we have that $\mathrm{Var}(X) = \alpha x_m^2/((\alpha-1)^2(\alpha-2))$ for $\alpha > 2$ and infinity otherwise.

The parameter α is sometimes called the *shape* parameter [Donice McCune and Luna McCune, 2000] of the Pareto distribution and x_m is sometimes referred to as the *threshold* [Donice McCune and Luna McCune, 2000] or *scale* [Fackler, 2013] parameter of the Pareto distribution.

Analysis

The field of analysis is a part of mathematics that concerns itself with the behaviour of sequences and series, among other subjects.

Definition 15 ([Kosmala, 2004, p. 66])**. (Convergence of a sequence)** A sequence of real numbers a_n converges to $a \in \mathbb{R}$ if and only if for all $\epsilon > 0$ there exists a $n_0 > 0$ such that for all $n \geq n_0$, $|a_n - a| < \epsilon$.

We say that a_n converges to a and that a is the limit of a_n when n tends to infinity. Conversely when such a value of a does not exist, in other words when $\lim_{n\to\infty} a_n = \pm\infty$ we say that a_n diverges.

Definition 16 ([Kosmala, 2004, p. 295])**. (Convergence of a series)** Suppose that a_n is a sequence of real numbers and let p be an arbitrary integer. We say that the series $\sum_{k=p}^{\infty} a_k$ converges to some real value S if and only if the sequence of partial sums $S_n = \sum_{k=p}^{n} a_k$ converges to S. If $\sum_{k=p}^{\infty} a_k$ converges to S, then S is called the sum of the series and we write $\sum_{k=p}^{\infty} a_k = S$.

A third notion of convergence we will encounter is *pointwise convergence*. Pointwise convergence tells us something about the convergence of a sequence of functions on a certain interval of the function.

Definition 17 ([Grimmett and Stirzaker, 2001, p. 306])**.** If for all $x \in [a, b]$, the sequence $f_n(x)$ of real numbers satisfies $f_n(x) \to f(x)$ as $n \to \infty$, then we say that $f_n \to f$ pointwise.

In order to determine whether a series diverges we have several tests at our disposal. In Theorem 18 and Theorem 19 we suppose that a_n is a sequence and when we say that $\sum a_k$ converges, we mean that $\lim_{n\to\infty} \sum_{k=p}^{n} a_k = a$ for some $a \in \mathbb{R}$.

Theorem 18 ([Kosmala, 2004, p. 296])**.** If $\sum a_k$ converges, then $\lim_{n\to\infty} a_n = 0$.

Proof. Let $S_n = \sum_{i=0}^{n} a_i$. Since S_n converges we say $\lim_{n\to\infty} S_n = S$ for some $S \in \mathbb{R}$. Furthermore $\lim_{n\to\infty} S_{n-1} = S$ as well. We know that for $n > 0$, we have $a_n = S_n - S_{n-1}$ and since both limits are finite, we can conclude that

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} (S_n - S_{n-1}) = \lim_{n\to\infty} S_n - \lim_{n\to\infty} S_{n-1} = 0. \qquad \square$$

The opposite of this theorem is the first way to find out if a series diverges.

Theorem 19 ([Kosmala, 2004, p. 296])**.** If a sequence a_n does not converge to zero, then the series $\sum a_k$ diverges.

1 Introduction

In this chapter, we will give the motivation behind this report. We first dive into the historical background of divergent series and probability theory.

1.1 Historical background

Mathematicians such as Euler, Newton and Leibniz were already familiar with the convergence of infinite series during the 18th century [Kline, 1983]. Mathematicians like Euler were curious to learn what series like $1 - 1 + 1 - 1 + ..$ converge to. Essentially the question they posed was "what is $1 - 1 + 1 - 1 + ..?$" [Ferraro, 1999]. It turned out that there was not one answer to this question. Already in 1703 [Bagni et al., 2005], Guido Grandi worked on solving this problem and he had the following approach:

$$(1-1) + (1-1) + (1-1) + (1-1) + \cdots = 0 + 0 + 0 + 0 + \cdots = 0. \tag{4}$$

However Grandi could also place the brackets around other terms to conclude that

$$1 + (-1+1) + (-1+1) + (-1+1) + \cdots = 1 + 0 + 0 + 0 + \cdots = 1. \tag{5}$$

Hence, Grandi found that the series could possibly be 1 or 0. Therefore he considered the series to be the average of equation (4) and (5), so $\frac{1}{2}$. Other mathematicians, like Euler, worked further on solving this and related problems. One method Euler used was to say that

$$s = 1 - 1 + 1 - 1 + \cdots = 1 - (1 + 1 - 1 + \cdots) = 1 - s, \tag{6}$$

and solving (6) for s gives $s = \frac{1}{2}$ [Hardy, 1949]. A third method that gave the same answer was to use the series

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots, \tag{7}$$

which for $x = 1$ gives

$$\frac{1}{2} = 1 - 1 + 1 - 1 + \cdots. \tag{8}$$

Mathematicians used these methods also for other series, and much in the same way as in (6), this provided values for series such as $s = 1 - 2 + 3 - 4 + \cdots$:

$$s = 1 - 2 + 3 - \cdots = 1 + (-2 + 3 - 4 - \cdots) = 1 - (2 - 3 + 4 + \cdots) \tag{9}$$

$$= 1 - (1 - 1 + 1 - 1 + \cdots) - (1 - 2 + 3 - \cdots) = 1 - \frac{1}{2} - s, \tag{10}$$

so that we may conclude that $s = \frac{1}{4}$ [Hardy, 1949]. Euler worked on with the power series and found that for $x = 2$ he could use the power series

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots, \tag{11}$$

having him conclude that $1 + 2 + 4 + 8 + \cdots = -1$ [Kline, 1983]. Nowadays we learn only to use (7) for $|x| < 1$ [Kosmala, 2004], and back then some mathematicians also questioned it.

Leibniz felt that these methods were "absurd" [Bagni et al., 2005] and started working on his solutions. He stated that the series $1 - 1 + 1 - 1 + \cdots$ was almost half of the time 1 and the other half 0, so it was logical to say it was $\frac{1}{2}$, but not mathematical [Bagni et al., 2005]. Thus Leibniz and others disagreed with Grandi's approach and changed the question from "what is $1 - 1 + 1 - 1 + ..?$" into "how shall we define $1 - 1 + 1 - 1 + ..?$" [Hardy, 1949]. This made that mathematicians in the 19th century worked on general procedures that gave unambiguous solutions to these series. Hardy bundled some of these techniques in his book *Divergent Series* [Hardy, 1949], which we will study in Chapter 3 of this report.

The foundations of probability theory as we know it now was mostly laid after the middle ages, right about at the same time that mathematicians started studying divergent series. The notions of probability and chance were made already by ancient Greeks philosophers, like Plato and Aristotle [Debnath and Basu, 2015]. It took until the end of the middle ages, around the 16th century [Debnath and Basu, 2015], that mathematicians began to investigate random events. In Italy, mathematicians including Galileo and Cardano started investigating dice games [Debnath and Basu, 2015]. Cardano was the first to introduce the idea of a probability p between 0 and 1. Furthermore, Cardano developed the first ideas that ultimately became the law of large numbers [Debnath and Basu, 2015]. Cardano stated that after n trials of an experiment with probability p, the number of times it will occur is close to np [Debnath and Basu, 2015].

After the death of Galileo, Pascal and Fermat started working on his unsolved problems concerning chance games. Pascal's investigation on dice throwing games led to *Pascals triangle* [Debnath and Basu, 2015], which was already discovered in 1303 by the Chinese Chu Shih-chieh [David, 1962].

Based on the work of Pascal, Fermat and others, Huygens wrote his book *On Reasoning in Games of Dice* in 1657 [Tijms, 2012]. In this book, Huygens was the first to create the notion of the expectation of an experiment. Huygens' book was the most important in probability theory for the next decades until Bernoulli's *The Art of Prediction* was published in 1713 [Debnath and Basu, 2015]. Bernoulli formulated the *law of large numbers* in his book, which was the base for the further development of probability theory. Furthermore, Bernoulli solved some of the problems Huygens did not solve, and Bernoulli formulated the probability of r successes in n trials of an experiment with probability p in the following form:

$$B(n, r, p) = \binom{n}{r} p^r (1-p)^{n-r}. \tag{12}$$

We now know (12) as the probability mass function of the *binomial distribution* [Grimmett and Stirzaker, 2001].

Around the same time De Moivre derived the distribution function of the *normal distribution* and De Moivre noted that various natural phenomena follow a normal distribution [Debnath and Basu, 2015]. Based on these observations he thought the first version of the *central limit theorem*, which did not get much attention until Laplace expanded it in 1812 [Debnath and Basu, 2015]. Gauss continued De Moivre's work on the normal distribution, and its shape would later become known as the bell curve [Debnath and Basu, 2015], which was later moreover renamed after him [Tijms, 2012]. The central limit theorem did initially not get much attention, and it took until the 20th century that Lyapunov defined it in general terms and managed to prove it [Tijms, 2012]. The work of Lyapunov has now made the central limit theorem one of the most important theorems in probability theory [Tijms, 2012].

In the 19th century, mathematics as a whole was further developed and this also led to improvements in probability theory. New distributions, such as the Poisson distribution, were derived and much work was done on logic by Venn and De Morgan [Debnath and Basu, 2015]. The *Venn diagrams* were used to describe the relations between subsets. The *De Morgan's laws* were used by Bayes to formulate the law of conditional probability of an event A, given that B has occurred [Debnath and Basu, 2015]:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{when} \quad \mathbb{P}(B) \neq 0. \tag{13}$$

The Russian mathematician Chebyshev contributed to probability theory by defining the concepts of mathematical expectation, variance and the arithmetic mean of random variables. He furthermore managed to prove the law of large numbers using a theorem that is now known as Chebyshev's inequality [Debnath and Basu, 2015].

Russian successors of Chebyshev, such as Kolmogorov and Markov, contributed further to the probability theory as we know it now. Markov did further research on sequences of dependent random variables, now known as Markov chains. Kolmogorov noted, in the 1930s that there was a lack of logical foundation on which probability theory and statistics was being built on for the last 300 years [Debnath and Basu, 2015]. Kolmogorov stated that probability theory should be developed from axioms in the same way as geometry and algebra [Kolmogorov and Bharucha-Reid, 2018]. In Kolmogorov's book *Foundations of Probability Theory* [Kolmogorov and Bharucha-Reid, 2018] he described the notions including *probability measure* and *probability space*, using measure theory as developed by Lebesgue [Debnath and Basu, 2015]. The foundations made by Kolmogorov are still the ones we use today [Grimmett and Stirzaker, 2001].

1.2 Motivation

In analysis, much research has been done on the convergence of series. In probability theory, also many convergence theorems are known, with the best known and most used being the *central limit theorem* and the *law of the large numbers*. However, it seems little is known about divergent series in probability theory, and now the question raises where divergence would appear in probability theory and how it could be applied.

Consider a random variable such as

$$X_i = \begin{cases} 2 \cdot (-1)^{i+1} & \text{with probability } \frac{1}{2}, \\ 0 & \text{with probability } \frac{1}{2}. \end{cases} \tag{14}$$

Clearly $\mathbb{E}[X_i] = \frac{1}{2} \cdot 2 \cdot (-1)^{i+1} = (-1)^{i+1}$, so the sum of these random variables, $S_n = \sum_{i=1}^n X_i$, has the property that $\mathbb{E}[S_n] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n (-1)^{i+1}$, which is a divergent series. We saw that we can assign values to divergent series and for $1 - 1 + 1 - 1 + \cdots$, we can assign $\frac{1}{2}$ just as Grandi did. We are specifically interested if a *law of the large numbers* can be applied to this random variable or whether this causes problems. The solution for $1 - 1 + 1 - 1 + \cdots$ given in (6) was of course not rigorous, and we will use methods as described in the *Divergent Series* book of Hardy [Hardy, 1949]. By analysing such examples we will try to find out where divergent series appear in probability theory and whether we can use summation methods as a solution.

1.3 Goals

We have two main research questions we focus on in this report:

- Is there a notion of divergent series in probability theory?
- Can we find mathematical applications of the summation techniques for divergent series in probability theory?

To answer these questions, we will set out to do a literature study on the most important convergence theorems in probability theory, including the law of large numbers and the central limit theorem. We will then locate the most important techniques for series from Hardy's book [Hardy, 1949].

1.4 Outline of the report

In Chapter 2 we describe and demonstrate the most important theorems and definitions of convergence in probability theory. In Chapter 3 we demonstrate a few summation techniques for divergent series. In Chapter 4 we will discuss the behavior of the random variable stated in (14) and a similar example. Furthermore, we report on a literature survey on the application of the summation techniques in probability theory. Finally, in Chapter 5 we will attempt to reformulate the law of large numbers in terms of a summation technique and we compare it to an existing reformulation.

2 Convergence in probability theory

In this chapter, we discuss the most well-known theorems and definitions of convergence in probability theory. Furthermore, we will illustrate some of them using examples.

2.1 Modes of convergence

The first mode of convergence we consider is called *almost sure convergence* and implies that the set of all events for which $X_n \to X$ pointwise has probability measure 1. In other words, any event in which X_n does not converge pointwise occurs with probability 0.

Definition 20 ([Grimmett and Stirzaker, 2001, p. 308])**.** Let $X, X_1, X_2, \cdots$ be random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $X_n \to X$ almost surely (a.s.) if $\mathbb{P}(\omega \in \Omega : X_n(\omega) \to X(\omega)$ as n $\to \infty) = 1$. We denote this as $X_n \xrightarrow{a.s.} X$.

