Random Harmonic Series

Randy Linder
Advisor: Adam Fieldsteel

A thesis in Mathematics submitted to the faculty of Wesleyan University in partial fulfillment of the requirements for the degree of Master of Arts

Wesleyan University

Middletown, Connecticut

Spring, 2014

Acknowledgements

I would like to start by thanking Wesleyan University. Studying mathematics at Wesleyan was an incredible opportunity to take a variety of interesting courses that were taught by passionate and engaging professors. Elementary number theory with Chan, fundamentals of analysis with Fieldsteel, abstract algebra II with Rasmussen, and combinatorics with Collins especially stood out. While I entered Wesleyan excited about mathematics only for its application to physics and other disciplines, I will graduate feeling thrilled to ramble to innocent bystanders about the beauty and depth of mathematics. To that I owe the contagiously ardent and welcoming department. Additionally, I am grateful for the opportunity to do a senior honor's research with Professor Collins learning a recent proof of Tychonoff's theorem and I am especially appreciative of the fifth year, allowing me to continue to take classes, and research a topic I am excited about.

Secondly, I would like to directly thank Professor Fieldsteel. I took four incredibly interesting and riveting analysis courses with him including fundamentals of analysis, analysis I (real), analysis I (complex), and analysis II (functional analysis). In addition, he spent hours of his time a week with me helping me understand, and effectively write up the mathematics pertaining to our research project. He is an incredible mathematician, and one I will always look up to.

Thirdly, I want to thank Professor Pollack and Professor Chan. Pollack the first mathematics professor I had as a transfer student to Wesleyan, where I took vectors and matrices and problem-solving for the Putnam. Since then he has inspired me to challenge myself and he consistently made time for me whenever I e-mailed him or stopped by his office for any math question I puzzled over. Professor Chan has incredible pedagogical understanding and explicated topics more clearly than any other teacher. His elementary number theory course sparked my passion for mathematics, and it is a course I will always remember.

Lastly, I want to thank my unrelentingly supportive and loving family. Mom, Dad, and Jason: I couldn't have ever been as successful in math, and everything else I am proud of, without your extraordinary confidence in me.

Contents

1 Prologue

We begin our discussion by recalling the elementary fact that the harmonic series diverges:

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

Further, by the alternating series test, we know that $\sum_{k=1}^{\infty} \frac{(-1)^k}{k}$ converges.

But what about assigning other coefficients to the terms of the harmonic series in some other way? In this thesis, we will examine the convergence properties of the sums of this form:

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

for various sorts of sequences $\{a_k\}_{k=1}^{\infty} \subset \{1, -1\}^{\mathbb{N}}$. Most of our work will concern *random* sequences $\{a_k\}_{k=1}^{\infty}$ in a sense that we will describe later. But here we begin by discussing some more elementary examples.

For example, we might consider a periodic sequence of *signs* ± 1, where an equal number of 1's and -1's is used in the repeated pattern. More generally suppose the sequence $\{a_k\}_{k=1}^{\infty}$ has the following property: for some $N \in \mathbb{N}$, and for every $k \in \mathbb{N}$, the sequence $\{a_{kN+i}\}_{i=1}^{N}$ has an equal number of 1's and -1's. That is, the sequence of coefficients can be divided into consecutive blocks of length N, in each of which there are equal numbers of 1's and -1's, though the sequence need not be periodic.

Proposition 1.1 *The series satisfying the above characteristic converges.*

To see this, we divide the series into N interwoven convergent alternating sub-series. One such sub-series would have the form

$$\sum_{j=1}^{\infty} \frac{a_{k_j}}{k_j},$$

where the k_j are chosen so that for each j,

$$k_j \in [jN, (j+1)N)$$

and so that the signs $\{a_{k_j}\}_{j=1}^{\infty}$ alternate. By the alternating series test, such a series would converge. The conditions on the whole series allow us to obtain N such seris, where the i^{th} series is given by a choice of indices $\{k_j^{(i)}\}_{j=1}^{\infty}$:

$$\sum_{j=1}^{\infty} \frac{a_{k_j}^{(i)}}{k_j^{(i)}}.$$

It follows that whole series can be re-written as an iterated sum of these convergent series:

$$\sum_{k=1}^{\infty} \frac{a_k}{k} = \sum_{i=1}^{N} \sum_{j=1}^{\infty} \frac{a_{k_j}^{(i)}}{k_j^{(i)}}.$$

A still more general result can be obtained by somewhat more elaborate, but still elementary arguments. Suppose that the sequence of signs $\{a_k\}_{k=1}^{\infty}$ has the property that the sequence of partial sums

$$\left\{ \sum_{k=1}^{n} a_k \right\}_{n=1}^{\infty}$$

is bounded. Then the series $\sum_{k=1}^{\infty} \frac{a_k}{k}$ converges.

Note that this result contains the previous one; if the coefficients a_k meet the blocked condition of the previous example, then their partial sums will be bounded.

In order to establish this result, we will need to consider general series of the form $\sum_{k=1}^{\infty} a_k b_k$ and use Dirichlet's test. We begin with the following result:

Theorem 1.2 (Summation by parts) *Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be two sequences of complex numbers, and let $s_n = a_1 + a_2 + ... + a_n$, where we define $s_0 = 0$.*

Then, for all $n \in \mathbb{N}$

$$\sum_{k=1}^{n} a_k b_k = s_n b_{n+1} - \sum_{k=1}^{n} s_k (b_{k+1} - b_k)$$

Proof: Observe that $s_{n-1} = a_1 + a_2 + ... + a_{n-1}$ and so $a_n = s_n - s_{n-1}$.

Therefore,

$$\sum_{k=1}^{n} a_k b_k = \sum_{k=1}^{n} (s_k - s_{k-1}) b_k.$$

Expanding, we get that

$$\sum_{k=1}^{n} a_k b_k = (s_1 - s_0) b_1 + (s_2 - s_1) b_2 + ... + (s_{n-1} - s_{n-2}) b_{n-1} + (s_n - s_{n-1}) b_n$$

$$= s_1 b_1 - s_0 b_1 + s_2 b_2 - s_1 b_2 + ... + s_{n-1} b_{n-1} - s_{n-2} b_{n-1} + s_n b_n - s_{n-1} b_n$$

$$= s_1 (b_1 - b_2) + s_2 (b_2 - b_3) + ... + s_{n-1} (b_{n-1} - b_n) + s_n b_n.$$

Reordering, we have that

$$\sum_{k=1}^{n} a_k b_k = s_n b_n - \sum_{k=1}^{n-1} s_k(b_{k+1} - b_k) = s_n b_n + s_n b_{n+1} - s_n b_{n+1} - \sum_{k=1}^{n-1} s_k(b_{k+1} - b_k)$$

$$= s_n b_{n+1} - s_n(b_{n+1} - b_n) - \sum_{k=1}^{n-1} s_k(b_{k+1} - b_k) = s_n b_{n+1} - \sum_{k=1}^{n} s_k(b_{k+1} - b_k)$$

and we obtain the summation by parts formula. ■

Observe that the summation by parts is the discrete analog of integration by parts. If we let

$$\Delta s_k = s_{k+1} - s_k = \frac{s_{k+1} - s_k}{(k+1) - k}$$

and substitute Δs_k for a_k, we obtain

$$\sum_{k=1}^{n} b_k \Delta s_{k-1} = s_n b_{n+1} - \sum_{k=1}^{n} s_k \Delta b_k$$

which is formally analogous to

$$\int b ds = sb - \int s db.$$

As a preliminary application to the summation by parts formula, we prove the following corollary, which will be used in chapter 3.

Corollary 1.3 *Suppose that $\{b_k\}$ is a sequence that decreases monotonically to $b > 0$. Further, suppose that $\sum_{k=1}^{\infty} a_k$ is a series of real terms whose partial sums tend to infinity:*

$$\lim_{n\to\infty} \sum_{k=1}^{n} a_k = \infty.$$

Then $\sum_{k=1}^{\infty} a_k b_k$ diverges.

Proof: Using the summation by parts formula, we have have that for each n

$$\begin{aligned} \sum_{k=1}^{n} a_k b_k &= s_n b_{n+1} - \sum_{k-1}^{n} s_k \left(b_{k+1} - b_k\right) \\ &= s_n b_{n+1} + \sum_{k-1}^{n} s_k \left(b_k - b_{k+1}\right). \end{aligned}$$

The terms $s_k \left(b_k - b_{k+1}\right)$ are all non-negative, once k is sufficiently large. But by assumption,

$$\lim_{n\to\infty} s_n b_{n+1} = \infty.$$

Therefore the sums $\sum_{k-1}^{n} s_k \left(b_k - b_{k+1}\right)$ are bounded below. Hence the overall sum must diverge to ∞. ■

In particular, if $\sum_{k=1}^{\infty} a_k \left(\frac{1}{k}\right)$ diverges to ∞, then since $\left\{\frac{k}{k-1}\right\}_{k=2}^{\infty}$ decreases to 1, we conclude that

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

diverges to ∞.

Now we prove a version of Abel's lemma, allowing for complex numbers.

Lemma 1.4 (Abel's lemma) *Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of complex numbers. For each n, let $s_n = \sum_{k=1}^{n} a_k$ and suppose there exists $M \in \mathbb{R}$ such that for all $n \in \mathbb{N}$, $|s_n| \leq M$. Let $\{b_n\}_{n=1}^{\infty}$ be a non-increasing sequence of non-negative reals. Then, for all $n \in \mathbb{N}$*

$$\left|\sum_{k=1}^{n} a_k b_k\right| \leq M b_1$$

Proof: By the summation by parts formula,

$$\sum_{k=1}^{n} a_k b_k = s_n b_{n+1} - \sum_{k=1}^{n} s_k (b_{k+1} - b_k)$$

Thus,

$$\begin{aligned}
\left|\sum_{k=1}^{n} a_k b_k\right| &\leq |s_n b_{n+1}| + \left|\sum_{k=1}^{n} s_k (b_{k+1} - b_k)\right| \\
&\leq M b_{n+1} + M \sum_{k=1}^{n} |(b_{k+1} - b_k)| \\
&= M b_{n+1} + M \sum_{k=1}^{n} (b_k - b_{k+1}) \\
&= M[(b_1 - b_2) + (b_2 - b_3) + \ldots + (b_n - b_{n+1}) + b_{n+1}] \\
&= M b_1. \blacksquare
\end{aligned}$$

We immediately obtain the following:

Corollary 1.5 *Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of complex numbers. For each n, let $s_n = \sum_{k=1}^{n} a_k$ and suppose there exists $M \in \mathbb{R}$ such that for all $n \in \mathbb{N}$, $|s_n| \leq M$. Let $\{b_n\}_{n=1}^{\infty}$ be a non-decreasing sequence of non-positive reals. Then, for all $n \in \mathbb{N}$*

$$\left|\sum_{k=1}^{n} a_k b_k\right| \leq M |b_1|.$$

Proof: Replace the a_k and b_k by their negatives and apply the previous lemma. ■

Now, using Abel's lemma, we can easily show Dirichlet's test for convergence. Again, we will allow some of the terms to be complex.

Theorem 1.6 (Dirichlet's test) *Suppose $\{a_k\}_{k=1}^{\infty}$ is a sequence of complex numbers whose partial sums $s_n = \sum_{k=1}^{n} a_k$ are bounded. That is, for some $M \in \mathbb{R}$ and for all n*

$$|s_n| \leq M.$$

Suppose that $\{b_k\}_{k=1}^{\infty}$ is monotonic sequence of real numbers converging to 0. Then the series $\sum_{k=1}^{\infty} a_k b_k$ converges.

Proof: A series is convergent if and only if it is Cauchy. Thus, we must show that for all $\epsilon > 0$, there exists $N \in \mathbb{N}$, such that for all $n, m \in \mathbb{N}$, with $n \geq m \geq N$,

$$\left| \sum_{k=m}^{n} a_k b_k \right| < \epsilon.$$

Now, by hypothesis, for all $n, m \in \mathbb{N}$ and $m \leq n$,

$$\left| \sum_{k=m}^{n} a_k \right| = |s_n - s_{m-1}| \leq |s_n| + |s_{m-1}| \leq 2M.$$

By Abel's Lemma, for all $n, m \in \mathbb{N}$ and $m \leq n$,

$$\left| \sum_{k=m}^{n} a_k b_k \right| \leq 2M \left| b_m \right|.$$

Now, because $b_k \to 0$, there exists $N \in \mathbb{N}$ such that for all $m \geq N$, $2Mb_m < \epsilon$.

Therefore, $\sum_{k=1}^{\infty} a_k b_k$ is Cauchy and converges. ■

Now we formally establish a claim we mentioned earlier.

Proposition 1.7 *Suppose that $\{a_k\}_{k=1}^{\infty}$ is a sequence of real numbers whose partial sums $s_n = \sum_{k=1}^{n} a_k$ are bounded. Then*

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

converges.

This follows directly from Dirichlet's test since $\{\frac{1}{k}\}_{k=1}^{\infty}$ decreases monotonically and converges to 0.

As another application, we can apply this test to a sequence of coefficients determined by revolutions about the unit circle. Let $z \in \mathbb{C}$, such that $|z| = 1$ but $z \neq 1$. Let $a_k = z^k$.

Summing the geometric series we have, for all n

$$\left|\sum_{k=1}^{n} a_k\right| = \left|\sum_{k=1}^{n} z^k\right| = \left|\frac{1 - z^{n+1}}{1 - z}\right| \leq \left|\frac{2}{1 - z}\right|.$$

Thus, by Dirichlet's Test,

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

Observe that if $z = -1$, $\sum_{k=1}^{\infty} \frac{a_k}{k}$ is the alternating harmonic series.

The following test will be useful in some of our later work.

Theorem 1.8 (Abel's test) *Suppose that $\sum_{k=1}^{\infty} a_k$ converges and that $\{b_k\}_{k=1}^{\infty}$ is a monotone and convergent sequence. Then the series $\sum_{k=1}^{\infty} a_k b_k$ converges.*

Proof: For each $k \in \mathbb{N}$, we let $c_k = b - b_k$, so that $\{c_k\}_{k=1}^{\infty}$ is a monotonic sequence converging to 0. Since $\sum_{k=1}^{\infty} a_k$ converges, its sequence of partial sums is bounded. Therefore, by Dirichlet's test, $\sum_{k=1}^{\infty} a_k c_k$ converges. Of course, the series $\sum_{k=1}^{\infty} a_k b$ also converges. But for each k,

$$a_k b_k = a_k b - a_k(b - b_k) = a_k b - a_k b_k.$$

So $\sum_{k=1}^{\infty} a_k b_k$ is the term-by-term difference of two convergent series, and so must converge. ■

Proposition 1.9 *Suppose that the sequences $\{a_k\}_{k=1}^{\infty}$ and $\{b_k\}_{k=1}^{\infty}$ are given such that $\sum_{k=1}^{\infty} a_k b_k$ converges and $\{\frac{b_{k-1}}{b_k}\}_{k=1}^{\infty}$ is monotonic and converges. Then the series*

$$\sum_{k=1}^{\infty} a_k b_{k-1}$$

converges.

This is an immediate application of Abel's test, since

$$\sum_{k=1}^{\infty} a_k b_{k-1} = \sum_{k=1}^{\infty} a_k b_k \left(\frac{b_{k-1}}{b_k}\right).$$

For example, if $\{b_k\}_{k=1}^{\infty}$ is the harmonic sequence: $b_k = \frac{1}{k}$, then $\{\frac{b_{k-1}}{b_k}\}_{k=1}^{\infty}$ converges monotonically to 1, so that whenever a series $\sum_{k=1}^{\infty} \frac{a_k}{k}$ converges, so does the series $\sum_{k=1}^{\infty} \frac{a_{k+1}}{k}$, which is obtained by shifting the sequence of coefficients one place to the left.

This observation will be used in later arguments.

What if the coefficient sequence $\{a_k\}_{k=1}^{\infty}$ are assigned randomly? What is the probability that the resulting random series will converge?

In order to make sense of this question and attempt to answer it, we must transition to the language of probability theory.

2 Random Harmonic Series With Independent Processes

The fundamental concepts of probability theory are expressed in measure-theoretic language. However, it is conventional to use language that is specific to the probabilistic point of view.

Definition 2.1 *Let $(\Omega, \mathcal{A}, P)$ be a measure space defined on the set Ω, σ-algebra $\mathcal{A}$, and measure P. $(\Omega, \mathcal{A}, P)$ is a* ***probability space****, and P is called a probability measure, if $P(\Omega) = 1$.*

Definition 2.2 *An* ***event*** *is an element of $\mathcal{A}$, i.e., a measurable set. The probability of event A is $P[A]$.*

Definition 2.3 *A* ***random variable*** *is a map $X : \Omega \to \mathbb{R}$ such that for all Borel sets B, $X^{-1}(B) \in \mathcal{A}$, i.e., a measurable function. We will often use the abbreviation r.v. for the term random variable.*

Given a random variable X on a probability space Ω, it is customary to abbreviate expressions like $\{\omega \in \Omega : X(\omega) > a\}$ by the more succinct $[X > a]$.

Definition 2.4 *The* ***expected value*** *of a random variable X is defined as the integral $\int_{\Omega} X dP$. We will simplify this notation to either $\int X$, or $E(X)$, which will be used interchangeably.*

Now we return to our discussion of randomly assigning coefficients 1 and -1 to the terms of the harmonic series. We formalize this by supposing that $\{\varepsilon_k\}_{k=1}^{\infty}$ is a sequence of random variables (a ***process***) on a probability space $(\Omega, \mathcal{A}, P)$, with which we form the series

$$\sum_{k=1}^{\infty} \varepsilon_k(\omega) \frac{1}{k}.$$

Given a process $\{\varepsilon_k\}_{k=1}^{\infty}$ one naturally asks about the probability with which this series converges. That is, we ask for the value

$$P\left[\sum_{k=1}^{\infty}\varepsilon_k(\omega)\frac{1}{k}\text{ converges}\right].$$

Perhaps the most natural choice for the sequence ε_k would be to simulate the flipping of a fair coin, assigning 1 or -1 as the coefficient of the term $\frac{1}{k}$ depending on the outcome of the k^{th} flip. In other words, the random variables would be chosen to take the values ± 1 independently and with

$$P[\varepsilon_k = 1] = P[\varepsilon_k = -1] = \frac{1}{2}.$$

We will now make precise this central notion of independence, and develop some of its essential properties. Once we have done this, our goal is to then to prove the Kolmogorov three series criterion, which will tell us (among other things) that the above series converges with probability 1.

Our exposition will closely follow chapter V, sections 16 and 17 of Probability Theory, written by Loève, item 3 in the references.

2.1 Sets and Classes

We assume that each r.v. is defined on a fixed probability space $(\Omega, \mathcal{A}, P)$.

Definition 2.5 *Let T be an arbitrary index set. A collection of events $\{A_t\}_{t\in T} \subset \mathcal{A}$ is said to be **independent** if, for every finite subset $\{t_1, t_2, ..., t_n\} \in T$*

$$P\left(\bigcap_{k=1}^{n} A_{t_k}\right) = \prod_{k=1}^{n} P(A_{t_k}).$$

Similarly, we can carry this definition to families of classes of events $\{\mathcal{C}_t\}_{t\in T}$.

Families of classes $\{\mathcal{C}_t\}_{t\in T} \subset 2^{\mathcal{A}}$, the power set of $\mathcal{A}$, are said to be independent if every collection $\{A_t\}_{t\in T}$, where each A_t is an arbitrarily selected event in $\mathcal{C}_t$, is independent.

Because subclasses of classes are also classes, it is clear subclasses of independent classes are independent.

Definition 2.6 *The **minimal σ-field** over a class $\mathcal{C}_t$ is the intersection of all σ-algebras that contain $\mathcal{C}_t$. In other words, the minimal σ-field is the σ-algebra generated by the events in $\mathcal{C}_t$.*

Theorem 2.7 (Extension theorem) *Minimal σ-fields over independent classes $\{C_t\}_{t\in T}$ that are closed under finite intersections are independent.*

The rest of this section will be dedicated to establishing the proof. We begin by defining Dynkin classes and π-classes.

Definition 2.8 *Let Ω be a set, and $\mathcal{D}$ a collection of subsets of Ω. Then $\mathcal{D}$ is called a **Dynkin class** (in Ω) if it has the properties:*

1. $\Omega \in \mathcal{D}$,
2. *Whenever $A, B \in \mathcal{D}$ and $A \supset B$, we have $A - B \in \mathcal{D}$, and*
3. *Whenever $\{A_i\}_{i=1}^{\infty}$ is an increasing sequence in $\mathcal{D}$, we have $\bigcup_{i=1}^{\infty} A_i \in \mathcal{D}$.*

Given any collection $\mathcal{C}$ of sets in Ω, we can form the intersection of all the Dynkin classes in Ω containing $\mathcal{C}$. This is then the smallest Dynkin class in Ω containing $\mathcal{C}$. That is, it has the property that it is contained in every Dynkin class that contains $\mathcal{C}$. We refer to this as the Dynkin class generated by $\mathcal{C}$, denoted by $\delta(\mathcal{C})$.

We will use the notation $\sigma(\mathcal{C})$ to denote the σ-algebra generated by $\mathcal{C}$.

Definition 2.9 *Let Ω be a set, and $\mathcal{C}$ a collection of subsets of Ω. Then $\mathcal{C}$ is called a π-**class** if it is closed under taking finite intersections.*

Now suppose we have a class of sets $\mathcal{C}$ in Ω. We can compare $\sigma(\mathcal{C})$ to $\delta(\mathcal{C})$. But every σ-algebra is a Dynkin class. In particular, $\sigma(\mathcal{C})$, being a σ-algebra, is a Dynkin class containing $\mathcal{C}$, and so contains the smallest Dynkin class containing $\mathcal{C}$. That is, $\sigma(\mathcal{C}) \supset \delta(\mathcal{C})$. The reverse containment holds if $\mathcal{C}$ is a π-class.

Lemma 2.10 *Let Ω be a set, and $\mathcal{C}$ a π-class in Ω. Then $\delta(\mathcal{C}) = \sigma(\mathcal{C})$.*

Proof: We want to show that $\delta(\mathcal{C}) \supset \sigma(\mathcal{C})$. By the minimality of $\sigma(\mathcal{C})$, it is enough to show that $\delta(\mathcal{C})$ is a σ-algebra. We begin by showing that $\delta(\mathcal{C})$ is closed under taking intersections.

First, we first define an auxiliary class $\mathcal{D}_1 \subset \delta(\mathcal{C})$ by setting

$$\mathcal{D}_1 = \{A \in \delta(\mathcal{C}) : (\forall C \in \mathcal{C})\, A \cap C \in \delta(\mathcal{C})\}.$$

One easily verifies that this class $\mathcal{D}_1$ is a Dynkin class containing $\mathcal{C}$. Then, by the minimality of $\delta(\mathcal{C})$, we get $\mathcal{D}_1 \supset \delta(\mathcal{C})$ and so $\mathcal{D}_1 = \delta(\mathcal{C})$. In other words, we now know that for all $A \in \delta(\mathcal{C})$ and all $C \in \mathcal{C}$, $A \cap C \in \delta(\mathcal{C})$.