The second mode of convergence in probability theory we consider is *convergence in probability*. This mode of convergence tells that X_n converges to X, when the probability that some notion of distance between X_n and X is greater than a given $\epsilon > 0$ converges to zero as $n \to \infty$.

Definition 21 ([Grimmett and Stirzaker, 2001, p. 308])**.** Let $X, X_1, X_2, \cdots$ be random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $X_n \to X$ in probability, if for all $\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$. We denote this as $X_n \xrightarrow{P} X$.

The last mode we mention here is *weak convergence*, which is also called *convergence in distribution* or *convergence in law*. This mode of convergence tells us that X_n converges to X in distribution if and only if the cumulative distribution function $F_n(x)$ converges to $F(x)$.

Definition 22 ([Koralov and Sinai, 2007, p. 111])**.** Let $X, X_1, X_2, \cdots$ be random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $X_n \to X$ in distribution if and only if $\lim_{n\to\infty} F_n(x) = F(x)$ for all points x at which the function $F = \mathbb{P}(X \le x)$ is continuous. We denote this as $X_n \xrightarrow{D} X$.

It turns out that there is a relation between these three modes of convergence. Almost sure convergence implies convergence in probability and convergence in probability implies convergence in distribution [Grimmett and Stirzaker, 2001, p. 310]. The converse of this statement does not hold in general, as we will see in Example 1.

Example 1. Let $X, X_1, X_2, \cdots$ be identically distributed random variables taking values 0 or 1 each with probability $\frac{1}{2}$. Clearly X_n converges to X in distribution. Defining $Y = 1 - X$, we see that Y has the same distribution as X, so X_n converges to Y in distribution as well. However, $|X_n - Y| = 1$, so X_n does not converge to Y in probability and also not almost surely.

In some cases, convergence in distribution implies convergence in probability as we see in Theorem 23.

Theorem 23 ([Grimmett and Stirzaker, 2001, p. 310])**.** *Let X_n be a sequence of random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $X_n \xrightarrow{D} c$, where $c \in \mathbb{R}$ is a constant, then $X_n \xrightarrow{P} c$.*

In Theorem 24 we see some consequences of convergence in distribution to a constant.

Theorem 24 ([Vaart, 1998, p. 11])**.** *Let X_n, Y_n be sequences of random variables and let X be a random variable. If $X_n \xrightarrow{D} X$ and $Y \xrightarrow{D} c$ for a constant $c \in \mathbb{R}$, then*

- $X_n + Y_n \xrightarrow{D} X + c$;
- $X_n Y_n \xrightarrow{D} cX$;
- $\dfrac{X_n}{Y_n} \xrightarrow{D} \dfrac{X}{c}$ *provided that* $c \neq 0$.

2.2 Law of large numbers

One of the most important convergence theorems in probability theory is the *law of large numbers.* The law of large numbers tells us that when we sample many, say n, random variables X_i with the same distribution, the expected value of $\sum_{i=1}^n X_i$ is close to n times the mean of the random variable. For instance, when we toss 1000 identical fair coins, the law of large numbers tells us that we can expect around 500 of them to show heads and 500 tails. This fact makes the law of large numbers practically useful: when we have for instance an unfair coin, then we can use the law of large numbers to find out what the probability is for the coin to show heads. This is exactly what is done in stochastic simulation, where we perform a large number of experiments so we can find an approximation for the mean of an unknown random variable.

There are two versions of the law of large numbers: the *weak* and *strong law.* The weak law only preserves weak convergence and the strong law implies almost sure convergence to the mean.

Theorem 25 ([Grimmett and Stirzaker, 2001, p. 193])**. (Weak law of large numbers)** Let $X_1, X_2, X_3, \cdots$ be a sequence of independent identically distributed random variables with finite mean μ. The partial sum $S_n = \sum_{i=1}^n X_i$ satisfies

$$\frac{1}{n} S_n \xrightarrow{D} \mu \text{ as } n \to \infty.$$

Note that by Theorem 23 the weak law of large numbers also implies $S_n/n \xrightarrow{P} \mu$.

Theorem 26 ([Grimmett and Stirzaker, 2001, p. 329])**. (Strong law of large numbers)** Let $X_1, X_2, X_3, \cdots$ be a sequence of independent identically distributed random variables with $\mathbb{E}[X_i] = \mu$. The partial sum $S_n = \sum_{i=1}^n X_i$ satisfies:

$$\frac{1}{n} S_n \xrightarrow{a.s.} \mu \text{ as } n \to \infty$$

for some constant μ if and only if $\mathbb{E}[|X_i|] < \infty$.

To illustrate the difference between the strong and weak law we give an example.

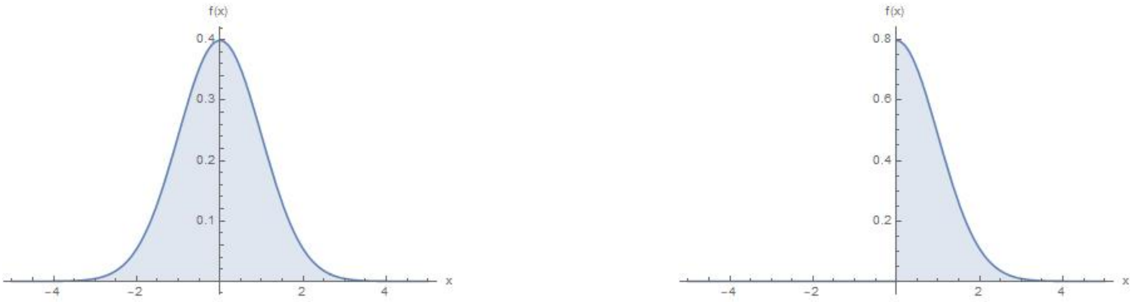
Example 2. Consider the sequence of random variables

$$X_i = \begin{cases} -1 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2}. \end{cases} \tag{15}$$

Clearly $\mathbb{E}[X_i] = -\frac{1}{2} + \frac{1}{2} = 0$, so the weak law tells us that the partial sum $S_n = \sum_{i=1}^n X_i$ satisfies $\frac{1}{n} S_n \xrightarrow{D} 0$ as $n \to \infty$. For the strong law we have that $\mathbb{E}[|X_1|] = \frac{1}{2} + \frac{1}{2} = 1$, so we also have that $\frac{1}{n} S_n \xrightarrow{a.s.} 0$ as $n \to \infty$.

Clearly the strong law of large numbers implies the weak law of large numbers, as almost sure convergence implies convergence in distribution. To find an example for which the weak law of large numbers holds, but the strong law of large numbers does not, we need to consider a distribution for which $\mathbb{E}[|X_i|] \to \infty$.

Example 3. Let $X_1, X_2, \cdots$ be a sequence of normally distributed random variables with mean zero. In Figure 1a we see the probability density function $f(x)$ for variance $\sigma^2 = 1$, also know as the *standard normal distribution.* Clearly, for an arbitrary variance $\mathbb{E}[X_i] = 0$, so according to the weak law of large numbers, the partial sum $S_n = \sum_{i=1}^n X_i$ satisfies $\frac{1}{n} S_n \xrightarrow{D} 0$ as $n \to \infty$. For the strong law of large numbers we examine $\mathbb{E}[|X_i|]$ and we see the probability density function of $|X_i|$ in Figure 1b. This distribution is also known as the *half normal distribution*, which has mean $\sigma\sqrt{\frac{\pi}{2}}$ [Leone et al., 1961]. Since we have that $\mathbb{E}[|X_i|] = \sigma\sqrt{\frac{\pi}{2}}$, we see that, for finite standard deviation σ, the partial sum $S_n = \sum_{i=1}^n X_i$ satisfies $\frac{1}{n} S_n \xrightarrow{a.s.} 0$ as $n \to \infty$. However when we let X_i have an infinite variance, $\mathbb{E}[|X_i|] \to \infty$, only the weak law of large numbers holds.



(a) (b)

Figure 1: Probability density functions for X_i (left) and $|X_i|$ (right).

2.2.1 Proof of the weak law of large numbers

One way of proving the weak law of large numbers is to use *Chebychev's inequality*. This inequality gives an upper bound for the probability that a random variable has a distance of a to its mean.

Theorem 27 ([Hofstad, 2018, p. 3])**. (Chebychev's inequality)** Let X be a random variable with $\mathbb{E}[X] = \mu$, $\text{Var}(X) < \infty$ and let $a > 0$. Then $\mathbb{P}(|X - \mu| \leq a) \leq \dfrac{\text{Var}(X)}{a^2}$.

We now prove Theorem 25.

Proof. Let $X_1, X_2, \cdots$ be independent identically distributed random variables with mean μ and let $\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$. Define $F_{\bar{X}_n}(x) = \mathbb{P}(\bar{X}_n \leq x)$ and $F_\mu(x) = \mathbb{1}_{\mu \leq x}$.

According to Theorem 22 we need to show that for $x < \mu$, $F_{\bar{X}_n}(x) \to F_\mu(x) = 0$, and for $x > \mu$ $F_{\bar{X}_n}(x) \to F_\mu(x) = 1$. Because $x = \mu$ is a discontinuity point, we do not need to prove convergence for it.

Note that for $x < \mu$,

$$F_{\bar{X}_n}(x) - F_\mu(x) = F_{\bar{X}_n}(x) \leq \mathbb{P}(|\bar{X}_n - \mu| > |x - \mu|), \tag{16}$$

and for $x > \mu$,

$$F_\mu(x) - F_{\bar{X}_n}(x) = 1 - F_{\bar{X}_n}(x) = \mathbb{P}(\bar{X}_n > x) \leq \mathbb{P}(|\bar{X}_n - \mu| > |x - \mu|). \tag{17}$$

Consequently

$$|F_{\bar{X}_n}(x) - F_\mu(x)| \leq \mathbb{P}(|\bar{X}_n - \mu| > |x - \mu|). \tag{18}$$

It is enough to show that for any $0 < a < \infty$, $\mathbb{P}(|\bar{X}_n - \mu| > a) \to 0$ as $n \to \infty$. For this we use Theorem 27:

$$\mathbb{P}(|\bar{X}_n - \mu| > a) \stackrel{i.i.d.}{=} \mathbb{P}(|\bar{X}_n - \mathbb{E}[\bar{X}_n]| > a) \leq \frac{\text{Var}(\bar{X}_n)}{a^2}. \tag{19}$$

Since $\mathbb{E}[\bar{X}_n] = \mu$ and $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$, we have

$$\mathbb{P}(|\bar{X}_n - \mu| > a) \leq \frac{\text{Var}(\bar{X}_n)}{a^2} = \frac{\sigma^2}{a^2 n} \to 0, \tag{20}$$

and this proves the statement. $\square$

2.3 Central limit theorem

The next important convergence theorem in probability theory that we consider is the *central limit theorem*. When we look at a sequence of independent identically distributed random variables $X_1, X_2, \cdots$ with mean μ, we find that for the partial sum $S_n = \sum_{i=1}^n X_i$, $\mathbb{E}[S_n] = n\mathbb{E}[X_1] = n\mu$. Furthermore we see that $\mathrm{Var}(S_n) = n\mathrm{Var}(X_1) = n\sigma^2$. Note that S_n therefore has standard deviation $\sqrt{n\sigma^2}$ while its mean equals $n\mu$. Therefore we can conclude that S_n is centered more around the mean for large n. The central limit theorem makes use of this notion, by looking at the scaled random variable $Z_n = (S_n - n\mu)/(\sqrt{n\sigma^2})$. We can directly see that $E[Z_n] = 0$, and that $\mathrm{Var}(Z_n) = 1$. The central limit theorem now states that Z_n converges to a standard normal distribution.

Theorem 28 ([Grimmett and Stirzaker, 2001, p. 194])**. (Central limit theorem)** Let $X_1, X_2, X_3, \cdots$ be a sequence of independent identically distributed random variables with finite mean μ and finite non-zero variance σ^2 and let $S_n = \sum_{i=1}^n X_i$. Then

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{D} X \text{ as } n \to \infty,$$

where X is a standard normal distributed random variable: $X \sim \mathcal{N}(0, 1)$.

2.4 Law of the iterated logarithm

A lesser known convergence theorem in probability theorem is the *law of the iterated logarithm*. This law tells us something about the behaviour of $\sum_{i=1}^n X_i$ for a random variable with expected value 0 and variance 1. While the law of large numbers tells us what S_n/n behaves like for large n, this theorem tells us something about the magnitude of fluctuations of S_n.

Theorem 29 ([Grimmett and Stirzaker, 2001, p. 332])**. (Law of iterated logarithm)** If $X_1, X_2, \cdots$ are independent identically distributed random variables with mean 0 and variance 1, then we have for the partial sums $S_n = \sum_{i=1}^n X_i$ that

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log \log n}} = 1 \text{ almost surely.}$$

Example 4. Consider the sequence of random variables $X_1, X_2, \cdots$ with

$$X_i = \begin{cases} -1 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2}. \end{cases} \tag{21}$$

Recall that $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = 1$, and so for $S_n = \sum_{i=1}^n X_i$ we have the almost sure bound $-n \leq S_n \leq n$. These bounds are loose, the law of the iterated logarithm tells us that S_n in the limit is almost surely less or equal to $\sqrt{2n \log \log n}$, which grows way slower than n. This can be seen by plotting both functions in Figure 2.

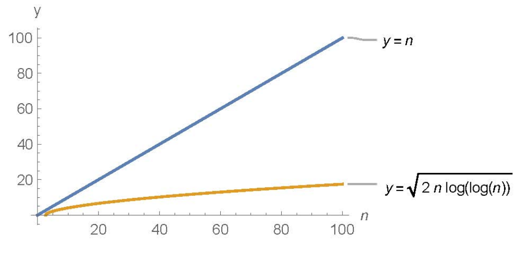


Figure 2: Plot of $\sqrt{2n \log \log n}$ against n.