Second we define the class

$$\mathcal{D}_2 = \{B \in \delta(\mathcal{C}) : (\forall A \in \delta(\mathcal{C}))\, A \cap B \in \delta(\mathcal{C})\}.$$

Note that A ranges over $\delta(\mathcal{C})$, not just $\mathcal{C}$ as before.

We've just shown that $\mathcal{D}_2 \supset \mathcal{C}$. Arguing as before, we verify that $\mathcal{D}_2$ is a Dynkin class. Therefore, by the minimality of $\delta(\mathcal{C})$, $\mathcal{D}_2 \supset \delta(\mathcal{C})$, and so $\mathcal{D}_2 = \delta(\mathcal{C})$. That is, $\delta(\mathcal{C})$ is closed under intersection.

But now we can see that $\delta(\mathcal{C})$ is an algebra, for conditions 1 and 2 of the definition of Dynkin class imply every Dynkin class is closed under complement. And using the above we see that $\delta(\mathcal{C})$ is closed under union. But then it is an algebra closed under increasing unions, and that implies it is a σ-algebra. ■

Now, we turn our attention to independent classes.

Lemma 2.11 *Let $(\Omega, \mathcal{A}, P)$ be a probability space, and suppose $\{\mathcal{B}_t\}_{t\in T}$ is a collection of independent classes in Ω, where each $\mathcal{B}_t$ is a π class. Let S_1 and S_2 be disjoint, non-empty subsets of T. Then $\sigma\left(\bigcup_{t\in S_1} \mathcal{B}_t\right)$ is independent of $\sigma\left(\bigcup_{t\in S_2} \mathcal{B}_t\right)$.*

Proof: Let $\mathcal{C}_1$ be the set of all finite intersections of elements of $\bigcup_{t\in S_1} \mathcal{B}_t$. Since the $\mathcal{B}_t$ are π-classes, each element of $\mathcal{C}_1$ has the form $\bigcap_{i=1}^n B_{t_i}$, where the indices $t_i \in S_1$ are distinct, and for each i, $B_{t_i} \in \mathcal{B}_{t_i}$.

We note that for each $t \in S_1$, $\mathcal{C}_1 \supset \mathcal{B}_t$, and so $\sigma(\mathcal{C}_1) = \sigma\left(\bigcup_{t\in S_1} \mathcal{B}_t\right)$. We repeat this definition and these comments for $\mathcal{C}_2$ in a corresponding way.

Note that $\mathcal{C}_1$ and $\mathcal{C}_2$ are π-classes. Here we make use of the fact that the $\mathcal{B}_t$ are π-classes.

Now let $\mathcal{D}_1 = \{A \in \sigma(\mathcal{C}_1) : (\forall C \in \mathcal{C}_2)\, \{A, C\} \text{ are independent}\}$. In other words, $\mathcal{D}_1 = \{A \in \sigma(\mathcal{C}_1) : (\forall C \in \mathcal{C}_2)\, P(A \cap C) = P(A)\, P(C)\}$.

We observe that $\mathcal{D}_1$ is a Dynkin class containing $\mathcal{C}_1$. Therefore, by our theorem above, $\mathcal{D}_1 = \sigma(\mathcal{C}_1)$. Thus, every element of $\sigma(\mathcal{C}_1)$ is independent of every element of $\mathcal{C}_2$. Now we repeat this sort of argument with $\mathcal{C}_2$. That is, we define

$$\mathcal{D}_2 = \{B \in \sigma(\mathcal{C}_2) : (\forall A \in \sigma(\mathcal{C}_1))\, \{A, B\} \text{ are independent.}\}$$

Note that A ranges over $\sigma(\mathcal{C}_1)$, not just $\mathcal{C}_1$. Again we verify that $\mathcal{D}_2$ is a Dynkin class containing $\mathcal{C}_2$, and so $\mathcal{D}_2 = \sigma(\mathcal{C}_2)$. Thus every element of $\sigma(\mathcal{C}_2)$ is independent of every element of $\sigma(\mathcal{C}_1)$. ■

Lastly, we establish the following lemma, which will conclude the proof of the extension theorem, Theorem 2.7.

Lemma 2.12 *Let $(\Omega, \mathcal{A}, P)$ be a probability space, and suppose $\{\mathcal{B}_t\}_{t\in T}$ is a collection of independent classes in Ω, where each $\mathcal{B}_t$ is a π-class. Let $\{S_r\}_{r\in R}$ be a pairwise disjoint collection of non-empty subsets of T. Then the σ-algebras $\left\{\sigma\left(\bigcup_{t\in S_r}\mathcal{B}_t\right)\right\}_{r\in R}$ are independent.*

Proof: We only need to prove that for every finite set $\{r_1, r_2, ..., r_n\} \subset R$, the $\sigma-$algebras $\left\{\sigma\left(\bigcup_{t\in S_{r_i}}\mathcal{B}_t\right)\right\}_{i=1}^{n}$ are independent. The previous lemma established that this is the case when $n = 2$.

We proceed by induction. Suppose we know that this is true for some $n \geq 2$. Now choose $n+1$ elements of R, say $\{r_1, r_2, ..., r_{n+1}\}$. We want to show that the $\sigma-$algebras $\left\{\sigma\left(\bigcup_{t\in S_{r_i}}\mathcal{B}_t\right)\right\}_{i=1}^{n+1}$ are independent.

Let $\mathcal{C}_{n+1}$ be the class consisting of all the finite intersections of elements of $\bigcup_{t\in S_{r_{n+1}}}\mathcal{B}_t$. This is a $\pi-$class containing $\bigcup_{t\in S_{r_{n+1}}}\mathcal{B}_t$. Each element of $\mathcal{C}_{n+1}$ has the form $\bigcap_{i=1}^{k} B_{t_i}$, where the indeces $t_i \in S_{n+1}$ are distinct, and for each i, $B_{t_i} \in \mathcal{B}_{t_i}$.

Let $\mathcal{C}'$ be the class consisting of all the finite intersections of elements chosen from the set $\left(\bigcup_{t\in\cup_{j=1}^{n}S_{r_j}}\mathcal{B}_t\right)$. That is, $\mathcal{C}'$ consists of sets of the form $\bigcap_{i=1}^{k} B_{t_i}$ where the indeces $t_i \in \cup_{j=1}^{n} S_{r_j}$ are distinct, and for each i, $B_{t_i} \in \mathcal{B}_{t_i}$. So $\mathcal{C}'$ is a $\pi-$ class containing $\bigcup_{t\in\cup_{j=1}^{n}S_j}\mathcal{B}_t$.

We know by assumption that $\mathcal{C}'$ and $\mathcal{C}_{n+1}$ are independent. So by the previous theorem, $\sigma(\mathcal{C}')$ and $\sigma(\mathcal{C}_{n+1})$ are independent.

Now for $j = 1, 2, ..., n+1$, choose sets $A_j \in \sigma\left(\bigcup_{t\in S_{r_j}}\mathcal{B}_t\right)$. We want to prove that $P\left(\bigcap_{j=1}^{n+1} A_j\right) = \prod_{j=1}^{n+1} P(A_j)$.

But since $\sigma(\mathcal{C}')$ and $\sigma(\mathcal{C}_{n+1})$ are independent (And $\sigma(\mathcal{C}')$ contains each $\sigma\left(\bigcup_{t\in S_{r_j}}\mathcal{B}_t\right)$, for $j = 1, 2, ..., n$), we know
$P\left(\bigcap_{j=1}^{n+1} A_j\right) = P\left(A_{n+1} \cap \bigcap_{j=1}^{n} A_j\right) = P(A_{n+1})\, P\left(\bigcap_{j=1}^{n} A_j\right)$.

Thus, by our inductive hypothesis, $P(A_{n+1})\, P\left(\bigcap_{j=1}^{n} A_j\right) = \prod_{j=1}^{n+1} P(A_j)$. ■

2.2 Multiplicative Properties

Definition 2.13 *A collection of r.v.'s* $\{X_t\}_{t\in T}$ *is said to be* ***independent*** *if for every finite class* $(S_{t_1},...,S_{t_n})$ *of Borel sets in* $\mathbb{R}$,

$$P\left[\bigcap_{k=1}^{n}[X_{t_k}\in S_{t_k}]\right]=\prod_{k=1}^{n}P[X_{t_k}\in S_{t_k}].$$

In other words, the classes are independent.

In order to prove the fundamental expectation property of independent r.v.'s, we need to prove the following lemma.

Lemma 2.14 (Multiplication lemma) *Let* $X_1,X_2,...,X_n$ *be independent nonnegative r.v.'s. Then the expectation of their product is the product of their expectations:*

$$E\left(X_1X_2...X_k\right)=\prod_{k=1}^{n}E\left(X_k\right).$$

Proof: We will prove the assertion for two independent r.v.'s X and Y, for the general case follows naturally by induction.

Let $X=\sum_{i=1}^{\infty}x_iI_{A_i}$ and $Y=\sum_{j=1}^{\infty}y_jI_{B_j}$ be nonnegative simple r.v.s, with $A_i=[X=x_j]$, $B_j=[Y=y_j]$. Because X,Y independent, we have that $P(A_iB_j)=P(A_i)P(B_j)$.

$E(XY)=\sum_{i=1,j=1}^{\infty}x_iy_jP(A_i)P(B_j)=\sum_{i=1}^{\infty}x_iP(A_i)\cdot\sum_{j=1}^{\infty}y_jPB_j=E(X)E(Y)$.

Now, let X, Y be independent nonnegative r.v.'s and let $A_{n,i}=[\frac{i-1}{2^n}\leq X<\frac{i}{2^n}]$ $B_{n,j}=[\frac{j-1}{2^n}\leq Y<\frac{j}{2^n}]$.

Now, set $X_n=\sum_{i=1}^{n2^n}\frac{i-1}{2^n}I_{A_{n,i}}$ and $Y_n=\sum_{j=1}^{n2^n}\frac{j-1}{2^n}I_{B_{n,j}}$.

Then, $0\leq\lim_{n\to\infty}X_n=X$, $0\leq\lim_{n\to\infty}Y_n=Y$, and $0\leq\lim_{n\to\infty}X_nY_n=XY$.

We already have that $E(X_nY_n)=E(X_n)E(Y_n)$. So, the by the monotone convergence theorem, $E(XY)=E(X)E(Y)$. ■

Theorem 2.15 (Multiplication theorem) *Let* $X_1,X_2,...,X_n$ *be independent r.v.'s. If these r.v.'s are integrable, so is their product and* $E(X_1X_2...X_n)=\prod_{k=1}^{n}E(X_k)$. *Conversely, if their product is integrable and none is 0, then they are integrable.*

Proof: It once again suffices to prove the assertion for two independent r.v.'s. X and Y. Because X and Y are independent, we know that the nonnegative r.v.'s $X'=X^+$ or X^- or $|X|$ and $Y'=Y^+$ or Y^- or $|Y|$ are independent.

So, by the previous lemma, $E(X'Y') = E(X')E(Y')$.