3 Summation techniques for divergent series

We will now summarize and demonstrate some of the summation for techniques divergent series discussed in [Hardy, 1949]. We discuss 4 *transformation* methods, that are closely related to each other and two of them we will see later in this report. Furthermore, we will discuss the *moment method* as it seems like we can link this summation technique to a concept from probability theory, *moments* of random variables.

Summation techniques assign values to series that diverge according to the regular criteria described in the introduction. These series are series such as $s = 1-1+1-1+\cdots$, $s = 1-2+3-4+\cdots$ and $s = 1+2+4+\cdots$. The summation techniques assign values to s according to their definition. In these definitions some new notation will be introduced. If we define the sum of a sequence $a_0, a_1, a_2, \cdots$ in some new sense, say the A sense, as s, we write that $\sum_{i=0}^{\infty} a_i$ is *summable* (A), we call s the A *sum* of $\sum_{i=0}^{\infty} a_i$, and we write

$$\sum_{i=0}^{\infty} a_i = s \text{ (A)}.$$

We also say that s is the A *limit* of the partial sum $s_n = \sum_{i=0}^{n} a_i$ and write

$$s_n \to s \text{ (A)}.$$

as $n \to \infty$. An important property of summation techniques is the *regularity* of the method.

Definition 30 ([Hardy, 1949, p. 10])**.** Suppose that $s_n = \sum_{i=0}^{n} a_i$ is a convergent series, with $s_n \to s$ and let A be an summation technique. If

$$s_n \to s \text{ (A)}.$$

then we say that a summation technique is *regular*.

3.1 Transformation methods

The first summation techniques we will discuss make make use of *transformations*.

Definition 31 ([Hardy, 1949, p. 43])**.** Let $a_0, a_1, \cdots$ be a sequence and define $s_n = \sum_{i=0}^{n} a_i$. A transformation t_m of the form

$$t_m = \sum_{n=0}^{m} c_{m,n} s_n$$

for $m = 0, 1, 2, 3, \cdots$, is said to be regular if

$$t_m \to s \text{ as } m \to \infty$$

whenever

$$s_n \to s \text{ as } n \to \infty.$$

3.1.1 Nörlund method

Definition 32 ([Hardy, 1949, p. 64])**.** (**Nörlund means method**) Let $a_0, a_1, a_2, \cdots$ be a sequence and define $s_n = \sum_{i=0}^{n} a_i$. Now suppose that $p_n \geq 0$, $p_0 > 0$ and $P_n = \sum_{i=0}^{n} p_i$. Define t_m by

$$t_m = N_m^{(p)}(s) = \frac{\sum_{i=0}^{m} p_{m-i} s_i}{P_m}.$$

If $t_m \to s$ as $m \to \infty$, we write

$$s_n \to s, \ \sum_{n=0}^{\infty} a_i = s \text{ (N,}p_n\text{)}.$$

Theorem 33 ([Hardy, 1949, p. 64])**.** The (N,p_n) method is regular if and only if $p_n/P_n \to 0$ as $n \to \infty$.

Example 5. The simplest example of the Nörlund method we can show is when we take $p_n = 0$ for $n \geq 1$ and $p_0 = 1$ such that $P_m = 1$. For an arbitrary convergent series s_n, we get

$$t_m = N_m^{(p)}(s) = s_m \to s \text{ as } m \to \infty. \tag{22}$$

This particular Nörlund method is regular, since $p_n/P_n \to 0$ as $n \to \infty$.

Example 6. Take the sequence $p_n = 1$ for all n and observe that $P_n = n + 1$. According to Theorem 33 the method is regular as $\frac{p_n}{P_n} = \frac{1}{n+1} \xrightarrow{n\to\infty} 0$. This implies the transformation

$$t_m = \frac{s_0 + s_1 + \cdots + s_n}{n+1}. \tag{23}$$

Now consider the series $1 - 1 + 1 - 1 + \cdots$, and let $s_n = \sum_{i=0}^{n}(-1)^n$. Note that $s_n = 0$ when n is odd and $s_n = 1$ when n is even. Using this particular Nörlund method we obtain

$$\frac{1 + 0 + 1 + 0 + \cdots}{n+1} \xrightarrow{n\to\infty} \frac{1}{2}, \tag{24}$$

and therefore $s_n \to \frac{1}{2}$ (N, p_n).

Example 7. Consider the sequence $p_n = (\frac{1}{2})^n$ and observe that then $P_n = 2 - (\frac{1}{2})^n$. By Theorem 32

$$t_m = N_m^{(p)}(s) = \frac{\sum_{i=0}^{m}(\frac{1}{2})^{m-i} s_i}{2 - (\frac{1}{2})^m}. \tag{25}$$

Clearly

$$\frac{p_n}{P_n} = \frac{(\frac{1}{2})^n}{2 - (\frac{1}{2})^n} = \frac{1}{2^{n+1} - 1} \xrightarrow{n\to\infty} 0, \tag{26}$$

so the according to Theorem 33 this method is regular. We consider the series $\sum(-1)^n = 1 - 1 + 1 - 1 + \cdots$ again, and substituting this series in equation (25), we obtain

$$t_m = \begin{cases} \dfrac{\sum_{i=0}^{m/2}(\frac{1}{2})^{2i}}{2 - (\frac{1}{2})^m} = \dfrac{\frac{4}{3} - \frac{1}{3}(\frac{1}{2})^m}{2 - (\frac{1}{2})^m} & \text{if } m \text{ even,} \\ \dfrac{\sum_{i=0}^{(m-1)/2}(\frac{1}{2})^{2i+1}}{2 - (\frac{1}{2})^m} = \dfrac{\frac{2}{3} - \frac{1}{3}(\frac{1}{2})^m}{2 - (\frac{1}{2})^m} & \text{if } m \text{ odd.} \end{cases} \tag{27}$$

We conclude that this Nörlund method is inconclusive for this series, as $t_m \to \frac{1}{3}$, when we assume that m is even, however $t_m \to \frac{2}{3}$, when m is odd.

3.1.2 Hölder's means method

The next summation method we will discuss is *Hölder's means method.* Note that in this and following definitions the indices H_n^k should not be read as powers.

Definition 34 ([Hardy, 1949, p. 94])**. (Hölder's means)** Let $a_0, a_1, a_2, \cdots$ be an arbitrary sequence and $s_n = \sum_{i=0}^{n} a_i$. Define H_n^k, for $k = 0, 1, 2, \cdots$, by $H_n^0 = s_n$, and let

$$H_n^k = \frac{H_0^{k-1} + H_1^{k-1} + \cdots + H_n^{k-1}}{n+1}.$$

If $H_n^k \to s$ as $n \to \infty$, we say that

$$s_n \to s, \ \sum_{n=0}^{\infty} a_i = s \text{ (H,}k\text{)}.$$

Theorem 35 ([Hardy, 1949, p. 95])**.** If $k' > k > -1$ and $\sum a_n = A$ (H,k) then $\sum a_n = A$ (H,k').

Note that the Nörlund method described in Example 6 is a (H,1)-summation. Therefore we can also say that $\sum_{i=0}^{\infty}(-1)^i = \frac{1}{2}$ (H,1).

Example 8. Consider the series $1 - 2 + 3 - 4 + ..$ and we say $s_n = \sum_{i=0}^{n}(-1)^i(i+1)$. Clearly $s_0 = 1$, $s_1 = -1$, $s_2 = 2$ etc. Computing H_n^1 gives

$$H_n^1 = \frac{s_0 + s_1 + \cdots + s_n}{n+1} = \begin{cases} \dfrac{s_n}{n+1} = \dfrac{\frac{1}{2}(n+2)}{n+1} & \text{if } n \text{ even, and } n = 0, \\ 0 & \text{if } n \text{ odd.} \end{cases} \tag{28}$$

We see that s_n is not (H,1)-summable, so we will go to the next step and compute H_n^2:

$$H_n^2 = \frac{H_0^1 + H_1^1 + H_2^1 + \cdots H_n^1}{n+1} = \frac{1 + \frac{2}{3} + \cdots + \dfrac{\frac{1}{2}(n+2)}{n+1}}{n+1}. \tag{29}$$

Observe that the limit of H_n^2 is a bit complicated, however in Example 9 we will see in a moment that $s_n \to \frac{1}{4}$ (H,2).

3.1.3 Cesàro method

Definition 36 ([Hardy, 1949, p. 96])**.** **(Cesàro means)** Let $a_0, a_1, \cdots$ be an arbitrary sequence and $s_n = \sum_{i=0}^{n} a_i$. Define A_n^k for $k = 0, 1, 2, \cdots$, by $A_n^0 = s_n$, and let

$$A_n^k = A_0^{k-1} + A_1^{k-1} + \cdots + A_n^{k-1}.$$

Let E_n^k be the value of A_n^k for the sequence $a_0 = 1$ and $a_n = 0$ for $n > 0$. In particular $E_n^0 = 1$ for all n. If

$$C_n^k = \frac{A_n^k}{E_n^k} \to s$$

as $n \to \infty$, then we say that

$$s_n \to s, \quad \sum_{n=0}^{\infty} a_i = s \text{ (C, } k\text{)}.$$

Theorem 37 ([Hardy, 1949, p. 101])**.** If $k > 0$, then the (C,k) method is regular.

Theorem 38 ([Hardy, 1949, p. 103])**.** **(Equivalence theorem)** The (C,k) and (H,k) means are equivalent, if $\sum a_n$ is (C,k) summable, then it is summable (H,k) to the same sum and conversely.

Theorem 39 ([Hardy, 1949, p. 100])**.** If $k' > k > -1$ and $\sum a_n = A$ (C,k) then $\sum a_n = A$ (C,k').

Example 9. We saw in Example 8 that the (H,1)-method was inconclusive, so we needed the (H,2)-method, but that limit was hard to compute. Note that the (C,1)-method is the same as the (H,1)-method, so we have

$$C_n^1 = H_n^1 = \frac{s_0 + s_1 + \cdots + s_n}{n+1} = \begin{cases} \dfrac{s_n}{n+1} = \dfrac{\frac{1}{2}(n+2)}{n+1} & \text{if } n \text{ even,} \\ 0 & \text{if } n \text{ odd.} \end{cases} \tag{30}$$

Now we compute E_n^2

$$\begin{aligned} E_n^2 =& E_0^1 + E_1^1 + \cdots + E_n^1 = E_0^0 + E_0^0 + E_0^1 + \cdots + \sum_{i=0}^{n} E_i^0 & (31) \\ =& 1 + 2 + 3 + \cdots + (n+1) = \frac{1}{2}(n+1)(n+2). & (32) \end{aligned}$$

For A_n^1 we get

$$A_n^1 = \begin{cases} \frac{s_n}{n+1} = \frac{1}{2}(n+2) & \text{if } n \text{ even,} \\ 0 & \text{if } n \text{ odd.} \end{cases} \tag{33}$$

Consequently for n even we have

$$A_n^2 = A_0^1 + A_1^1 + \cdots + A_n^1 = 1 + 2 + 3 + \cdots + \frac{1}{2}(n+2) = \frac{1}{8}(n+2)(n+4), \tag{34}$$

and for odd n we have

$$A_n^2 = A_0^1 + A_1^1 + \cdots + A_{n-1}^1 + A_n^1 = 1 + 2 + 3 + \cdots + \frac{1}{2}(n+1) = \frac{1}{8}(n+1)(n+3). \tag{35}$$

Hence we find that

$$C_n^2 = \frac{A_n^2}{E_n^2} = \begin{cases} \frac{\frac{1}{8}(n+2)(n+4)}{\frac{1}{2}(n+1)(n+2)} = \frac{\frac{1}{4}(n+4)}{n+1} & \text{if } n \text{ even,} \\ \frac{\frac{1}{8}(n+1)(n+3)}{\frac{1}{2}(n+1)(n+2)} = \frac{\frac{1}{4}(n+3)}{n+2} & \text{if } n \text{ odd,} \end{cases} \tag{36}$$

and we conclude that

$$C_n^2 \to \frac{1}{4} \tag{37}$$

as $n \to \infty$. We therefore say that $\sum_{n=0}^{n}(-1)^n(i+1) = \frac{1}{4}$ (C,2) and by Theorem 38 we also have that $\sum_{n=0}^{n}(-1)^n(i+1) = \frac{1}{4}$ (H,2).

Calculating A_n^k and E_n^k step-by-step requires k steps. However, it turns out that we can express A_n^k and E_n^k in general terms. For an arbitrary sequence a_n, [Hardy, 1949, p. 96] derives that

$$A_n^k = \sum_{i=0}^{n} \binom{n-i+k}{k} a_i = \sum_{i=0}^{n} \binom{i+k}{k} a_{n-i}. \tag{38}$$

For E_n^k we have that $a_0 = 1$ and for $n \geq 1$ we have $a_n = 0$, so we find the general expression

$$E_n^k = \binom{n+k}{k} = \frac{(k+1)(k+2)\cdots(k+n)}{n!}, \tag{39}$$

which we can rewrite as

$$E_n^k = \binom{n+k}{k} = \frac{(n+1)(n+2)\cdots(n+k)}{k!}. \tag{40}$$

Equation (40) behaves for large n as $\frac{n^k}{k!}$. This leads to the following alternative definition for the Cesàro means method.