If X, Y integrable, then so must be X', Y' by same reasons as above. Therefore we have that

$$\begin{aligned} E(XY) &= E((X^+ + X^-)(Y^+ + Y^-)) \\ &= E(X^+)E(Y^+) + E(X^+)E(Y^-) + E(X^-)E(Y^+) + E(X^-)E(Y^-) \\ &= (E(X^+) + E(X^-))(E(Y^+) + E(Y^-)) \\ &= E(X)E(Y). \end{aligned}$$

Conversely, if XY is integrable with $E(|XY|) = E(|X|)E(|Y|) < \infty$, and neither X nor Y is 0, then $E(|X|), E(|Y|) \neq 0$, then $E(|X|), E(|Y|) < \infty$. ■

2.3 Sequences of Independent Events

Definition 2.16 *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of events. Then the $\lim_{n\to\infty} \sup A_n$ is defined to be $\bigcap_{k=1}^{\infty}(\bigcup_{n=k}^{\infty} A_n)$. Similarly, $\lim_{n\to\infty} \inf A_n$ is defined to be $\bigcup_{k=1}^{\infty}(\bigcap_{n=k}^{\infty} A_n)$.*

It follows that,

$$\lim_{n\to\infty} \sup A_n = \lim_{k\to\infty} P(\bigcup_{n=k}^{\infty} A_n)$$

and

$$\lim_{n\to\infty} \inf A_n = \lim_{k\to\infty} P(\bigcap_{n=k}^{\infty} A_n).$$

Theorem 2.17 (Borel zero-one criterion) *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of independent events. If $\sum_{n=1}^{\infty} P[A_n] < \infty$, then $P[\lim_{n\to\infty} \sup A_n] = 0$, and if $\sum_{n=1}^{\infty} P[A_n] = \infty$, then $P[\lim_{n\to\infty} \sup A_n] = 1$.*

Proof: Suppose that $\sum_{n=1}^{\infty} P[A_n]$ converges. Then, we have that

$$P[\lim_{n\to\infty} \sup A_n] = \lim_{k\to\infty} P[\bigcup_{n=k}^{\infty} A_n] \leq \lim_{k\to\infty} \sum_{n=k}^{\infty} P[A_n] = 0.$$

This first part of the theorem is known as the Borel-Cantelli lemma. Notice that it does not require independence of the collection of r.v.'s. In fact, we will use this lemma later in the thesis when we are dealing with a collection of more general r.v.'s.

Conversely, suppose that $\sum_{n=1}^{\infty} P[A_n]$ diverges. We again have that

$$P[\lim_{n\to\infty} \sup A_n] = \lim_{k\to\infty} P[\bigcup_{n=k}^{\infty} A_n].$$

By De Morgan's Laws,

$$P(\bigcup_{n=k}^{\infty} A_n) = 1 - P[\bigcap_{n=k}^{\infty} A_n^c],$$

and by independence

$$1 - P[\bigcap_{n=k}^{\infty} A_n^c] = 1 - \prod_{n=k}^{\infty}(1 - P[A_n]).$$

So,

$$P[\lim_{n\to\infty} \sup A_n] = \lim_{k\to\infty}(1 - \prod_{n=k}^{\infty}(1 - P[A_n])).$$

Using the fact that, $e^{-x} \geq 1 - x$ for all $x \in \mathbb{R}$, we have that

$$e^{-\sum_{n=k}^{\infty} P[A_n]} \geq \prod_{n=k}^{\infty}(1 - P[A_n]).$$

Therefore we have that

$$1 - \lim_{k\to\infty} e^{-\sum_{n=k}^{\infty} P[A_n]} \leq 1 - \lim_{k\to\infty} \prod_{n=k}^{\infty}(1 - P[A_n]) = P[\lim_{n\to\infty} \sup A_n].$$

Since, $\sum_{n=k}^{\infty} P[A_n] \to \infty$, the inequality above gives us that $1 \leq P[\lim_{n\to\infty} \sup A_n]$. Therefore, $P[\lim_{n\to\infty} \sup A_n] = 1$. ■

Definition 2.18 *We say that* $\lim_{n\to\infty} A_n = A$ *if* $\lim_{n\to\infty} \sup A_n = \lim_{n\to\infty} \inf A_n = A$.

Corollary 2.19 *Suppose that* $\{A_n\}_{n=1}^{\infty}$ *is a sequence of independent events such that* $A_n \to A$. *Then* $P[A] = 0$ *or* $P[A] = 1$.

This is a direct consequence of the theorem.

Corollary 2.20 *Suppose that* $\{X_n\}_{n=1}^{\infty}$ *is an independent sequence of random variables such that* $X_n \to 0$ *a.s. Then for all finite values* $c > 0$, $\sum_{n=1}^{\infty} P[|X_n| \geq c]$ *converges.*

Proof: Fix $c > 0$ finite. Let $A_n = [|X_n| \geq c]$. Then because $X_n \to 0$ *a.s.*, $P[\lim_{n\to\infty} \sup A_n] = 0$. By the previous theorem, $\sum_{n=1}^{\infty} P[A_n] < \infty$. ■

Definition 2.21 *Let* X *be a complex r.v. We define* $\mathbb{B}(X)$ *to be the* σ*-field generated by the pre images of the Borel sets in* $\mathbb{C}$. *We define the* ***tail*** $\boldsymbol{\sigma}$***-field*** *of a sequence* $\{X_n\}_{n=1}^{\infty}$, *to be* $\lim_{n\to\infty} \sup \mathbb{B}(X_n, X_{n+1}, X_{n+2}, ...) = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} \mathbb{B}(X_n, X_{n+1}, X_{n+2}, ...)$.

More explicitly, the sequence $\mathbb{B}(X_n, X_{n+1}, X_{n+2}, ...), \mathbb{B}(X_{n+1}, X_{n+2}, X_{n+3}...), ...$ *is a non-increasing sequence of* σ*-fields and its limit, or intersection, is a* σ*-field contained in* $\mathbb{B}(X_n, X_{n+1}, X_{n+2}, ...)$ *for all* $n \in \mathbb{N}$*. Elements of the tail* σ*-field are naturally called* ***tail events****. Measurable functions which induce sub* σ*-fields of events contained in the tail* σ*-field are called* ***tail functions****.*

Theorem 2.22 (Kolmogorov's zero-one law) *For a sequence of independent r.v.'s, the tail events have probability* 0 *or* 1*, and the tail functions are degenerate.*

Proof: We begin with the simple observation that if A is an event that is independent of itself, the $P(A) = P(A \cap A) = P(A)P(A) = P(A)^2$, which implies $P(A) = 0$ or 1.

Now, let σ be the tail σ-field of a sequence of independent r.v.'s $\{X_n\}_{n=1}^{\infty}$, and let $S \in \sigma$. Then by the definition of tail σ-field, for all $n \in \mathbb{N}$, $\sigma \subset \mathbb{B}(X_n, X_{n+1}, ...)$, and thus $S \in \mathbb{B}(X_n, X_{n+1}, ...)$.

But for all $n \in \mathbb{N}$, $S \in \mathbb{B}\left(X_{n+1}, X_{n+1}, ...\right)$ which is independent of $\mathbb{B}\left(X_1, X_2, ..., X_n\right)$. Therefore by the extension theorem S is independent of $\mathbb{B}\left(X_1, X_2, X_3...\right)$. But $S \in \mathbb{B}\left(X_1, X_2, X_3...\right)$, so that S is independent of itself. Therefore S has probability 0 or 1, which proves the first assertion.

The second follows, since if X is a tail function, then it is $\{\emptyset, \Omega\}$-measurable *a.s.*, and thus degenerates. ■

Corollary 2.23 *If* X_n *are independent r.v.'s, then the sequence* X_n*, and series* $\sum_{n=1}^{\infty} X_n$ *either converge or diverge a.s. Moreover, the limits of the sequences* X_n *and* $\frac{X_1+X_2+...+X_n}{b_n}$*, where* $b_n \to \infty$*, are degenerate.*

Proof: By the previous theorem, all we must verify is that the set on which the sequence X_n and the sum $\sum_{n=1}^{\infty} X_n$ converge is a tail event.

Observe that $\{X_n(\omega)\}_{n=1}^{\infty}$ and $\sum_{n=1}^{\infty} X_n(\omega)$ converge if and only if $\{X_n(\omega)\}_{n=i}^{\infty}$ and $\sum_{n=i}^{\infty} X_n(\omega)$ respectively converge for all $i \in \mathbb{N}$. Therefore, the set of $\omega \in \Omega$ on which the sequence and sum converge is a tail event.

Similarly, the set on which $\frac{X_1+X_2+...+X_n}{b_n}$ converges is a tail event since $\frac{X_1+X_2+...+X_n}{b_n} = \frac{X_1+X_2+...+X_i}{b_n} + \frac{X_i+X_2+...+X_n}{b_n}$. ■

2.4 Series of Independent Random Variables

A r.v. is **centered** at c if we replace X by $X-c$. If X is integrable, or $E(|X|)<\infty$, we can center X at its expectation $E(X)$, replacing X with $X-E(X)$. In other words, a r.v. is centered at its expectation if, and only if, its expectation exists and equals 0.

Definition 2.24 *Let X be integrable. The **variance** of a r.v., $E(X-EX)^2 = EX^2-(EX)^2$, is denoted $\sigma^2 X$.*

Observe that for every c, $\sigma^2(X-c) = E(X-c-E(X-c))^2 = E(X-c-EX+Ec)^2 = E(X-EX)^2$. Therefore, centerings do not change variances.

It is obvious X must be integrable in order to define its variance. However, even then $(X-EX)^2$ need not be integrable, which allows the variance of an integrable r.v. to be infinite. We handle this problem with truncation.

Fix $c>0$ finite. Let X be a r.v. We say that X^c is X **truncated** at $c>0$ finite, where

$$X^c = \begin{cases} X, & \text{if } |X|<c. \\ 0, & \text{if } |X|\geq c. \end{cases}$$

It follows that if F is the distribution function of X, then

$$E(X^c) = \int_{|X|<c} X dF, \text{ and } E(X^c)^2 = \int_{|X|<c} X^2 dF$$

exist and are finite.

We can always select c large enough so that $P[X\neq X^c] = P[|X|\geq c]$ is arbitrarily small. Furthermore, given any sequence of r.v.'s $\{X_k\}_{k=1}^{\infty}$, and given $\epsilon>0$, we can select a sufficiently large sequence $\{c_k\}_{k=1}^{\infty}$ so as to make

$$P[\bigcup_{k=1}^{\infty}[X_k\neq X_k^{c_k}]] \leq \sum_{k=1}^{\infty} P[|X_k|\geq c_k] \leq \sum_{k=1}^{\infty}\frac{\epsilon}{2^k} = \epsilon.$$

Therefore, to every countable family of r.v.'s we can construct a family of bounded r.v.'s which differs from the first family on an event of arbitrarily small probability.

Definition 2.25 *Let the two sequences X_n and X'_n of r.v.'s be called **tail-equivalent** if they differ a.s. only by a finite number of terms; in other words, if for a.e. $\omega\in\Omega$, there is a number $n(\omega)$ such that for $n\geq n(\omega)$ the two sequences $X_n(\omega)$ and $X'_n(\omega)$ are the same.*

Further, we abbreviate the phrase *infinitely often* by *i.o.*. Two sequences are **tail-equivalent** if $P[X_n \neq X'_n \text{ i.o.}] = 0$. If the sequences X_n and X'_n converge on the same event up to some null subset, then we say that they are **convergence equivalent**.

Let $S_n = \sum_{k=1}^n X_k$ and $S'_n = \sum_{k=1}^n X'_k$.

Lemma 2.26 (Equivalence lemma) *If the series $\sum_{k=1}^\infty P[X_n \neq X'_n]$ converges, then the sequences X_n and X'_n are tail-equivalent. Therefore the series $\sum_{k=1}^\infty X_n$ and $\sum_{k=1}^\infty X'_n$ are convergence-equivalent, and the sequence $\frac{S_n}{b_n}$ and $\frac{S'_n}{b_n}$ where $b_n \to \infty$, converge on the same event and to the same limit, up to a null event.*

Proof: Suppose that $\sum_{k=1}^\infty P[X_n \neq X'_n] < \infty$. Then by Borel-Cantelli, theorem 2.17, $P[X_n \neq X'_n \text{ i.o.}] = 0$ and so X_n, and X'_n are tail-equivalent.

It immediately follows that $\sum_{k=1}^\infty X_n$ and $\sum_{k=1}^\infty X'_n$, and thus $\frac{S_n}{b_n}$ and $\frac{S'_n}{b_n}$, where $b_n \to \infty$ converge on the same event up to some null subset, and are thus convergence-equivalent. ■

For the rest of the chapter, we will abide by the convention $S_0 = 0$, and $S_n = \sum_{k=1}^n X_k$ for $n = 1, 2, 3, ...$, and the summands $X_1, X_2, ...$ are independent r.v.'s.

Now, let $\{X_i\}_{i=1}^\infty$ be integrable. Since centering r.v.'s at do not modify the variances, we can assume, when computing variances, that these r.v.'s are centered at expectations.

Theorem 2.27 (Bienaymé equality) $\sigma^2 S_n = \sum_{k=1}^n \sigma^2 X_k$.

Proof: We know

$$\sigma^2 S_n = ES_n^2 = \sum_{k=1}^n E(X_k^2) + \sum_{j,k=1, j\neq k}^n E(X_j X_k)$$

and because we are assuming that each r.v. is centered at its expectation, $E(X_j X_k) = E(X_j)E(X_k) = 0$. Thus,

$$\sum_{k=1}^n E(X_k^2) + \sum_{j,k=1, j\neq k}^n E(X_j)E(X_k) = \sum_{k=1}^n E(X_k^2) = \sum_{k=1}^n \sigma^2 X_k. \blacksquare$$

Theorem 2.28 (Basic inequality) *Let X be an arbitrary r.v. and let g be a non-negative Borel function on $\mathbb{R}$. For each a, define $A = [|X| \geq a]$. If g is even and nondecreasing on $[0, +\infty)$ then, for every $a \geq 0$,*

$$\frac{Eg(X) - g(a)}{a.s. \sup g(X)} \leq P(A) \leq \frac{Eg(X)}{g(a)}.$$

Proof: Since g is a Borel function on $\mathbb{R}$, it follows that $g(X)$ is a measurable function on Ω and, since g is nonnegative on $\mathbb{R}$, its integral exists. If g is even and is nondecreasing on $[0, \infty)$, then

$$\int_A g(X) + \int_{A^c} g(X)$$

and

$$g(a)P(A) \leq \int_A g(X) \leq a.s. \sup g(X)P(A)$$

and

$$0 \leq \int_{A_c} g(X) \leq g(a)P(A^c) \leq g(a).$$

Thus,

$$g(a)P(A) \leq Eg(X) \leq a.s. \sup g(X)P(A) + g(a).$$

So after putting together the right hand side and left hand side of this inequality,

$$\frac{Eg(X) - g(a)}{a.s. \sup g(X)} \leq P(A) \leq \frac{Eg(X)}{g(a)}$$

which concludes the proof. ■

Now, let $X = S_n - E(S_n)$, $g(t) = t^2$, and $a = \epsilon$. Then, $A = [|X| \geq \epsilon] = [|S_n - E(S_n)| \geq \epsilon]$.

The basic inequality gives us that

$$\frac{E(S_n - E(S_n))^2 - g(\epsilon)}{a.s. \sup(S_n - E(S_n))^2} \leq P(A) \leq \frac{E(S_n - E(S_n))^2}{g(\epsilon)},$$

which says that

$$\frac{\sigma^2(S_n) - \epsilon^2}{a.s. \sup(S_n - E(S_n))^2} \leq P(A) \leq \frac{\sigma^2(S_n)}{\epsilon^2}.$$

Thus,

$$\frac{\sum_{k=1}^n \sigma^2(X_k) - \epsilon^2}{a.s. \sup(S_n - E(S_n))^2} \leq P(A) \leq \frac{\sum_{k=1}^n \sigma^2(X_k)}{\epsilon^2},$$

which establishes the following inequality,

$$\frac{\sum_{k=1}^n \sigma^2(X_k) - \epsilon^2}{a.s. \sup(S_n - E(S_n))^2} \leq P[|S_n - E(S_n)| \geq \epsilon] \leq \frac{\sum_{k=1}^n \sigma^2(X_k)}{\epsilon^2}.$$

The right hand side inequality is known as the Bienaymé-Tchebichev Inequaility. We will use it to prove the following lemma.

Lemma 2.29 *If the series* $\sum_{k=1}^{\infty} \sigma^2 X_k$ *converges, then the series* $\sum_{k=1}^{\infty}(X_k - EX_k)$ *converges in probability. In particular, if* $\frac{\sum_{k=1}^{n} \sigma^2 X_k}{{b_n}^2}$ *converges to* 0*, then* $\frac{S_n - ES_n}{b_n}$ *converges to* 0 *in probability.*

Proof: The Bienaymé-Tchebichev inequality tells us that for all n and k, if we set $S_{n,k} = \sum_{i=n+1}^{n+k} X_i$, we get, for all $\varepsilon > 0$,

$$P\left(|S_{n,k} - E\left(S_{n,k}\right)| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{i=n+1}^{n+k} \sigma^2\left(X_i\right).$$

But we are given that $\lim_{n\to\infty} \sum_{i=n+1}^{\infty} \sigma^2\left(X_i\right) = 0$, so for fixed $\varepsilon > 0$ and $\delta > 0$, there is an N so that for all $n \geq N$ and all $k \geq 1$,

$$\frac{1}{\varepsilon^2} \sum_{i=n+1}^{n+k} \sigma^2\left(X_i\right) < \delta.$$

Since

$$S_{n,k} - E\left(S_{n,k}\right) = \sum_{i=n+1}^{n+k} \left(X_i - E\left(X_i\right)\right)$$

this says that the partial sums of the series

$$\sum_{i=1}^{\infty} \left(X_i - E\left(X_i\right)\right)$$

are Cauchy in probability. Therefore, the partial sums (and hence the series) must converge in probability. Now suppose $b_n \geq 0$ and

$$\frac{1}{b_n^2} \sum_{k=1}^{n} \sigma^2\left(X_k\right) \to 0.$$

If we replace ε by εb_n in the Bienaymé-Tchebichev we see that, for each k,

$$\begin{aligned} P\left[\left|\frac{S_n - ES_n}{b_n}\right| \geq \varepsilon\right] &= P\left[|S_n - ES_n| \geq \varepsilon b_n\right] \\ &\leq \frac{1}{\varepsilon^2 b_n} \sum_{i=1}^{n} \sigma^2\left(X_k\right) \\ &= \frac{1}{\varepsilon^2} \left(\frac{1}{b_n} \sum_{i=n+11}^{n+k} \sigma^2\left(X_k\right)\right) \to 0 \end{aligned}$$

and this says precisely that

$$\frac{S_n - ES_n}{b_n} \to 0 \text{ in probability.} \blacksquare$$

We note that in the case each X_k takes the value 1 with probability p and takes the value 0 with probability $q = 1 - p$, and where we set $b_n = n$. We then get that $E(X_k) = p$, $\sigma^2(X_k) = pq$ and

$$\frac{1}{n^2}\sum_{k=1}^{n}\sigma^2(X_k) = \frac{1}{n}pq \to 0$$

so

$$\frac{S_n - np}{n} = \frac{S_n}{n} - p \to 0$$

in probability. This is Bernoulli's weak law of large numbers.

Note that the second inequality below strengthens the Bienaymé-Tchebichevcan inequality.

Theorem 2.30 (Kolmogorov inequalities) *If the independent r.v.'s X_k are integrable and the $|X_k| \le c$, c finite, then for every $\epsilon > 0$, and for all $n \in \mathbb{N}$,*

$$1 - \frac{(\epsilon + 2c)^2}{\sum_{k=1}^{n}\sigma^2 X_k} \le P[\max_{k\le n}|S_k - E(S_k)| \ge \epsilon] \le \frac{\sum_{k=1}^{n}\sigma^2 X_k}{\epsilon^2}.$$

Proof: We can assume, without restricting the generality, that the X_k, and hence S_k are centered at their expectations provided that we note that $|X_k| \le c$, implies that $E(S_k) \le c$, and so $|X_k - E(X_k)| \le 2c$. Recall from an earlier calculation, centering X_k at its expectation does not affect $S_k - E(S_k)$.

Therefore, we suppose that the collection of X_k are centered at their expectation and bounded by $2c$.

Let $A_k = [\max_{j\le k}|S_j| < \epsilon]$. In other words, A_k is the subset of Ω that satisfies $\forall j \le k, |S_j| < \epsilon$.

Let $B_k = A_{k-1} - A_k = [|S_1| < \epsilon, ..., |S_{k-1}| < \epsilon, S_k \ge \epsilon]$. In particular,
$B_k \subset [S_{k-1} < \epsilon, |S_k| \ge \epsilon]$.

Observe that $A_0 = \Omega$. So,

$$A_n^c = (\Omega - A_n) = (A_0 - A_n) = (A_0 - A_1) \cup (A_1 - A_2) \cup ... \cup (A_{n-1} - A_n) = \bigcup_{k-1}^{n} B_k.$$

Now, $E(S_n I_{B_k})^2 = E((S_n - S_k) + S_k)I_{B_k})^2 = E((S_n - S_k)I_{B_k} + S_k I_{B_k})^2$.

Which implies that

$$\int_{B_k} S_n^2 = \int_{B_k}((S_n - S_k) + S_k)^2 = \int_{B_k}(S_n - S_k)^2 + 2\int_{B_k}(S_n - S_k)S_k + \int_{B_k} S_k^2.$$

But we are assuming that the X_k are centered at their expectation. So, because $I_{B_k}S_k$ is independent from $S_n - S_k$, we have that

$$\begin{aligned}\int_{B_k} (S_n - S_k)\, S_k &= \int_\Omega (S_n - S_k)\, I_{B_k} S_k = \left(\int_\Omega (S_n - S_k)\right)\left(\int_\Omega I_{B_k} S_k\right)\\ &= 0 \cdot \left(\int_\Omega I_{B_k} S_k\right)\\ &= 0.\end{aligned}$$

Therefore,

$$\int_{B_k} S_n^2 = \int_{B_k} (S_n - S_k)^2 + \int_{B_k} S_k^2 \geq \int_{B_k} S_k^2$$

and by definition, $B_k \subset [|S_{k-1}| < \epsilon, |S_k| \geq \epsilon]$. So,

$$\int_{B_k} S_n^2 \geq \int_{B_k} S_k^2 \geq \epsilon^2 P[B_k].$$

Observe that summing over all $k \leq n$, we obtain

$$\sum_{k=1}^{n} \sigma^2 X_k = E(S_n^2) = \int_\Omega S_n^2 \geq \int_{A_n^c} S_n^2$$

$$= \sum_{k=1}^{n} \int_{B_k} S_n^2 \geq \epsilon^2 \sum_{k=1}^{n} P[B_k] = \epsilon^2 P[\bigcup_{k=1}^{n} B_k] = \epsilon^2 P[A_n^c].$$

which implies that

$$P[A_n^c] \leq \frac{\sum_{k=1}^n \sigma^2 X_k}{\epsilon^2}$$

and thus

$$P[\Omega - A_n] = P[\max_{j \leq n} |S_j| \geq \epsilon] \leq \frac{\sum_{k=1}^n \sigma^2 X_k}{\epsilon^2},$$

which establishes the inequality on the right in the statement of the theorem.