Definition 40 ([Hardy, 1949, p. 96])**.** Let $a_0, a_1, \cdots$ be an arbitrary sequence and $s_n = \sum_{i=0}^{n} a_i$. If

$$C_n^k = \frac{k!}{n^k} \sum_{i=0}^{n} \binom{n-i+k}{k} a_i = \frac{k!}{n^k} \sum_{i=0}^{n} \binom{i+k}{k} a_{n-i} \to s$$

when $n \to \infty$, then we say that

$$s_n \to s \text{ (C,}k\text{)}, \qquad \sum_{n=0}^{\infty} a_i = s \text{ (C,}k\text{)}.$$

3.1.4 Riesz method

Riesz worked on improvements for the Cesàro's method and one of them closely resembles Nörlund's method.

Definition 41 ([Kozlov, 2003, p. 1474])**.** Let $a_0, a_1, \cdots$ be an arbitrary sequence and $s_n = \sum_{i=0}^{n} a_i$. Let $p_0 > 0$, $p_n \geq 0$ and $\sum p_n = \infty$. The *Riesz averages* of a sequence are given by

$$t_n = \frac{p_0 s_0 + \cdots + p_n s_n}{p_0 + \cdots + p_n}.$$

If $t_n \to s$ for $n \to \infty$ then we say that

$$s_n \to s \text{ (R,} p_n\text{)}, \quad \sum_{n=0}^{\infty} a_i = s \text{ (R,} p_n\text{)}.$$

We see that this method looks similar to the Norlund method of Definition 32. The main difference is that Definition 41 asks for a sequence p_n that satisfies $\sum p_n = \infty$, such that the method is regular, while Definition 32 does not ask for this property. Observe that when we take the sequence $p_n = 1$ for all n, that then the (R,p_n)-summation is the same as (H,1)- and (C,1)-summation.

3.2 Moment method

We will now discuss one method that does not use a transformation.

Definition 42 ([Hardy, 1949, p. 81])**.** A moment constant μ_n is a number of the form

$$\mu_n = \int_0^\infty x^n f(x) dx, \tag{41}$$

where $F(x) = \int_0^x f(u) du$ is a bounded and increasing function of x such that the integral (41) converges for all n.

Definition 43 ([Hardy, 1949, p. 81, 82])**. (Moment constant methods)** Define

$$a(x) = \sum_{n=1}^{\infty} \frac{a_n}{\mu_n} x^n,$$

where a_n is a sequence. Term-by-term integration yields

$$\int a(x) f(x) dx = \int \sum_{n=0}^{\infty} \frac{a_n}{\mu_n} x^n f(x) dx = \sum_{n=0}^{\infty} \frac{a_n}{\mu_n} \int x^n f(x) dx = \sum_{n=0}^{\infty} a_n,$$

and this suggest that we take the integral on the left as the basis of a definition of the sum of $\sum_{n=0}^{\infty} a_n$. We write

$$\int a(x) f(x) dx = s,$$

and

$$\sum_{n=0}^{\infty} a_n = s \ (\mu_n).$$

Theorem 44 ([Hardy, 1949, p. 82])**.** The (μ_n) method is regular.

This method grabs our attention as we are familiar with moments from probability theory, so we will investigate whether we can use them in a moments constant method. We know the nth moment of a probability density function $f(x)$ equals the expectation of the random variable X^n.

Definition 45 ([Grimmett and Stirzaker, 2001, p. 51])**.** If n is a positive integer, the nth moment μ_n of a random variable X, with probability density function $f(x)$, is defined as

$$\mu_n = \int_{-\infty}^{\infty} x^n f(x)dx = \mathbb{E}[X^n].$$

We first try it with moments from a distribution with relative easy moments.

Example 10. The uniform distribution on $[0, 1]$ has the following probability density function

$$f(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1, \\ 0 & \text{for } x < 0 \text{ and } x > 1. \end{cases} \tag{42}$$

Consequently the moments of this uniform distribution are given by

$$\mu_n = \int_{-\infty}^{\infty} x^n f(x)dx = \int_0^1 x^n dx = \frac{1}{n+1}, \tag{43}$$

and therefore

$$a(x) = \sum_{n=0}^{\infty} \frac{a_n}{\mu_n} x^n = \sum_{n=0}^{\infty} (n+1)a_n x^n. \tag{44}$$

We take the series of the sequence $a_n = \frac{1}{(n+1)^2}$, such that

$$a(x) = \sum_{n=0}^{\infty} \frac{a_n}{\mu_n} x^n = \sum_{n=0}^{\infty} \frac{x^n}{n+1}, \tag{45}$$

which is convergent for $|x| < 1$. Numerical computing suggest that $s = \sum_{n=1}^{\infty} \frac{1}{(n+1)^2}$ converges to $\frac{1}{6}(\pi^2 - 6)$. Now we will compute this using the moment constant method:

$$\int a(x)f(x)dx = \int \sum_{n=1}^{\infty} \frac{n+1}{(n+1)^2} x^n f(x)dx = \int_0^1 \sum \frac{x^n}{n+1}dx = \int_0^1 \frac{1}{x} \sum \frac{x^{n+1}}{n+1}dx \tag{46}$$

$$= \int_0^1 \frac{1}{x} \sum_{n=1}^{\infty} \frac{x^{n+1}}{n+1}dx = \int_0^1 \frac{-\log(1-x) - x}{x}dx = \frac{1}{6}(\pi^2 - 6). \tag{47}$$

We see that this moment method gives the same answer as direct numerically calculation. Although the moments of the uniform distribution are quite simple, integrating them is only possible using complex integral theory, and this is beyond the scope of this project, which is the reason why we did it numerically.

We will consider another example using the Pareto distribution.

Example 11. For a *Pareto distribution* with parameters α and x_m, the probability density function and moments are given by

$$f(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}} & \text{for } x \geq x_m \\ 0 & \text{for } x < x_m, \end{cases}, \text{ and } \mu_n = \begin{cases} \infty & \text{for } \alpha \leq n \\ \frac{\alpha x_m^n}{\alpha - n} & \text{for } \alpha > n, \end{cases} \tag{48}$$

respectively. We see that for this distribution a_n/μ_n tends to zero from the αth term onward. If we now assign values for α and x_m, we can perform the moment method using this distribution. For example if we take $\alpha = 3$ and $x_m = 1$, we obtain

$$a(x) = \sum_{n=0}^{\infty} \frac{a_n}{\mu_n} x^n = a_0 + \frac{a_1 x}{3/2} + \frac{a_2 x}{3/1} = a_0 + \frac{2a_1 x}{3} + \frac{a_2 x^2}{3}. \tag{49}$$

When we execute the moment constant method with these values we obtain

$$\int a(x)f(x)dx = \int_1^\infty \left(a_0 + \frac{2a_1 x}{3} + \frac{a_2 x^2}{3}\right)\frac{3}{x^4}dx = \int_1^\infty \frac{3a_0}{x^4} + \frac{2a_1}{x^3} + \frac{a_2}{x^2}dx \tag{50}$$

$$= \left[-\frac{a_0}{x^3} - \frac{a_1}{x^2} - \frac{a_2}{x}\right]_{x=1}^{x\to\infty} = a_0 + a_1 + a_2. \tag{51}$$

It is clear that this example is not a regular method. We can explain this as definition 42 requires that the integral (41) converges for all n, and by (48) μ_n only converges for $n = 0, 1, 2$, for these values of α and x_m. We also see in general that the moments of the Pareto distribution only converge for $\alpha \leq n$, as the moments are defined by (48) to be ∞ when $\alpha \leq n$. In general we see that the method is not regular, since we have that

$$\int a(x)f(x)dx = \int_{x_m}^\infty \sum_{n=0}^{\alpha-1} \frac{(\alpha-n)a_n x^n}{\alpha x_m^n} \frac{\alpha x_m^\alpha}{x^{\alpha+1}}dx = \int_{x_m}^\infty \sum_{n=0}^{\alpha-1}(\alpha-n)a_n x_m^{\alpha-n} x^{n-\alpha-1}dx \tag{52}$$

$$= \left[-\sum_{n=0}^{\alpha-1} a_n x_m^{\alpha-n} x^{n-\alpha}\right]_{x=x_m}^{x\to\infty} = \sum_{n=0}^{\alpha-1} a_n x_m^{\alpha-n} x_m^{n-\alpha} = \sum_{n=0}^{\alpha-1} a_n. \tag{53}$$

Therefore we must conclude that as μ_n does not converge for all n, that the moments of the Pareto distribution do not fit the requirements of Definition 42 and therefore cannot be used in the moments method.

4 Divergent series in probability theory

In this Chapter, we will discuss two sequences of random variables that have a diverging expectation. We will also describe the findings of a brief literature study on the applications of the summation techniques in probability theory.

4.1 Divergent expectations

4.1.1 Alternating random variable

In Example 6 we saw the alternating series $s_n = 1 - 1 + 1 - 1 + \cdots$ that equals $\frac{1}{2}$ (C,1). We are interested in constructing a random variable that behaves similar to this series. Consider for example

$$X_i = \begin{cases} (-1)^{i+1} \cdot 2 & \text{with probability } \frac{1}{2}, \\ 0 & \text{with probability } \frac{1}{2}, \end{cases} \tag{54}$$

so that $\mathbb{E}[X_i] = (-1)^{i+1}$. We now define $S_n = \sum_{i=1}^{n} X_i$ and we obtain

$$\mathbb{E}[S_n] = \mathbb{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} (-1)^{i+1}. \tag{55}$$

Equation (55) diverges according to normal convergence theorems used in probability theory. However, we can say that

$$\mathbb{E}[S_n] \xrightarrow{n\to\infty} \frac{1}{2} \text{ (C,1)}. \tag{56}$$

We will now prove that according to Theorem 25 (the weak law of large numbers) $\frac{1}{n} S_n \xrightarrow{D} 0$ as $n \to \infty$.

Proof. We split S_n into an odd and an even part. For the odd part we have random variables X_i with mean 1 and for the even part we have mean -1. If we take $E_n = \sum_{i=1}^{\lfloor n/2 \rfloor} X_{2i} = X_2 + X_4 + \cdots$, then we obtain using Theorem 25 that $\frac{2}{n} E_n \xrightarrow{D} -1$ as $n \to \infty$. With $O_n = \sum_{i=0}^{\lfloor (n-1)/2 \rfloor} X_{2i+1} = X_1 + X_3 + \cdots$ we find according to Theorem 25 that $\frac{2}{n} O_n \xrightarrow{D} 1$ as $n \to \infty$. Since $S_n = E_n + O_n$ we find that

$$\frac{1}{n} S_n = \frac{1}{n} (E_n + O_n) = \frac{1}{2} \left(\frac{2}{n} E_n + \frac{2}{n} O_n \right). \tag{57}$$

Now by applying Theorem 24 on (57), we find that

$$\frac{1}{2} \left(\frac{2}{n} E_n + \frac{2}{n} O_n \right) \xrightarrow{D} \frac{1}{2} (-1 + 1) = 0. \tag{58}$$

Hence we conclude that $\frac{1}{n} S_n \xrightarrow{D} 0$. □

It seems like we have found an interesting random variable in S_n as we need a (C,1)-summation to find the expected value. We will try to find out more properties of S_n in the remainder of this section.

Probability mass function for S_n

We will first try to derive the probability mass function for S_n and we will guess a candidate expression by inspection. When we calculate the possible outcomes for S_n, for instance for $n = 6$, we find that $\mathbb{P}(S_6 = -6) = \mathbb{P}(S_6 = 6) = \frac{1}{64}$, $\mathbb{P}(S_6 = -4) = \mathbb{P}(S_6 = 4) = \frac{6}{64}$, $\mathbb{P}(S_6 = -2) = \mathbb{P}(S_6 = 2) = \frac{15}{64}$ and $\mathbb{P}(S_6 = 0) = \frac{20}{64}$ so $\mathbb{E}[S_6] = 0$. Likewise for an odd n, like 7, we get

$\mathbb{P}(S_7 = -6) = \mathbb{P}(S_7 = 8) = \frac{1}{128}$, $\mathbb{P}(S_7 = -4) = \mathbb{P}(S_7 = 6) = \frac{7}{128}$, $\mathbb{P}(S_7 = -2) = \mathbb{P}(S_7 = 4) = \frac{21}{128}$ and $\mathbb{P}(S_7 = 0) = \mathbb{P}(S_7 = 2) = \frac{35}{128}$ such that

$$\mathbb{E}[S_7] = \frac{1}{128}\left(-6 + 8 + 7\cdot(-4+6) + 21\cdot(-2+4) + 35\cdot 2\right) = 1. \tag{59}$$

We see that the problem we have behaves like binomial coefficients for $n = 6$ and $n = 7$ and it turns out that this is in general the case. We therefore can find an expression for the probability that $S_n = m$, by taking the values out of the Pascal's triangle, dividing them by 2^n. We guess the following expression for n even

$$f_{n,m} = \frac{1}{2^n}\binom{n}{(n-m)/2}, \tag{60}$$

and for odd n we find

$$f_{n,m} = \frac{1}{2^n}\binom{n}{(n-m+1)/2}. \tag{61}$$

We will prove that (60) and (61) in fact satisfy $f_{n,m} = \mathbb{P}(S_n = m)$ by induction.

Proof. Base for odd: For $n = 1$ we know that S_1 can take the values $m = 0$ and $m = 2$, both with probability $\frac{1}{2}$. We fill in both values for m and $n = 1$ in (61) and we get

$$f_{1,0} = \frac{1}{2}\binom{1}{(1-0+1)/2} = \frac{1}{2} = \mathbb{P}(S_1 = 0), \quad f_{1,2} = \frac{1}{2}\binom{1}{(1-2+1)/2} = \frac{1}{2} = \mathbb{P}(S_1 = 2). \tag{62}$$

Base for even: For $n = 2$ we know that S_2 can take the values $m = 0$, $m = 2$ and $m = -2$, with probabilities $\frac{1}{2}$, $\frac{1}{4}$ and $\frac{1}{4}$ respectively. We now fill these values for m and $n = 2$ in (60) and we get

$$f_{2,0} = \frac{1}{4}\binom{2}{(2-0)/2} = \frac{1}{2} = \mathbb{P}(S_2 = 0), \quad f_{2,2} = \frac{1}{4}\binom{2}{(2-2)/2} = \frac{1}{4} = \mathbb{P}(S_2 = 2), \tag{63}$$

$$f_{2,-2} = \frac{1}{4}\binom{2}{(2+2)/2} = \frac{1}{4} = \mathbb{P}(S_2 = -2). \tag{64}$$

Induction step: We suppose that (60) and (61) satisfy, for all m in the sample space, $f_{n,m} = \mathbb{P}(S_n = m)$ for $1 \le n \le k$, and we assume that without loss of generality that k is even. Using the combinatorial identity

$$\binom{a}{b} + \binom{a}{b+1} = \binom{a+1}{b+1}, \tag{65}$$

we obtain that

$$\begin{aligned}
\mathbb{P}(S_{k+1} = m) &\overset{(54)}{=} \frac{1}{2}\mathbb{P}(S_k = m) + \frac{1}{2}\mathbb{P}(S_k = m-2) && (66)\\
&= \frac{1}{2^{k+1}}\binom{k}{(k-m)/2} + \frac{1}{2^{k+1}}\binom{k}{(k-m+2)/2} && (67)\\
&\overset{(65)}{=} \frac{1}{2^{k+1}}\binom{k+1}{(k+1-m+1)/2}, && (68)
\end{aligned}$$

as the claim says. For k odd, we obtain

$$\begin{aligned}
\mathbb{P}(S_{k+1} = m) &\overset{(54)}{=} \frac{1}{2}\mathbb{P}(S_k = m+2) + \frac{1}{2}\mathbb{P}(S_k = m) && (69)\\
&= \frac{1}{2^{k+1}}\binom{k}{(k-m-2+1)/2} + \frac{1}{2^{k+1}}\binom{k}{(k-m+1)/2} && (70)\\
&\overset{(65)}{=} \frac{1}{2^{k+1}}\binom{k+1}{(k+1-m)/2}, && (71)
\end{aligned}$$

which also corresponds to the claim, so by induction $\mathbb{P}(S_n = m)$ is equal to (60) for all even n and to (61) for all odd n. $\square$

Expected value for S_n

Now that we have found a probability mass function for S_n, we can also compute its expectation for all n. We find that for n even

$$\mathbb{E}[S_n] = \sum_{i=-n/2}^{n/2} 2i\frac{1}{2^n}\binom{n}{(n-2i)/2} = \sum_{i=1}^{n/2} 2i\frac{1}{2^n}\binom{n}{(n-2i)/2} - \sum_{i=1}^{n/2} 2i\frac{1}{2^n}\binom{n}{(n-2i)/2} = 0, \quad (72)$$

and for n odd

$$\mathbb{E}[S_n] = \sum_{i=-(n-1)/2}^{(n+1)/2} 2i\frac{1}{2^n}\binom{n}{(n-2i+1)/2} \quad (73)$$

$$= \sum_{i=0}^{(n-1)/2} (2i+2)\frac{1}{2^n}\binom{n}{(n-2i+1)/2} - \sum_{i=0}^{(n-1)/2} 2i\frac{1}{2^n}\binom{n}{(n-2i+1)/2} \quad (74)$$

$$= \sum_{i=0}^{(n-1)/2} 2\frac{1}{2^n}\binom{n}{(n-2i+1)/2} = 1, \quad (75)$$

which is indeed how we constructed S_n.

Variance of S_n

We are interested in the question whether the possible outcomes are centered around 0, so we will try to calculate the variance of S_n. For this we use that the X_i as stated in (54) are independent of each other. As we have that $\mathbb{E}[X_i] = (-1)^{i+1}$ and $\mathbb{E}[X_i^2] = \frac{1}{2}\left(2(-1)^{i+1}\right)^2 = 2$, we obtain that $\text{Var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = 2 - 1 = 1$, both for i odd and i even. Hence we get

$$\text{Var}(S_n) = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) = n. \quad (76)$$

Hence we found that S_n has probability mass functions (61) and (60), (C,1) mean $\frac{1}{2}$, and variance n.

4.1.2 Sequence of random variables each with another distribution

Another interesting series we saw is $1-2+3-4+5-\cdots$. We can construct a sequence of random variables behaving like this series. We take the sequence of random variables $X_1, X_2, \cdots$ with $X_i \sim (-1)^{i+1} \cdot 2i \cdot \text{Ber}(\frac{1}{2})$, in other words

$$\mathbb{P}[X_i = x] = \begin{cases} \frac{1}{2} & \text{for } x = (-1)^{i+1} \cdot 2i, \\ \frac{1}{2} & \text{for } x = 0. \end{cases} \tag{77}$$

Clearly the partial sum $S_n = \sum_{i=1}^{n} X_i$ has the expected value

$$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = 1 - 2 + 3 - 4 + 5 - \cdots + (-1)^{i+1} \cdot n, \tag{78}$$

which is exactly the series we described in Examples 8 and 9, which is (H,2)- and so (C,2)-summable to $\frac{1}{4}$. Hence we can conclude that

$$\mathbb{E}[S_n] \xrightarrow{n\to\infty} \frac{1}{4} \text{ (C,2)}. \tag{79}$$

Now we are of course interested in whether we can find a general probability mass function for S_n. We do not see a pattern in this directly, such as in the previous example, and we can not resort to guessing an expression for the probability mass function. Instead we can use recursions to find the probability density function, for n odd

$$\mathbb{P}(S_n = i) = \frac{1}{2}\mathbb{P}(S_{n-1} = i) + \frac{1}{2}\mathbb{P}(S_{n-1} = i + 2n), \tag{80}$$

and for n even

$$\mathbb{P}(S_n = i) = \frac{1}{2}\mathbb{P}(S_{n-1} = i) + \frac{1}{2}\mathbb{P}(S_{n-1} = i - 2n). \tag{81}$$

We performed the recursion in (80), and (81) for a total of 28 times, and in Table 1 we see the probability that S_n is less or equal to 0, the expectation and the variance of S_n. The values for $\mathbb{P}(S_n) \leq 0$ are of course fractions, but we give them as decimals with at most seven decimal places.

n	$\mathbb{P}(S_n \leq 0)$	$\mathbb{E}(S_n)$	$\text{Var}(S_n)$	n	$\mathbb{P}(S_n \leq 0)$	$\mathbb{E}(S_n)$	$\text{Var}(S_n)$	n	$\mathbb{P}(S_n \leq 0)$	$\mathbb{E}(S_n)$	$\text{Var}(S_n)$
1	0.5	1	1	11	0.4155273	6	506	20	0.5795889	-10	2870
2	0.75	-1	5	12	0.6044922	-6	650	21	0.4324436	11	3311
3	0.375	2	14	13	0.4196777	7	819	22	0.5757031	-11	3795
4	0.6875	-2	30	14	0.5961304	-7	1015	23	0.4348333	12	4324
5	0.40625	3	55	15	0.4234619	8	1240	24	0.5723321	-12	4900
6	0.65625	-3	91	16	0.5895538	-8	1496	25	0.4369939	13	5525
7	0.4062500	4	140	17	0.4268188	9	1785	26	0.5693720	-13	6201
8	0.6289062	-4	204	18	0.5841370	-9	2109	27	0.4389589	14	6930
9	0.4101562	5	285	19	0.4297867	10	2470	28	0.5667461	-14	7714
10	0.6152344	-5	385								

Table 1: Theoretical values S_n.

We observe from Table 1 that it seems like the probability that S_n is less or equal to 0 tends to 0.5. Furthermore it seems like the variance tends to infinity, which is logical as

$$\text{Var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = \sum x^2 \mathbb{P}(X_i = x) - \mathbb{E}[X_i]^2 = \frac{1}{2}(2i)^2 - i^2 = i^2, \tag{82}$$

and therefore

$$\text{Var}(S_n) = \text{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Var}(X_i) = \sum_{i=1}^{n} i^2 = \frac{1}{6}n(n+1)(2n+1). \tag{83}$$

We conclude that S_n has a (C,2) expectation $\frac{1}{4}$ and variance $\frac{1}{6}n(n+1)(2n+1)$.

4.2 Summation techniques applied on probability theory

A lot is known about the applications of the summation techniques for divergent series, see for instance [Mursaleen, 2014]. However [Mursaleen, 2014] lets us think that the applications in probability theory are limited, as other mathematical subjects, like statistics and analysis, have a lot more theorems described in this book. In [Mursaleen, 2014] some theory regarding *Abel summability*, a summation technique we did not mention in Chapter 3, of random variables is given. Furthermore, a theorem regarding the strong law of large numbers is given, which we will discuss further in Chapter 5. In the remainder of this Chapter, we will give some theories that we found in other articles.

4.2.1 Cesàro law of the iterated logarithm

In [Gaposhkin, 1965] a reformulation for the law of the iterated logarithm (Theorem 29) using (C,k)-summation is given.

Theorem 46 ([Gaposhkin, 1965, p. 412])**.** Let $X_1, X_2, \cdots$ be a sequence of independent identically distributed random variables with mean 0 and variance 1. Let C_n^k be the (C,k)-sum of these random variables for $k > 0$,

$$C_n^k = \frac{1}{\binom{n+k}{k}} \sum_{i=1}^{n} \binom{n-i+k}{k} X_i. \tag{84}$$

Then

$$\limsup_{n\to\infty} \frac{C_n^k}{\sqrt{2n/(2k+1)\log\log n}} = 1 \text{ almost surely.} \tag{85}$$

It is remarkable to see that Cesàro summation is used in a different way then we saw before in 3.1.3. In Theorem 46 (C,1)-summation yields

$$C_n^1 = \frac{X_1 + X_2 + \cdots + X_n}{n+1}, \tag{86}$$

while we were familiar using a sequence $s_n = \sum_{i=0}^n a_i$ instead of X_n. The Cesàro summation technique is therefore not used to sum a sequence, but to give a mean of the sequence $X_1, X_2, \cdots$. Comparing Theorem 46 with Theorem 29 we see that in (85) an extra division by $2k+1$ is done in the square root. For $k = 0$ we see in (85) the normal formulation of the law of the iterated logarithm, as stated in Theorem 29, as (84) just gives $C_n^0 = \sum_{i=0}^n X_i = S_n$. For higher k it seems like Theorem 46 gives a more tight bound than Theorem 29

4.2.2 Riesz methods for random variables

In [Kozlov, 2003] we find some interesting theorems regarding random variables with mean zero combined Riesz summation (Theorem 41). Just like in Theorem 46, a summation technique, in Theorems 47 and 48 Riesz summation, is used to give a mean to the sequence $X_1, X_2, \cdots$. In Theorems 47 and 48 we see that $X_n \xrightarrow{a.s.} 0$ (R,p_n), which means that $\mathbb{P}(X_n \to 0 \text{ (R,}p_n))= 1$.

Theorem 47 ([Kozlov, 2003, p. 1479])**.** Let $X_0, X_1, \cdots$ be independent, identically distributed random variables with mean zero and finite moments of order $\leq 2k$, and let $p_0, p_1, \cdots$ satisfy

$$\sum_{n=0}^{\infty} \left[\frac{p_0^2 + \cdots + p_n^2}{(p_0 + \cdots + p_n)^2}\right]^k < \infty.$$

Then $X_n \xrightarrow{a.s.} 0$ (R,p_n).

Theorem 48 ([Kozlov, 2003, p. 1479])**.** Let $X_0, X_1, \cdots$ be independent random variables with mean zero and variances $\sigma_1^2, \sigma_2^2, \cdots$. If

$$\sum_{j=0}^{\infty} \frac{p_j^2 \sigma_j^2}{(p_0 + \cdots + p_j)^2} < \infty,$$

then $X_n \xrightarrow{a.s.} 0$ (R,p_n).

Example 12. We investigate the random variables $X_0, X_1, \cdots$ with

$$X_n = \begin{cases} -\frac{1}{n} & \text{with probability } \frac{1}{2}, \\ \frac{1}{n} & \text{with probability } \frac{1}{2}. \end{cases}$$

We see that $\mathbb{E}[X_n] = 0$ and $\mathrm{Var}(X_n) = \frac{1}{n^2}$. If we now take the Riesz means for sequence $p_n = 1$ for all n, we obtain by Theorem 48 that $X_n \xrightarrow{a.s.} 0$ (R,p_n), as we have that

$$\sum_{j=0}^{\infty} \frac{p_j^2 \sigma_j^2}{(p_0 + \cdots + p_j)^2} = \sum_{j=0}^{\infty} \frac{\frac{1}{j^2}}{(1 + \cdots + 1)^2} = \sum_{j=0}^{\infty} \frac{1}{j^2 (j+1)^2} < \infty. \tag{87}$$

Note that this particular Riesz means is a (H,1)- and (C,1)-summation, so we find that

$$\lim_{n \to \infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = 0 \quad \text{almost surely.} \tag{88}$$

5 Law of large numbers using Cesàro summation

In this section, we discuss an attempt to state an alternative version of the weak law of large numbers (Theorem 25) when we use Cesàro summability instead of normal convergence. We already saw that Theorem 46 is a reformulation of the law of the iterated logarithm (Theorem 29) that is using (C,k)-summation. Therefore it seems logical to us that we can reformulate the weak law of large numbers similarly. Finally, it turns out that there exists a reformulation of the strong law of large numbers in literature, and we will compare it to our approach.