We begin the other side of the inequality with the following observation:

$$(S_{k-1} + X_k) I_{A_{k-1}} = S_k I_{A_{k-1}} = S_k I_{A_k + B_k} = S_k I_{A_k} + S_k I_{B_k}.$$

Because X_k is centered at its expectation, and X_k and $S_{k-1}I_{A_{k-1}}$ are independent, we have that $S_{k-1}I_{A_{k-1}}$ and X_k are orthogonal, and $I_{A_k}I_{B_k} = 0$.

It follows that:

$$\begin{aligned}
\int (S_k I_{A_{k-1}})^2 &= \int ((S_{k-1} + X_k) I_{A_{k-1}})^2 \\
&= \int (S_{k-1} I_{A_{k-1}})^2 + \int (X_k I_{A_{k-1}})^2 + 2 \int S_{k-1} X_k I_{A_{k-1}} \\
&= \int (S_{k-1} I_{A_{k-1}})^2 + \int (X_k I_{A_{k-1}})^2 \\
&= \int (S_{k-1} I_{A_{k-1}})^2 + E(X_k)^2 P[A_{k-1}] \\
&= E(S_{k-1} I_{A_{k-1}})^2 + \sigma^2 X_k P[A_{k-1}].
\end{aligned}$$

But the above observation then also gives

$$\begin{aligned}
\int (S_k I_{A_{k-1}})^2 &= \int (S_k I_{A_k} + S_k I_{B_k})^2 \\
&= \int (S_k I_{A_k})^2 + \int (S_k I_{B_k})^2 + 2 \int S_k I_{A_k} I_{B_k} \\
&= E(S_k I_{A_k})^2 + E(S_k I B_k)^2.
\end{aligned}$$

Now, since $|X_k| \leq 2c$, and $P[A_{k-1}] \geq P[A_n]$, we obtain

$$\begin{aligned}
|S_k I_{B_k}| &\leq |(S_{k-1} + X_k) I_{B_k}| \\
&\leq |S_{k-1} I_{B_k}| + |X_k I_{B_k}| \leq (\epsilon + 2c) I_{B_k}.
\end{aligned}$$

Therefore, it follows that

$$\begin{aligned}
E(S_{k-1} I_{A_{k-1}})^2 + \sigma^2 X_k P[A_{k-1}] &\leq E(S_{k-1} I_{A_{k-1}})^2 + \sigma^2 X_k P[A_n] \\
&\leq E(S_k I_{A_k})^2 + (\epsilon + 2c)^2 P[B_k].
\end{aligned}$$

Summing over $1 \leq k \leq n$, we obtain

$$\sum_{k=1}^{n} E(S_{k-1} I_{A_{k-1}})^2 + \sum_{k=1}^{n} \sigma^2 X_k P[A_n] \leq \sum_{k=1}^{n} E(S_k I_{A_k})^2 + \sum_{k=1}^{n} (\epsilon + 2c)^2 P[B_k].$$

Therefore,

$$\begin{aligned}
\sum_{k=1}^{n}\sigma^2 X_k P[A_n] &\le \sum_{k=1}^{n} E(S_k I_{A_k})^2 - \sum_{k=1}^{n} E(S_{k-1} I_{A_{k-1}})^2 + \sum_{k=1}^{n}(\epsilon+2c)^2 P[B_k] \\
&= E(S_n I_{A_n})^2 + (\epsilon+2c)^2 \sum_{k=1}^{n} P[B_k] \\
&\le \epsilon^2 P[A_n] + (\epsilon+2c)^2 P[A_n^c] \\
&\le (\varepsilon+c)^2 P[A_n] + (\varepsilon+c)^2 P[A_n^c] \\
&= (\varepsilon+c)^2 (P[A_n] + P[A_n^c]) \\
&= (\varepsilon+c)^2 .
\end{aligned}$$

Therefore,

$$P[A_n^c] = P[\max_{k\le n}|S_k - E(S_k)| \ge \epsilon] = 1 - P[A_n] \ge 1 - \frac{(\epsilon+2c)^2}{\sum_{k=1}^{n}\sigma^2 X_k}$$

which establishes the inequality on the left statement of the theorem. ■

Corollary 2.31 *Suppose that $\sum_{k=1}^{\infty}\sigma^2 X_k$ converges. Then $\sum_{k=1}^{\infty}(X_k - EX_k)$ converges a.s. If $\sum_{k=1}^{\infty}\sigma^2 X_k$ diverges and the X_k are uniformly bounded, then $\sum_{k=1}^{\infty}(X_k - EX_k)$ diverges a.s. Thus, if the X_k are uniformly bounded, then $\sum_{k=1}^{\infty}(X_k - EX_k)$ converges a.s. if and only if $\sum_{k=1}^{\infty}\sigma^2 X_k$ converges.*

Proof: Suppose that $\sum_{k=1}^{\infty}\sigma^2 X_k$ converges. Fix $\varepsilon > 0$ and $\delta > 0$. Then by the Kolmogorov inequality (on the right), if n is sufficiently large, $(\forall m \ge n)$

$$P\left[\max_{n\le k\le m}\left|\sum_{i=n}^{k}(X_i - E(X_i))\right| \ge \varepsilon\right] < \delta.$$

Note that these events form an increasing sequence of events as $m \to \infty$. Therefore, if n is sufficiently large,

$$P\left[\bigcup_{m=n}^{\infty}\left[\max_{n\le k\le m}\left|\sum_{i=n}^{k}(X_i - E(X_i))\right| \ge \varepsilon\right]\right] \le \delta.$$

In other words, if n is sufficiently large, then for most points, the tails of the series $\sum_{k=1}^{\infty}(X_k - EX_k)$ after n are bounded by ε.
We write

$$B_{\varepsilon,n} = \left[\bigcup_{m=n}^{\infty}\left[\max_{n\le k\le m}\left|\sum_{i=n}^{k}(X_i - E(X_i))\right| \ge \varepsilon\right]\right].$$

So if we fix a sequence of ε's decreasing to 0, say $\left\{\frac{1}{j}\right\}_{j=1}^{\infty}$, we choose, for each j an integer n_j so that the $n_j \uparrow \infty$ and

$$P\left[B_{\frac{1}{j},n_j}\right] \leq \frac{1}{2^j}.$$

So by the Borel-Cantelli lemma

$$P\left[B_{\frac{1}{j},n_j} \text{ i.o}\right] = 0.$$

So for almost every ω, ω is in only finitely many $B_{\frac{1}{j},n_j}$. So for such ω, and all sufficiently large j, $\omega \notin B_{\frac{1}{j},n_j}$, we have that for all $k \geq n_j$

$$\left|\sum_{i=n_j}^{k} \left(X_i\left(\omega\right) - E\left(X_i\right)\right)\right| \leq \frac{1}{j}.$$

In other words, the partial sums of the series $\sum_{k=1}^{\infty} \left(X_k\left(\omega\right) - EX_k\right)$ are Cauchy, and hence the series must converge.

Now suppose that the X_k are uniformly bounded by c, and $\sum_{k=1}^{\infty} \sigma^2 X_k$ diverges. Fix $\varepsilon > 0$. (We may choose ε as large as we like). Then for all n,

$$\sum_{k=n}^{\infty} \sigma^2 X_k$$

diverges, so

$$\lim_{m\to\infty} \frac{\left(\varepsilon + 2c\right)^2}{\sum_{k=n}^{m} \sigma^2 X_k} = 0,$$

and so we can choose m_n so that

$$1 - \frac{\left(\varepsilon + 2c\right)^2}{\sum_{k=n}^{m_n} \sigma^2 X_k} > 1 - \frac{1}{2^n}.$$

Therefore, for each n

$$1 - \frac{1}{2^n} < P\left[\max_{n\leq k\leq m_n} \left|\sum_{k=n}^{k} \left(X_k - EX_k\right)\right| \geq \varepsilon\right].$$

Suppose we let

$$G_{\varepsilon,n} = \left[\max_{n\leq k\leq m_n} \left|\sum_{k=n}^{k} \left(X_k - EX_k\right)\right| \geq \varepsilon\right].$$

Then by the Borel-Cantelli lemma,

$$P\left[\bigcup_{r=1}^{\infty}\bigcap_{n\geq r} G_{\varepsilon,n}\right] = 1.$$

That is, almost every ω is, for all sufficiently large n, in $G_{\varepsilon,n}$. So for almost every ω, and all sufficiently large n, there is some $k \geq n$ for which

$$\left|\sum_{k=n}^{k}(X_k(\omega) - EX_k)\right| \geq \varepsilon.$$

So the partial sums of the series are not Cauchy, and the series diverges. ■

We note that we could repeat this argument for a sequence of ε increasing to infinity, and find that for almost every ω, the partial sums fluctuate by arbitrarily large amounts infinitely often.

Corollary 2.32 *If the X_k are uniformly bounded by c and $\sum_{k=1}^{\infty} X_k$ converges a.s., then $\sum_{k=1}^{\infty}\sigma^2 X_k$ and $\sum_{k=1}^{\infty} E(X_k)$ converge.*

Proof: Let X_n' denote a set of r.v.'s bounded by c such that X_n and X_n' are identically distributed for every n, and $X_1, X_1', X_2, X_2', ...$ is a sequence of independent r.v.'s.

One way to construct such a collection is by forming the product measure space $(\Omega\times\Omega, P\times P)$. Then, consider the functions Y_n and Y_n' on $\Omega\times\Omega$ given by

$$Y_n(\omega_1,\omega_2) = X_n(\omega_1)$$

$$Y_n'(\omega_1,\omega_2) = X_n(\omega_2).$$

Then, it is evident that both of the sequences Y_n and Y_n' have the same distribution as the original sequence X_n, and Y_n and Y_n' are independent.

Now, we form the *symmetrized* sequence of independent r.v.'s

$$X_n^s = X_n - X_n'.$$

Then we have that

$$X_n^s \leq |X_n| + |X_n'| \leq 2c$$

and because the two sequences are identically distributed $E(X_n^s) = E(X_n) - E(X_n') = 0$ and

$$\sigma^2(X_n^s) = \sigma^2(X_n) - \sigma^2(X_n') = 2\sigma^2 X_n.$$

Now, suppose that $\sum_{k=1}^{\infty} X_k$ converges $a.s.$ Then because the primed sequence is identically distributed, $\sum_{k=1}^{\infty} X_k'$ converges $a.s.$ and so $\sum_{k=1}^{\infty} X_k^s$ converges $a.s.$

Therefore, $\sum_{k=1}^{\infty}(X_n^s - E(X_n^s)) = \sum_{k=1}^{\infty} X_n^s$ converges $a.s.$, and from the previous corollary we have that $\sum_{k=1}^{\infty} \sigma^2 X_k^s$, and thus $\sum_{k=1}^{\infty} \sigma^2 X_k$ converges.

Applying the previous corollary again, we have that $\sum_{k=1}^{\infty}(X_k - E(X_k))$ converges $a.s.$ So, $\sum_{k=1}^{\infty} X_k - \sum_{k=1}^{\infty}(X_k - E(X_k)) = \sum_{k=1}^{\infty} E(X_k)$ converges and the assertion is proved. ■

We now prove the theorem we are after.

Theorem 2.33 (Kolmogorov's three series criterion) *The series $\sum_{k=1}^{\infty} X_k$ of independent summands converges a.s. to a r.v. if, and only if, for some $c > 0$, the following three series converge:*

1. $\sum_{k=1}^{\infty} P[|X_k| \geq c]$.
2. $\sum_{k=1}^{\infty} \sigma^2 X_k^c$.
3. $\sum_{k=1}^{\infty} E(X_k^c)$.

Proof: We start with the assumption that the three series above converge. In particular, $\sum_{k=1}^{\infty} \sigma^2 X_k^c$ and $\sum_{k=1}^{\infty} E(X_k^c)$ converge, and by Corollary 2.31, $\sum_{k=1}^{\infty}(X_k^c - E(X_k^c))$ converges $a.s.$ which implies that $\sum_{k=1}^{\infty} X_k^c$ converges $a.s.$

Lastly, $\sum_{k=1}^{\infty} P[|X_k| \geq c]$ converges, which implies that $\sum_{k=1}^{\infty}[X_k \neq X_k^c]$ converges, and thus by the equivalence lemma $\sum_{k=1}^{\infty} X_k$ and $\sum_{k=1}^{\infty} X_k^c$ are convergent-equivalent. Therefore, $\sum_{k=1}^{\infty} X_k$ converges $a.s.$

Conversely, suppose that $\sum_{k=1}^{\infty} X_k$ converges $a.s.$ Then $X_k \to 0$ $a.s.$ and by Corollary 2.20 of the Borel zero-one criterion, we have the convergence of $\sum_{k=1}^{\infty} P[|X_k| \geq c]$. Now, again applying the equivalence lemma, we have convergence equivalence of $\sum_{k=1}^{\infty} X_k$ and $\sum_{k=1}^{\infty} X_k^c$.

Therefore, $\sum_{k=1}^{\infty} X_k^c$ converges $a.s.$ and by Corollary 2.32, $\sum_{k=1}^{\infty} \sigma^2 X_n^c$ and $\sum_{k=1}^{\infty} E(X_n^c)$ converge. ■

We note that the three series theorem could be formulated to say for all $c > 0$. This follows from the proof since the Borel zero one criterion allows $c > 0$ to be arbitrary.

Corollary 2.34 *If at least one of the three series in Kolmogorov's theorem does not converge, then $\sum_{k=1}^{\infty} X_k$ diverges a.s.*

Proof: By Corollary 2.20, $\sum_{k=1}^{\infty} X_k$ either converges $a.s.$ or diverges $a.s.$ ■

2.5 Consequence

Now that we have finally proven the Kolmogorov's three series criterion, we can return to our fundamental question.

Let $X_k = \epsilon_k(\omega)\left(\frac{1}{k^s}\right)$, where $s \in \mathbb{R}^+$ and $P[\epsilon_k = 1] = P[\epsilon_k = -1] = \frac{1}{2}$ denote a collection of independent r.v.'s defined on $(\Omega, \mathcal{A}, P)$.

Choose $c = 2$. Recall that

$$X^c = \begin{cases} X, & \text{if } |X| < c. \\ 0, & \text{if } |X| \geq c. \end{cases}$$

So, observe that in this case, $X_k^{c=2} = X_k$ since $\left|\epsilon_k(\omega)\left(\frac{1}{k^s}\right)\right| \leq 1$.

When $\frac{1}{2} < s \leq 1$, we calculate that

1. $\sum_{k=1}^{\infty} P[|X_k| \geq 2] = \sum_{k=1}^{\infty} P[\epsilon_k(\omega)\left(\frac{1}{k^s}\right) \geq 2] = \sum_{k=1}^{\infty} P[\emptyset] = 0 < \infty$.
2. $\sum_{k=1}^{\infty} \sigma^2 X_k^{c=2} = \sum_{k=1}^{\infty} \sigma^2 \epsilon_k(\omega)\left(\frac{1}{k^s}\right) = \sum_{k=1}^{\infty} E\left(\frac{1}{k^{2s}}\right) = \sum_{k=1}^{\infty} \frac{1}{k^{2s}} < \infty$.
3. $\sum_{k=1}^{\infty} E(X_k^{c=2}) = \sum_{k=1}^{\infty} E(X_k) = \sum_{k=1}^{\infty} E(\epsilon_k(\omega)\left(\frac{1}{k^s}\right)) = \sum_{k=1}^{\infty} 0 = 0 < \infty$.

We conclude that in this case, $\sum_{k=1}^{\infty} X_k$ converges $a.s.$ However, if we select $-\infty < s \leq \frac{1}{2}$, we see $\sum_{k=1}^{\infty} \sigma^2 X_k \geq \sum_{k=1}^{\infty} \frac{1}{k} = \infty$. Therefore, by Corollary 2.34, $\sum_{k=1}^{\infty} X_k$ diverges $a.s.$

Therefore:

- $\sum_{k=1}^{\infty} \epsilon_k(\omega)\left(\frac{1}{k^s}\right)$ diverges $a.s.$ when $-\infty < s \leq \frac{1}{2}$.
- $\sum_{k=1}^{\infty} \epsilon_k(\omega)\left(\frac{1}{k^s}\right)$ converges $a.s.$ when $\frac{1}{2} < s \leq 1$.
- $\sum_{k=1}^{\infty} \epsilon_k(\omega)\left(\frac{1}{k^s}\right)$ converges absolutely when $s > 1$.

The three series theorem also tells us following fact about the logarithmic series.

Claim 2.35 *For all $s \in \mathbb{R}$, the series $\sum_{k=2}^{\infty} \epsilon_k(\omega)\left(\frac{1}{\ln^s k}\right)$ diverges a.s.*

All we must show is that one of the three series diverges. Again, let $c = 2$, and let $X_k = \epsilon_k(\omega)\left(\frac{1}{\ln^s k}\right)$. Observe

$$\sum_{k=2}^{\infty} \sigma^2(X_k^{c=2}) = \sum_{k=2}^{\infty} E\left(\frac{1}{\ln^{2s} k}\right) = \sum_{k=2}^{\infty} \frac{1}{\ln^{2s} k}.$$

One could easily show with calculus techniques that for all $s \in \mathbb{R}$,

$$\sum_{k=2}^{\infty} \frac{1}{\ln^{2s} k} > \sum_{k=1}^{\infty} \frac{1}{k},$$

which implies that $\sum_{k=2}^{\infty} \frac{1}{\ln^{2s} k}$ diverges. Therefore, the claim is satisfied and we have that $\sum_{k=2}^{\infty} \epsilon_k(\omega) \left(\frac{1}{\ln^s k}\right)$ diverges *a.s.*

3 Random Harmonic Series With Ergodic Processes

In this chapter we approach the problem of random harmonic series from a different viewpoint. In the previous chapter we rigorously went through the necessary steps to prove the Kolmogorov's three series criterion. This theorem showed us explicitly that the random harmonic series with coefficients 1's and -1 assigned by flipping a fair coin converges *a.s.* It also gives us the exact circumstance in which a series of independent summands converge *a.s.*

This motivates exploring summands with other properties. In the present chapter we will consider non-independent processes $\{X_k\}$ and investigate the pointwise convergence of the series $\sum X_k(\omega)\left(\frac{1}{k}\right)$. The processes we will consider will be stationary and ergodic, and for the most part will only take the values 1 and -1. In order to describe these processes, we will need to introduce a new point of view.

Definition 3.1 *Let $(\Omega, \mathcal{A}, P)$ be a probability space. A measure preserving transformation T on Ω is a measurable map $T : \Omega \to \Omega$ such that for all $A \in \mathcal{A}$, $P\left[T^{-1}A\right] = P[A]$. A set $A \in \mathcal{A}$ is ($T-$) invariant if $T^{-1}(A) = A$. T is ergodic if every invariant set has probability 0 or 1.*

Given a measure preserving transformation T on Ω and a random variable $X : \Omega \to \mathbb{C}$, we obtain a process $\{X_k\}_{k=0}^{\infty}$ by setting $X_k = X\left(T^k\right)$. This process is stationary in the sense that, for all n and m, the joint distribution of the r.v.'s $\{X_k\}_{k=1}^{n}$ equals the joint distribution of $\{X_{k+m}\}_{k=1}^{n}$.

We will be concerned with random series of the form

$$\sum_{k=1}^{\infty} X\left(T^k \omega\right)\left(\frac{1}{k}\right),$$

where T is an ergodic measure preserving transformation.

We will make use of the following fundamental fact, which is a special case of the Birkhoff ergodic theorem.

Theorem 3.2 *Suppose that T is an ergodic transformation on $(\Omega, \mathcal{A}, P)$ and $X \in L^1(\Omega, \mathcal{A}, P)$. Then*

$$\frac{1}{n}\sum_{k=1}^{n} X\left(T^k\omega\right) \to E(X) \text{ a.e.}$$

A proof for the theorem can be found in the Walters textbook, item 8 in the references.

As an example, we consider the case where X is an integrable r.v. and $E(X) \neq 0$.

Proposition 3.3 *Suppose that T is an ergodic measure preserving transformation of $(\Omega, \mathcal{A}, P)$ and that $X : \Omega \to \mathbb{C}$ is integrable with $E(X) \neq 0$. Then for a.e. $\omega \in \Omega$, the series $\sum_{k=1}^{\infty} X\left(T^k\omega\right)\frac{1}{k}$ diverges.*

Proof: Since $E(X) \neq 0$, we can choose ε so that

$$0 < \varepsilon < \frac{1}{2}\max\{\mathrm{Re}(E(X)), \mathrm{Im}(E(x))\}.$$

Thus, in $\{z : |z - E(X)| < \varepsilon\}$, either $\mathrm{Re}\, z$ is of one sign or $\mathrm{Im}\, z$ is of one sign and bounded away from 0.

Choose $\omega \in \Omega$ so that

$$\frac{1}{n}\sum_{k=1}^{n} X\left(T^k\omega\right) \to E(X).$$

Choose N_0 so that for all $n \geq N_0$,

$$\left|\frac{1}{n}\sum_{k=1}^{n} X\left(T^k\omega\right) - E(X)\right| < \varepsilon.$$

For each n, we write

$$s_n = \sum_{k=1}^{n} X\left(T^k\omega\right).$$

By summation by parts we have, for each n,

$$\begin{aligned}
\sum_{k=1}^{n} X\left(T^k\omega\right)\frac{1}{k} &= s_n\left(\frac{1}{n+1}\right) - \sum_{k=1}^{n} s_k\left(\frac{1}{k+1} - \frac{1}{k}\right) \\
&= s_n\left(\frac{1}{n+1}\right) + \sum_{k=1}^{n} s_k\left(\frac{1}{k(k+1)}\right) \\
&= s_n\left(\frac{1}{n+1}\right) + \sum_{k=1}^{n} \frac{s_k}{k}\left(\frac{1}{k+1}\right).
\end{aligned}$$

We have

$$\lim_{n\to\infty} s_n\left(\frac{1}{n+1}\right) = \lim_{k\to\infty}\frac{s_k}{k} = E(X) \neq 0.$$

But for all $k \geq N_0$,

$$\left|\frac{s_k}{k} - E(X)\right| < \varepsilon.$$

So either $\operatorname{Re}\left(\frac{s_k}{k}\right)$ or $\operatorname{Im}\left(\frac{s_k}{k}\right)$ is of constant sign and bounded away from 0. Therefore, either $\sum_{k=1}^{n}\operatorname{Re}\left(\frac{s_k}{k}\left(\frac{1}{k+1}\right)\right)$ diverges or $\sum_{k=1}^{n}\operatorname{Im}\left(\frac{s_k}{k}\left(\frac{1}{k+1}\right)\right)$ diverges. In either case $\sum_{k=1}^{n}\left(\frac{s_k}{k}\left(\frac{1}{k+1}\right)\right)$ diverges. Therefore, $\sum_{k=1}^{n} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$ diverges. ■

In view of this, we will concentrate on integrable r.v.'s X with $E(X) = 0$. (We will not address the case of non-integrable r.v.'s). In fact, for the most part, we will restrict attention to r.v.'s $X : \Omega \to \{1, -1\}$ and $E(X) = 0$.