5.1 Formulation of an statement

We first repeat both the weak law of large numbers and the definition of Cesàro summation (Definition 36). The weak law of large numbers tells us that when we have a sequence of independent identically distributed random variables with mean μ, the partial sums satisfies

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}S_n \xrightarrow{D} \mu \tag{89}$$

as $n \to \infty$. The Cesàro method tells us that when we have a sequence $a_0, a_1, a_2, \cdots$ for which for a certain k

$$C_n^k = \frac{A_n^k}{E_n^k} = \frac{1}{\binom{n+k}{k}} \sum_{i=0}^{n} \binom{n-i+k}{k} a_i \to s \tag{90}$$

when $n \to \infty$, then we say that $\sum a_i$ is (C,k) summable to s.

Our approach will be to replace S_n first by a Cesàro sum, and afterwards we will investigate whether we still have that $\tilde{C}_n^k/n$ converges in distribution to μ. We are interested in taking a sequence of random variables $X_1, X_2, \cdots$ as the sequence we are interested in. Therefore we want to replace S_n by

$$\tilde{C}_n^k = \frac{1}{\binom{n-1+k}{k}} \sum_{i=1}^{n} \binom{n-i+k}{k} X_i. \tag{91}$$

Note that we shift the index in the binomial coefficient of E_n^k as well. When we now do a division by n, we obtain

$$\frac{1}{n}\tilde{C}_n^k = \frac{1}{n\binom{n-1+k}{k}} \sum_{i=1}^{n} \binom{n-i+k}{k} X_i. \tag{92}$$

Substituting $k = 0$ in (92) gives

$$\frac{1}{n}\tilde{C}_n^0 = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}S_n \xrightarrow{D} \mu, \tag{93}$$

and it appears that (92) is an alternative formulation of the week law of large numbers. However, for $k > 0$ it does not work correctly. When we consider the expected value

$$\mathbb{E}\left[\frac{1}{n}C_n^k\right] = \frac{1}{n\binom{n-1+k}{k}} \sum_{i=1}^{n} \binom{n-i+k}{k} \mathbb{E}[X_i] \tag{94}$$

should equal μ. However for $k = 1$ we get, according to the Definition 36, that

$$\frac{1}{n}\tilde{C}_n^1 = \frac{S_1 + S_2 + \cdots + S_n}{n^2}, \tag{95}$$

and since $\mathbb{E}[S_n] = n\mu$, we find that

$$\mathbb{E}\left[\frac{1}{n}\tilde{C}_n^1\right] = \frac{\mu + 2\mu + \cdots + n\mu}{n^2} = \frac{\mu\sum_{i=1}^{n} i}{n^2} = \frac{\frac{1}{2}\mu n(n+1)}{n^2} \xrightarrow{n\to\infty} \frac{1}{2}\mu. \tag{96}$$

This is not exactly what we were hoping for, and we observe for $k = 2$ similarly that

$$\frac{1}{n}\tilde{C}_n^2 = \frac{S_1 + (S_1 + S_2) + \cdots + (\sum_{i=1}^n S_i)}{\frac{1}{2}n^2(n+1)} \tag{97}$$

such that the expectation

$$\mathbb{E}[\frac{1}{n}\tilde{C}_n^2] = \frac{\mu \sum_{i=1}^n \frac{1}{2}i(i+1)}{\frac{1}{2}n^2(n+1)} = \frac{\frac{1}{6}\mu n(n+1)(n+2)}{\frac{1}{2}n^2(n+1)} \xrightarrow{n\to\infty} \frac{1}{3}\mu. \tag{98}$$

Summarizing $2\mathbb{E}[\frac{1}{n}\tilde{C}_n^1] \to \mu$ and $3\mathbb{E}[\frac{1}{n}\tilde{C}_n^2] \to \mu$. Thus it seems like we need to add a factor $k+1$ in (92), which indeed makes no difference for $k = 0$. In general we see that

$$\mathbb{E}\left[\frac{k+1}{n}\tilde{C}_n^k\right] = \frac{k+1}{n}\frac{1}{\binom{n-1+k}{k}}\sum_{i=1}^n \binom{n-i+k}{k}\mathbb{E}[X_i] = \frac{k+1}{n}\frac{1}{\binom{n-1+k}{k}}\sum_{i=1}^n \binom{n-i+k}{k}\mu. \tag{99}$$

Now as $\sum_{i=1}^n \binom{n-i+k}{k} = \frac{k+n}{k+1}\binom{n-1+k}{k}$ we can simplify (99):

$$\mathbb{E}\left[\frac{k+1}{n}\tilde{C}_n^k\right] = \mu\frac{k+1}{n}\frac{1}{\binom{n-1+k}{k}}\frac{k+n}{k+1}\binom{n-1+k}{k} = \mu\frac{(n+k)}{n}. \tag{100}$$

Therefore we have indeed that, as long as k is fixed, $\mathbb{E}\left[\frac{k+1}{n}\tilde{C}_n^k\right] \to \mu$ as $n \to \infty$. Now we have created a sequence that just like $\mathbb{E}[\frac{1}{n}S_n] \to \mu$ as $n \to \infty$, which leads us to an idea for a Cesàro weak law of large numbers as stated in Idea 1.

Unproven idea 1. Let $X_1, X_2, X_3, \cdots$ be a sequence of i.i.d. random variables with $\mathbb{E}[X_1] = \mu$. For all finite $k \geq 0$,

$$\frac{k+1}{n}\tilde{C}_n^k = \frac{k+1}{n}\frac{1}{\binom{n-1+k}{k}}\sum_{i=1}^n \binom{n-i+k}{k}X_i \xrightarrow{D} \mu, \tag{101}$$

as $n \to \infty$.

Now it would be nice if we were able to prove Idea 1 analogously to the proof for the normal weak law of large numbers in Section 2.2.1. A crucial part in this proof is (20), where it is used that for $\mathrm{Var}(X_i) = \sigma^2$, we have that $\mathrm{Var}(S_n) = \sigma^2/n \to 0$ as $n \to \infty$. For $(k+1)\tilde{C}_n^k/n$ we calculate the variance, analogously to the expected value:

$$\mathrm{Var}\left(\frac{k+1}{n}\tilde{C}_n^k\right) = \left(\frac{k+1}{n}\frac{1}{\binom{n-1+k}{k}}\right)^2 \mathrm{Var}\left(\sum_{i=1}^n \binom{n-i+k}{k}X_i\right) \tag{102}$$

$$= \left(\frac{k+1}{n}\frac{1}{\binom{n-1+k}{k}}\right)^2 \sum_{i=1}^n \binom{n-i+k}{k}^2 \mathrm{Var}(X_i) \tag{103}$$

$$= \left(\frac{k+1}{n}\frac{1}{\binom{n-1+k}{k}}\right)^2 \sum_{i=1}^n \binom{n-i+k}{k}^2 \sigma^2. \tag{104}$$

Unfortunately we do not manage to simplify $\sum_{i=1}^n \binom{n-i+k}{k}^2$ in (104). Therefore the best we can do is give an upperbound for (104):

$$\mathrm{Var}\left(\frac{k+1}{n}\tilde{C}_n^k\right) \leq \left(\frac{k+1}{n}\frac{1}{\binom{n-1+k}{k}}\right)^2 \left(\sum_{i=1}^n \binom{n-i+k}{k}\right)^2 \sigma^2 = \frac{(n+k)^2}{n^2}\sigma^2, \tag{105}$$

which does not converge to 0 and we must conclude that we presently have insufficient identities to prove our idea.

Despite not being able to prove our idea, we will build a simulation to find out how (101) behaves.

5.1.1 Simulation

We made a simulation in which we generate random variables from a certain distribution. Then we calculate $(k+1)\tilde{C}_n^k/n$ for a fixed value of n and for different values of k. Of course, for a random variable with finite mean, we expect $(k+1)\tilde{C}_n^k/n$ to be around the same value for all values of k. We will simulate a certain number of independent identically distributed random variables and compute $(k+1)\tilde{C}_n^k/n$ for different k, and repeat that, to find an estimate and a 95% confidence interval for the value of $(k+1)\tilde{C}_n^k/n$. The pseudo-code described in Algorithm 1 represents what the simulation does exactly.

Algorithm 1: Pseudo-code of the simulation

Input : Distribution for X_i, $runs$ = number of runs, n = number of random variables each run, v = vector containing values for k

Output: mean and sd

for *j in 1:runs* **do**
    Build vector X with n samples from distribution;
    Build empty vector C ;
    Build empty matrix C-complete ;
    for *k in v* **do**
        Build vector A with n positions with 0;
        for *i in 1:n* **do**
            $A[i] = \binom{n-i+k}{k} X[i]$;
        end
        add $\dfrac{k+1}{n\binom{n+k}{n}} \sum A$ to vector C ;
    end
    Add C as row in C-complete ;
end
Build empty vector results ;
Build empty vector sd ;
for *h in 1:length(v)* **do**
    results[h]= mean(C-complete[column h]);
    sd[h]=standard deviation(C-complete[column h]);
end
return: results, sd

5.1.2 Simulation results

We first inspect whether our simulation works correctly, by setting $X_i = (-1)^{i+1}$ such that we get the sequence $1 - 1 + 1 - 1 + \cdots$, which is (C,1)-summable to $\frac{1}{2}$ (Example 6). We need to change the computation of C a little, so we change line 10 in Algorithm 1 in add $(\sum A)/\binom{n+k}{n}$ to vector C. Table 2 contains the results.

From Table 2, we see that the (C,0)-sum is equal to 0 for n even and 1 for n odd. Furthermore, we see that the (C,1)-sum goes to $\frac{1}{2}$, as expected. For higher k we see that the larger the n, the more accurate the results are. For the simulations, using sequences of random variables we will take sequences of 10000 random variables, Increasing the simulation length to 100000 makes the simulation too slow, especially when we want to do more runs.

k	0	1	2	10	20	30	40
C_{1000}^k	0	0.5	0.5004995	0.5024888	0.5049530	0.5073929	0.5098088
C_{1001}^k	1	0.5004995	0.5004995	0.5024863	0.5049481	0.5073856	0.5097992
C_{10000}^k	0	0.5	0.50005	0.5002499	0.5004995	0.5007489	0.5009981
C_{100000}^k	0	0.5	0.500005	0.500025	0.50005	0.500075	0.5001

Table 2: C_n^k for $1 - 1 + 1 - 1 + \cdots$.

Bernoulli distribution

We will now investigate a random variable with a known finite mean. In particular we take $X_i \sim \text{Ber}(\frac{1}{2})$. With 10000 runs we obtain the results with 95% confidence intervals in Figure 3. We have added a zoomed in situation in the frame, because otherwise the confidence intervals are too small to be visible. We see that the results are mostly in line with what we expect, $\mu = 0.5$. For larger k we observe that $(k+1)\tilde{C}_{10000}^k/10000$ becomes a little larger then we expect. This can be declared by (100), for instance for $k = 40$ we have that $E[(k+1)\tilde{C}_{10000}^k/10000] = \mu \cdot 10040/40 = 1.004\mu = 0.502$.

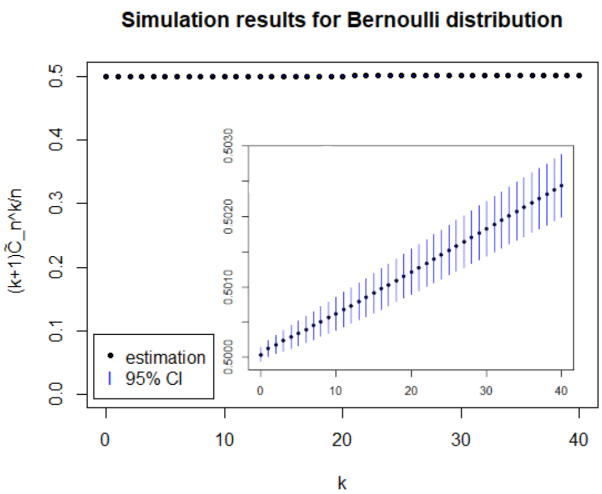


Figure 3: Estimations for $(k+1)\tilde{C}_{10000}^k/10000$ for identically distributed Bernoulli random variables.

Pareto distribution

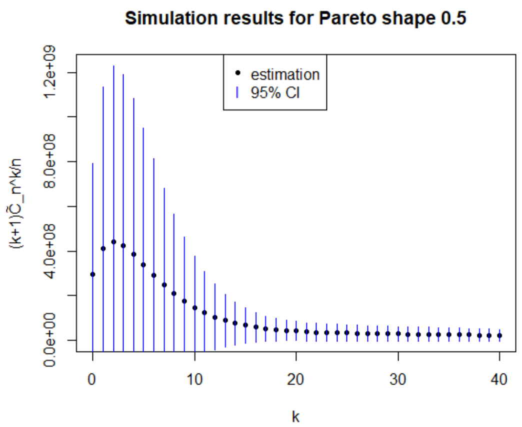
Next we are interested in the Pareto distribution because it has some interesting properties. The Pareto distribution depends on a scale $x_m > 0$ and a shape $\alpha > 0$ (Definition 14). For $\alpha \leq 1$ the expected value of a Pareto distributed random variable equals ∞, furthermore for $\alpha \leq 2$ the variance is ∞.

We will do a simulation for different values of the shape α, we chose those as we wanted to see what happens for Pareto random variables when the expectation is ∞, or when the variance is ∞ (for $\alpha \leq 2$). In Figure 4 we see the results of 10000 runs in which each 10000 random Pareto random variables with scale $x_m = 1$ and different shapes are simulated.

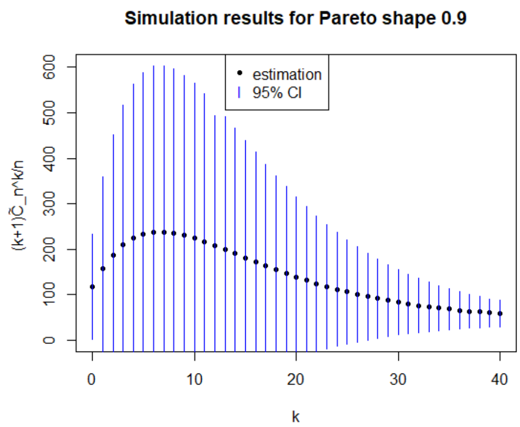
We see in Figure 4a that all estimations are at least of order 10^7. By (100) we know that $E[(k+1)\tilde{C}^k_{10000}/10000] = \mu(10000+k)/10000 \approx \mu$, so we see that μ is quite large, as expected. The Pareto distribution is a *heavy tailed* distribution [Fackler, 2013] for these three parameters, which means in practice that for one single run with shape 0.5, 9000 of the 10000 X_i can be less than 100, but the others can take values up to 10^{10}. The bigger the shape parameter α is, the lower the number of these outliers is. For Figure 4b and 4c we see a relative small mean, despite expecting an infinite mean. This is probably because we only generated 10000 random variables, so not many outliers occurred.

For shape $\alpha = 1.5$ in Figure 4d we already see small confidence intervals, why we had to add the zoomed in version to make it visible. The estimated values in Figure 4d are in line with what we expect, namely $\mu = 3$, the infinite variance does not affect that. In comparison for what we said for shape $\alpha = 0.5$, for shape $\alpha = 1.5$ we have only 10 to 20 of such outliers that are bigger than 100 and they are probably responsible for the bigger confidence intervals than for shape $\alpha = 2$ and shape $\alpha = 2.5$ in Figure 4e and 4f. For shape $\alpha = 2$ we only have 0 or 1 X_i that is larger than 100. For shape 2.5 we do not have any, in most cases.

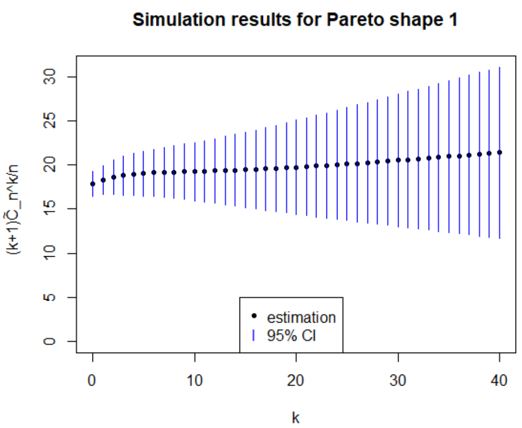
Focusing only on the estimated values in Figure 4e, and 4f we see the estimations we expected, for shape $\alpha = 2$ we indeed have $\mu = 2$, and for shape $\alpha = 2.5$ we have $\mu \approx 1.67$. In the three simulations for Pareto distribution with finite mean (Figures 4d, 4e and 4f), we see the same as in Figure 3: for higher k the estimations are a little increasing.



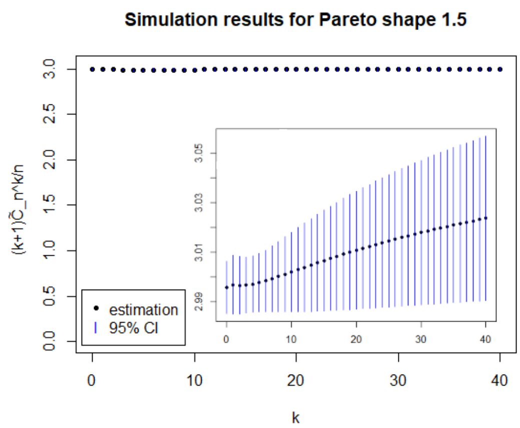
(a) Pareto distribution with shape $\alpha = 0.5$.



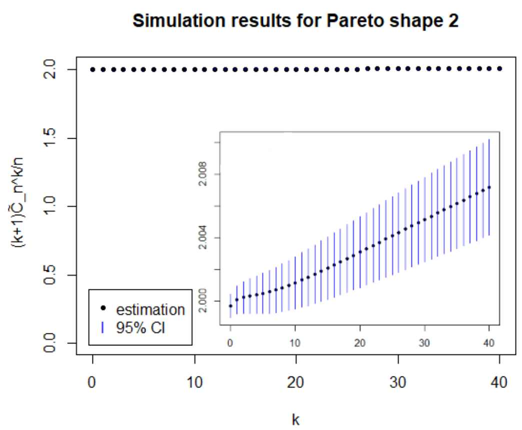
(b) Pareto distribution with shape $\alpha = 0.9$.



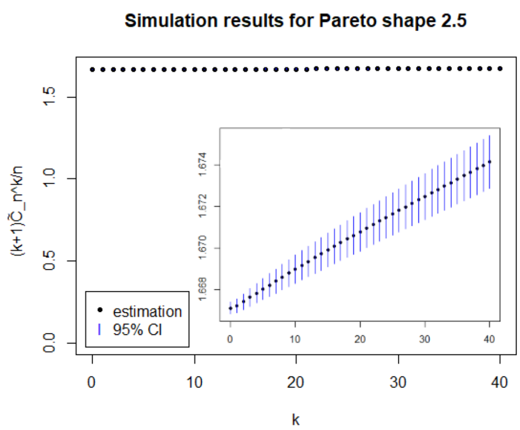
(c) Pareto distribution with shape $\alpha = 1$.



(d) Pareto distribution with shape $\alpha = 1.5$.



(e) Pareto distribution with shape $\alpha = 2$.



(f) Pareto distribution with shape $\alpha = 2.5$.

Figure 4: Estimations for $(k+1)\tilde{C}^k_{10000}/10000$ for Pareto distributed random variables.

5.2 Cesàro strong law of large numbers

After we attempted to formulate a Cesàro law of large numbers by ourselves, it turned out there already exists one. In [Lai, 1974] and [Mursaleen, 2014] a theorem is given regarding summability in the Cesàro sense of a sequence of random variables.

Theorem 49 ([Lai, 1974])**.** If $X_1,\ X_2, \cdots$ is a sequence of independent, identically distributed random variables, and $k \geq 1$, then the following are equivalent:

i. $\mathbb{E}[X_1] = \mu$

ii. $X_n \xrightarrow{a.s.} \mu$ (C,1), i.e.,

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu \text{ a.s.} \tag{106}$$

iii. $X_n \xrightarrow{a.s.} \mu$ (C,k), i.e.,

$$\lim_{n\to\infty} C_n^k = \lim_{n\to\infty} \frac{1}{\binom{n+k}{k}} \sum_{i=1}^{n-1} \binom{i+k-1}{i} X_{n-i} = \mu \text{ a.s.} \tag{107}$$

We see that just like in Sections 4.2.1 and 4.2.2, the summation technique is used to give a mean to the sequence $X_1, X_2, \cdots$. Almost sure convergence of X_n to μ (C,k) in Theorem 49 thus means that (107) is an event with probability 1, so $\mathbb{P}(X_n \to \mu(\mathrm{C}, k)) = 1$

Note that Theorem 49 implies a Cesàro strong law of large numbers. Furthermore, as almost sure convergence implies convergence in distribution, the weak law also holds in the form of Theorem 49. The complete proof of Theorem 49 is given by [Mursaleen, 2014, p. 51] and we of course see that and iii. follows from ii. by Theorem 39.

The approach of Theorem 49 is different from our own approach, as we took S_n as our (C,0)-sum, while Theorem 49 takes $S_n/(n+1)$ as the (C,1)-sum. It seems like our approach was more complicated, as we needed to add the factor $k+1$, while the expectation of (107) converges to μ, which we will see in (111).

To compare the behaviour of this Cesàro law of large numbers with our own approach we will run the same simulation on this form. We need to adjust Algorithm 1 only on lines 8 and 10. Line 8 we replace by $A[i] = \binom{i+k-1}{i} X[n-i]$ and Line 10 by add $\binom{n+k}{k} \sum A$ to vector C.

Before we start the simulation we are interested in the expected value of C_n^k:

$$\mathbb{E}[C_n^k] = \frac{1}{\binom{n+k}{k}} \sum_{i=1}^{n-1} \binom{i+k-1}{i} \mathbb{E}[X_{n-i}] = (n-1)\mu \frac{1}{\binom{n+k}{k}} \sum_{i=1}^{n-1} \binom{i+k-1}{i}. \tag{108}$$

Filling in that $\sum_{i=1}^{n-1} \binom{i+k-1}{i} = (n/k) \cdot \binom{n-1+k}{n} - 1$ in (108) gives

$$\mathbb{E}[C_n^k] = (n-1)\mu \frac{1}{\binom{n+k}{k}} \Big(\frac{n\binom{n-1+k}{n}}{k} - 1 \Big). \tag{109}$$

Since $\binom{n+k}{n+1} / \binom{n-1+k}{k} = k/(n+k)$, we find that (109) gives

$$\mathbb{E}[C_n^k] = (n-1)\mu \Big(\frac{1}{n+k} - \frac{1}{\binom{n+k}{k}} \Big). \tag{110}$$

From (110) we expect that the estimations we find for C_n^k are probably decreasing as k increases. Furthermore as $\lim_{n\to\infty} 1/\binom{n+k}{k} = 0$ we conclude that

$$\lim_{n\to\infty} \mathbb{E}[C_n^k] = \lim_{n\to\infty} (n-1)\mu \left(\frac{1}{n+k} - \frac{1}{\binom{n+k}{k}} \right) = \lim_{n\to\infty} \frac{\mu(n-1)}{n+k} = \mu. \tag{111}$$

5.2.1 Simulation results

For these simulations we did not calculate C_n^k for $k = 0$ as Theorem 49 is defined for $k \geq 1$.

Bernoulli distribution

We will again investigate $X_i \sim \text{Ber}(\frac{1}{2})$. With 10000 runs we obtain the results with 95% confidence intervals in Figure 5. Again we give a zoomed in version, as the confidence intervals were invisible otherwise. What is most remarkable in contrast to Figure 3 is that Figure 5 is decreasing for higher k, while 3 is increasing. Our prognosis based on (110) that C_n^k decreases as k increases is indeed right. Besides this, we see not much difference between Figure 5 and Figure 3, the estimations for C_{10000}^k are also close to the expected value of the distribution, 0.5.

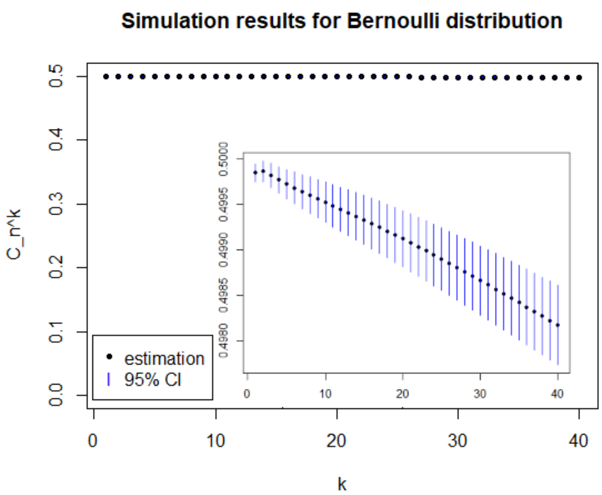


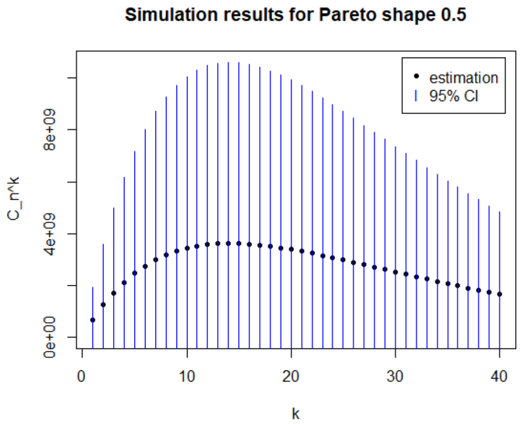
Figure 5: Estimations for C_{10000}^k for identically distributed Bernoulli random variables.

Pareto distribution

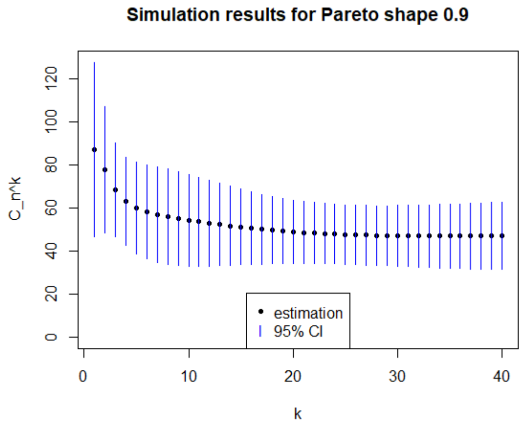
Again we do the simulation for Pareto distributed random variables with different shapes α, just as we did in Section 5.1.2. The results are shown in Figure 6.

We see that the estimated values in Figures 6a, 6b and 6c have the same size as we saw before in Figures 4a, 4b and 4c. The shapes of the confidence intervals in Figures 6a and 6b differ from those we saw in Figures 4a and 4b, which can be explained by how the Cesàro transformations are done. When an outlier happens in the beginning of the sequence of random variables, it is summed more often in the Cesàro transformation. For instance for (C,2), X_1 is summed n times, while X_n is summed only 1 time. Therefore when an outlier happens in the beginning it has more consequence to the C_n^k of that run, resulting in a larger confidence interval. Besides this, we need to note that [Mursaleen, 2014] states that a Cesàro method requires $\mathbb{E}[|X_1|] < \infty$ for summability. In our situation we of course have for $\alpha = 0.5$, $\alpha = 0.9$, and $\alpha = 1$ that $\mathbb{E}[X_1] = \infty$. Therefore we must probably conclude that a Cesàro method does not sum these three to a value, so the limit of C_n^k as $n \to \infty$ is ∞.

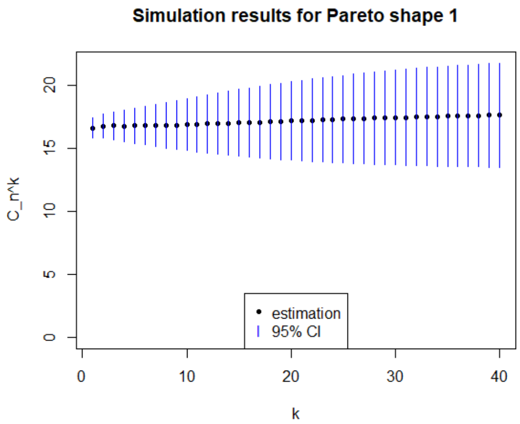
Just as for the Bernoulli distributed random variables, we see that all estimations of the random variables with finite mean decrease in Figures 6d, 6e and 6f. Except for this difference with Figure 4, we see that the confidence intervals have the same behavior as explained in section 5.1.2.



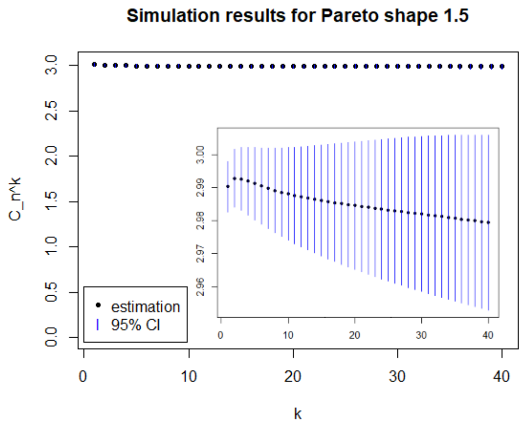
(a) Pareto distribution with shape $\alpha = 0.5$.



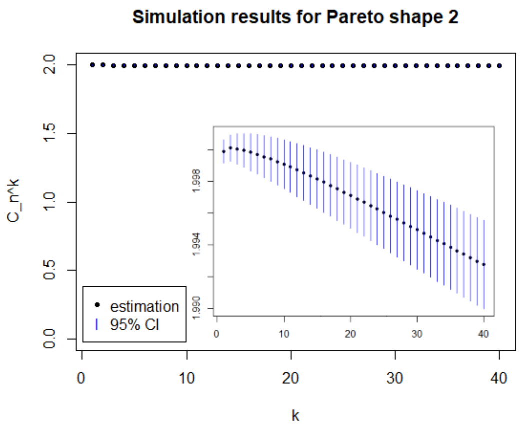
(b) Pareto distribution with shape $\alpha = 0.9$.



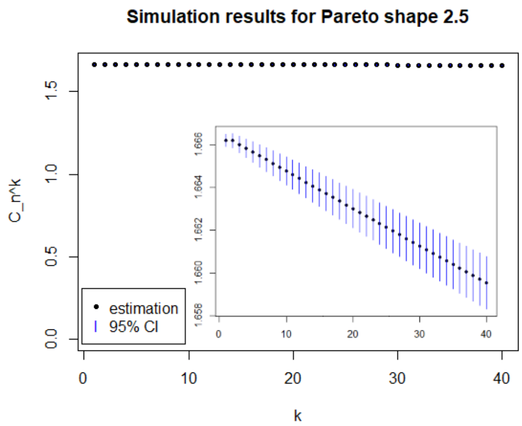
(c) Pareto distribution with shape $\alpha = 1$.



(d) Pareto distribution with shape $\alpha = 1.5$.



(e) Pareto distribution with shape $\alpha = 2$.



(f) Pareto distribution with shape $\alpha = 2.5$.

Figure 6: Estimations for C_{10000}^k for Pareto distributed random variables.

5.3 Comparison between Idea 1 and the Cesàro law of large numbers

Formulating Idea 1 we faced the fact that we needed to add a factor $k+1$ to let $\tilde{C}_n^k/n$ have expectation μ. When we look at the differences between Idea 1 and Theorem 49 it becomes clear that we needed to add this factor by the choice of our (C,0)-sum, S_n. In Idea 1 we thought it was needed to change the index of E_n^k in (91), however, we see in Theorem 49 that this difference does not matter in the limit. We see that both $\binom{n+k}{k}$ and $\binom{n-1+k}{k}$ go to infinity as $n \to \infty$.

For the sum part we see that [Lai, 1974] takes $\sum_{i=1}^{n-1} \binom{i+k-1}{i} X_{n-i}$ in Theorem 49, while in Idea 1 we took $\sum_{i=1}^{n} \binom{n-i+k}{n} X_i$. What is not entirely clear is why [Lai, 1974] sums only up to $n-1$, while Theorem 46 sums $\sum_{i=1}^{n} \binom{n-i+k}{k} X_i$ to n, nevertheless in the limit this makes no difference. Note that furthermore only the indexes are swifted, and that $\binom{n+k}{k} = \binom{n+k}{n}$, so both sums have the same limit.

In Idea 1 we took S_n as our (C,0)-sum, and by dividing this by n we got of course $(X_1+\cdots+X_n)/n$, which is almost the (C,1)-sum as defined in Theorem 49, the only difference is the division by $n+1$ instead of n. Now the (C,1)-sum in Idea 1 is $2(S_1+\cdots+S_n)/n^2$, while the (C,2)-sum as defined in Theorem 49 gives $(S_1+\cdots S_{n-1})/(\frac{1}{2}(n+1)(n+2)) = 2(S_1+\cdots S_{n-1}))/((n+1)(n+2))$. Now it becomes clear where the factor $k+1$ comes from, as we saw in subsection 3.1.3 in (40) that $E_n^k = \binom{n+k}{k}$ behaves like $n^k/k!$ as $n \to \infty$ and therefore $\lim_{n\to\infty} E_n^{k+1} = \lim_{n\to\infty} n/(k+1)E_n^k = n^{k+1}/(k!(k+1)) = n^{k+1}/((k+1)!)$. Therefore we must conclude that our choice of $A_n^0 = S_n$ made things more difficult, as we needed to adjust E_n^k and the manner of Theorem 49, taking $A_n^0 = X_n$, is to be preferred.

From both our own Idea 1 as Theorem 49 we must conclude that Cesàro methods do not sum random variables with infinite mean. Moreover, we know from [Mursaleen, 2014] that for Theorem 49 the Cesàro method does only sum random variables with $\mathbb{E}[|X_1|] < \infty$.

6 Conclusion and discussion

In the introduction of this report we stated two research questions that we hoped to answer:

- Is there some notion of divergent series in probability theory?
- Can we find mathematical applications of the summation techniques for divergent series in probability theory?

To answer these research questions we first went through the most important convergence theorems in probability theory in Chapter 2, which is considered an important background for this report. We saw different modes of convergence, the law of large numbers, the central limit theorem, and the law of the iterated logarithm. In Chapter 3 we went through summation techniques for divergent series; most we considered were so-called transformation techniques. We also attempted to link probability theory to the moment constant method, and in the example where we used moments of the uniform distribution this seemed to work correctly. However, the moments of the Pareto distribution in general do not suit the requirements for the moment method.

In Chapter 4 we searched for applications and notions of divergent series in probability theory. We thought of random variables that have divergent expectations and we investigated two examples. The sequences of random variables $X_i \sim (-1)^{i+1} \cdot 2{\cdot}\mathrm{Ber}(\frac{1}{2})$ and $X_i \sim (-1)^{i+1} \cdot 2i{\cdot}\mathrm{Ber}(\frac{1}{2})$ have the property that $S_n = \sum_{i=1}^{n} X_i$ has expected value $\mathbb{E}[S_n] = \frac{1}{2}$ (C,1) and $\mathbb{E}[S_n] = \frac{1}{4}$ (C,2) respectively.

Besides this, we saw in Chapter 4 that there exist theorems that combine the summation techniques with convergence theorems in probability theory. We saw that the law of the iterated logarithm has a form were S_n was replaced by a (C,k)-sum. We furthermore found some theorems regarding so-called Riesz summability of random variables. What was remarkable about the theorems we found, was that the summation techniques were used to give a mean to sequences of random variables and not to sum their series. Moreover, the Theorems in which summation techniques are used within probability theory are only applicable to random variables with finite means.

In Chapter 5 we tried to formulate an alternative version of the weak law of large numbers, applying a (C,k)-sum instead of a normal limit of S_n. We saw that by adding a factor $k+1$ and division by n we indeed had that $\mathbb{E}[(k+1)\tilde{C}_n^k/n) \to \mu$ as $n \to \infty$. Unfortunately we did not manage to prove Idea 1 that $(k+1)\tilde{C}_n^k/n \xrightarrow{D} \mu$ as $n \to$. The problem we encountered was that the variance of $(k+1)\tilde{C}_n^k/n$ was hard to calculate and we could not find a bound that was sharp enough for the variance to converge to 0. Maybe a proof of the weak law of large numbers using the characteristic function and not Chebychev's inequality will allow one to prove Idea 1. A bit later, we found out that a Cesàro strong law of large numbers already existed.

One of the things we were interested was to investigate our own Cesàro law of large numbers and compare it to the one stated in literature, and see how it behaves for random variables with infinite means. By doing simulations on multiple values of k, we saw that a random variable with infinite mean, both our own idea and the Cesàro strong law of large numbers give this infinite mean. The summation techniques do not seem to help in solving this problem, as they only seem to help solving alternating series that are not absolutely convergent. Furthermore, we saw that for the Cesàro strong law of large numbers a finite mean is needed to make it Cesàro summable.

For further applications of the summation techniques, we may have to look into the central limit theorem. For instance for the random variables we considered, $X_i \sim (-1)^{i+1} \cdot 2{\cdot}\mathrm{Ber}(\frac{1}{2})$ and $X_i \sim (-1)^{i+1} \cdot 2i{\cdot}\mathrm{Ber}(\frac{1}{2})$, one could investigate how the central limit theorem behaves with it. Another thing one could try is to reformulate the central limit theorem, by replacing the normal sum by a Cesàro sum, in a similar way as we saw for the law of the iterated logarithm and the law of large numbers.

References

[Bagni et al., 2005] Bagni, G. T. et al. (2005). Infinite series from history to mathematics education. *International Journal for Mathematics Teaching and Learning*, pages 1473–0111.

[David, 1962] David, F. N. (1962). *Games, gods and gambling: The origins and history of probability and statistical ideas from the earliest times to the Newtonian era.* Hafner Publishing Company.

[Debnath and Basu, 2015] Debnath, L. and Basu, K. (2015). A short history of probability theory and its applications. *International Journal of Mathematical Education in Science and Technology*, 46(1):13–39.

[Donice McCune and Luna McCune, 2000] Donice McCune, E. and Luna McCune, S. (2000). Estimation of the pareto shape parameter. *Communications in Statistics-Simulation and Computation.*, 29(4):1317–1324.

[Fackler, 2013] Fackler, M. (2013). Reinventing pareto: Fits for both small and large losses. In *ASTIN Colloquium, Den Haag.*

[Ferraro, 1999] Ferraro, G. (1999). The first modern definition of the sum of a divergent series: an aspect of the rise of 20th century mathematics. *Archive for history of exact sciences*, 54(2):101–135.

[Gaposhkin, 1965] Gaposhkin, V. F. (1965). The law of the iterated logarithm for Cesàro's and Abel's methods of summation. *Theory of Probability & Its Applications*, 10(3):411–420.

[Grimmett and Stirzaker, 2001] Grimmett, G. and Stirzaker, D. (2001). *Probability and random processes.* Oxford university press.

[Hardy, 1949] Hardy, G. H. (1949). *Divergent series.* Clarendon Press.

[Hofstad, 2018] Hofstad, R. v. d. (2018). *Handout limietstellingen Kansrekening 2WS20.*

[Kline, 1983] Kline, M. (1983). Euler and infinite series. *Mathematics Magazine*, 56(5):307–315.

[Kolmogorov and Bharucha-Reid, 2018] Kolmogorov, A. N. and Bharucha-Reid, A. T. (2018). *Foundations of the theory of probability: Second English Edition.* Courier Dover Publications.

[Koralov and Sinai, 2007] Koralov, L. and Sinai, Y. G. (2007). *Theory of probability and random processes.* Springer Science & Business Media.

[Kosmala, 2004] Kosmala, W. A. (2004). *A friendly introduction to analysis.* Pearson Prentice Hall.

[Kozlov, 2003] Kozlov, V. (2003). Summation of divergent series and ergodic theorems. *Journal of Mathematical Sciences*, 114(4):1473–1490.

[Lai, 1974] Lai, T. L. (1974). Summability methods for independent identically distributed random variables. *Proceedings of the American Mathematical Society*, 45(2):253–261.

[Leone et al., 1961] Leone, F., Nelson, L., and Nottingham, R. (1961). The folded normal distribution. *Technometrics*, 3(4):543–550.

[Mursaleen, 2014] Mursaleen, M. (2014). *Applied summability methods.* Springer.

[Tijms, 2012] Tijms, H. (2012). *Understanding probability.* Cambridge University Press.

[Vaart, 1998] Vaart, A. W. v. d. (1998). *Stochastic Convergence*, page 5–24. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.