We begin with the following observation:

Proposition 3.4 *Suppose then that T is an ergodic, measure preserving transformation of the probability space (Ω, A, P), and $X : \Omega \to \{1, -1\}$ has $E(X) = 0$. Suppose that $\{b_k\}_{k=1}^{\infty}$ is sequence such that $\left\{\frac{b_{k-1}}{b_k}\right\}_{k=2}^{\infty}$ is monotonic and converges. Then*

$$P\left\{\omega \in \Omega : \sum_{k=1}^{\infty} X\left(T^k\omega\right) b_k \text{ converges}\right\}$$

equals 0 *or* 1.

Proof: By proposition 1.9 in the Prologue, the series

$$\sum_{k=1}^{\infty} X\left(T^k\omega\right) b_k$$

converges if and only if the series

$$\sum_{k=1}^{\infty} X\left(T^{k+1}\omega\right) b_k.$$

In other words, the set

$$\left\{\omega \in \Omega : \sum_{k=1}^{\infty} X\left(T^k\omega\right) b_k \text{ converges}\right\}$$

is $T-$ invariant. So since T is ergodic, this set has probability 0 or 1. ■

In particular, the above proposition applies to the sequence $\{a_k\}_{k=1}^{\infty} = \left\{\frac{1}{k}\right\}_{k=1}^{\infty}$. Thus, the series we are are studying, series of the form $\sum_{k=1}^{\infty} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$ where T is ergodic and measure preserving, and where $E(X) = 0$, must converge *a.e.* or diverge *a.e.*

In the case that $X : \Omega \to \{1, -1\}$, and $E(X) = 0$, the ergodic theorem says that for *a.e.* $\omega \in \Omega$, there are, asymptotically, the same number of 1's and -1's in the sequence $\left\{X\left(T^k\omega\right)\right\}_{k=1}^{\infty}$. In other words, the series $\sum_{k=1}^{\infty} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$ resembles (in some sense), an alternating series. It might seem plausible that this would be enough to ensure that the series $\sum_{k=1}^{\infty} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$ should converge *a.e.* But we will show that this is not the case. In fact, we will prove the following result of B. Weiss, which is the main objective of this chapter:

Theorem 3.5 *(B. Weiss) For every ergodic measure preserving transformation T on a probability space (Ω, A, P), there is a r.v. $X : \Omega \to \{1, -1\}$ with $E(X) = 0$, such that $\sum_{k=1}^{\infty} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$ diverges a.e.*

This is an unpublished result, which was communicated by B. Weiss to A. Fieldsteel. The proof, which we will give below, is similar to arguments given in a paper written by Thouvenot and Weiss and can be found as item 7 in the references.

This theorem will be proved by first constructing such a particular transformation T and r.v. X, where T is of a special type, called *rank-1*, and then showing how to use that construction to produce the general result.

From here on, for the sake of brevity, we may make use of the following language.

Definition 3.6 *Given an ergodic measure preserving transformation T on a probability space (Ω, A, P), and an integrable r.v X on Ω, we say that (T, X) is a **convergent** process if $\sum_{k=1}^{\infty} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$ converges a.e. and that (T, X) is a **divergent** process if $\sum_{k=1}^{\infty} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$ diverges a.e.*

3.1 A Rank-1 Construction

We begin by making a construction of a divergent process (T, X). This construction will be of a type that is called rank-1, but rather than give a general definition of such constructions, we will proceed directly with this particular construction, and afterwards outline the general concept.

The space on which our transformation will be defined will be a disjoint union of intervals in $\mathbb{R}$, each endowed with Lebesgue measure μ. The construction will ensure that the measure of this space will be finite, and so can be normalized to have measure 1.

Step 1: Let B be a bounded interval in $\mathbb{R}$. Our construction will be governed by an increasing sequence of natural numbers $\{c_n\}_{n=0}^{\infty}$, which will be specified later. Subdivide B into $2c_0$ subintervals of equal length, which we label as $\{B_{0,i}\}_{i=0}^{2c_0-1}$, and for each $i < 2c_0 - 1$ we define T on to be a linear map that carries $B_{0,i}$ to $B_{0,i+1}$. We view this set of intervals as a *tower* whose i^{th} level is the interval $B_{0,i}$ and on which T acts by moving points upward by one level. T is as yet undefined on the top level $B_{0,2c_0-1}$. We refer to this tower (or to the union of its levels) as τ_0 and to the number of levels in τ_0 as its height h_0, so that $h_0 = 2c_0$. (Throughout this construction we may discard the endpoints of these intervals. The set of all the endpoints of all the intervals in this construction is a countable set, and hence has measure zero, and so discarding them will not affect our argument).

We define a function X on τ_0 by setting $X \equiv (-1)^{i+1}$ on $B_{0,i}$. That is, if ω is a point in the base $B_{0,0}$ of τ_0, then along its $T-$orbit $\left\{T^i\omega\right\}_{i=0}^{2c_0-1}$, the values of X simply alternate between 1 and -1. We refer to this sequence of values as a $0-$block.

$$\tau_0 \left\{ \begin{array}{|c|} \hline T^{2c_0-1}(B_{0,0}) \\ \hline \vdots \\ \hline T(B_{0,0}) \\ \hline B_{0,0} \\ \hline \end{array} \right. \quad \begin{array}{c} 1 \\ \vdots \\ -1 \\ 1 \end{array}$$

Step 2: We add one additional level, which we call D_1, to the top of τ_0 and extend T so that it maps $B_{0,2c_0-1}$ linearly to D_1. We denote this larger tower by τ_0'. Now we divide τ_0' into $2c_1$ columns of equal width. That is, we divide the base $B_{0,0}$ of τ_0' into $2c_1$ subintervals of equal length, $\{B_{1,i}\}_{i=0}^{2c_1-1}$, each of which can be viewed as the base of a sub-tower or column of τ_0'.

Let $C_t(\tau_0')$ denote the $(t-1)^{th}$ column of τ_0'.

$$C_t(\tau_0') \left\{ \begin{array}{|c|} \hline D^1 \\ \hline T^{2c_0-1}(B_{1,t}) \\ \hline \vdots \\ \hline T(B_{1,t}) \\ \hline B_{1,t} \\ \hline \end{array} \right. \quad \begin{array}{c} \\ 1 \\ \vdots \\ -1 \\ 1 \end{array}$$

We extend the definition of T by concatenating these columns to form a single tower τ_1. That is, for each subinterval $B_{1,i}$ of $B_{0,0}$, where $i < 2c_1 - 1$, T is defined to linearly map $T^{2c_0}B_{1,i}$ to $B_{1,i+1}$. This tower τ_1 has height $h_1 = 2c_1(2c_0 + 1)$, and we have defined T on all but the top level of τ_1.

Thus far X has been defined on τ_1, but not on the added interval D_1. Note that D_1 has been subdivided into $2c_1$ subintervals of equal length, which are the top levels of the columns described in the previous paragraph. Now we extend the definition of X to D_1 by setting $X \equiv 1$ on each of the subintervals of D_1 that lie in the bottom half of τ_1 and setting $X \equiv -1$ on all those in the top half of τ_1. Thus, for each x in the base $B_{1,0}$ of τ_1 the values of X along the orbit $\left\{T^i\omega\right\}_{i=0}^{2c_1-1}$ is a sequence of $2c_1$ copies of the $0-$block, with a single 1 following each of the first c_1 copies, and a single -1 following the rest. We refer to this sequence as a $1-$block.

τ_1	$C_{2c_n-1}(\tau_0')$	-1
	$\vdots$	$\vdots$
	$C_{2c_n-2}(\tau_0')$	-1
	$C_{c_1}(\tau_0')$	-1
	—	—
	$C_{c_1-1}(\tau_0')$	1
	$\vdots$	$\vdots$
	$C_1(\tau_0')$	1
	$C_0(\tau_0')$	1

Step 3: We continue this procedure indefinitely. After the n^{th} tower τ_n has been constructed and T defined on all but its top level and X defined on τ_n, we add an additional level D_{n+1} to the top of the tower to form τ_n', we cut τ_n' into $2c_{n+1}$ columns of equal measure, and we concatenate them (or stack them) to form the next tower τ_{n+1}. This indicates how to extend T to all but the top level of τ_{n+1}. We also extend X to τ_{n+1} by setting X equal to 1 on the half of D_{n+1} that lies in the bottom half of τ_{n+1}, and equal to -1 on the rest of D_{n+1}.

Let Ω be the union of the towers τ_n. (We are using τ_n to denote both a set of levels and the union of those levels). The set Ω may be viewed either as a subset of $\mathbb{R}$, with Lebesgue measure, or as an abstract disjoint union of measure spaces, but in either case, we have

$$\mu(\Omega) = \mu(B) + \sum_{n=1}^{\infty} \mu(D_n).$$

For each n, $\mu(D_{n+1}) \le \frac{1}{2}\mu(D_n)$, so this series converges, and $\mu(\Omega) < \infty$. Thus (once the sequence $\{c_n\}_{n=0}^{\infty}$ is chosen), we can choose B so that $\mu(\Omega) = 1$. Going forward, we will assume this has been done, and we will revert to our usual notation, using P to denote this probability measure.

For each n, at the n^{th} stage of our construction, T is defined everywhere except at the top level of τ_n, and these levels form a decreasing sequence of sets, whose measure approaches 0. Hence T is defined almost everywhere. In fact, if we delete all the endpoints of our intervals, the top levels of these towers can be made to have empty intersection, so that T (and T^{-1}) is defined everywhere.

We next show that T is measure preserving, and that T is ergodic. We will note that the proofs of these assertions do not depend on the choice of the sequence $\{c_n\}_{n=0}^{\infty}$.

Lemma 3.7 *The transformation T constructed above is measure preserving.*

Proof: Fix a measurable set $A \subset \Omega$ with $P[A] > 0$. Fix $\varepsilon > 0$ with $\varepsilon < P[A]$. For each n let A_n denote the portion of A that is contained in τ_n and not in the top nor the bottom level of τ_n. Choose n so that

$$P[A_n] > P[A] - \varepsilon.$$

Then it is immediate that

$$P[T^{-1}A_n] = P[A_n] = P[TA_n].$$

So

$$P\left[T^{-1}A\right] \geq P[A] - \varepsilon \text{ and } P[TA] \geq P[A] - \varepsilon.$$

Since this is true for all $\varepsilon > 0$, we have

$$P[T^{-1}A] \geq P[A] \text{ and } P[TA] \geq P[A].$$

Since this is true for all A of positive measure, this implies that for all A of positive measure

$$P[A] = P\left[T^{-1}A\right].$$

Hence for sets A of measure 0 as well

$$P[A] = P\left[T^{-1}A\right]. \blacksquare$$

Lemma 3.8 *The transformation T constructed above is ergodic.*

Proof: Fix $A \subset \Omega$ with $P[A] > 0$. We will show that if we write

$$\tilde{A} = \bigcup_{n=-\infty}^{\infty} T^n A$$

we have

$$P\left[\tilde{A}\right] = 1.$$

From this it will follow that every invariant set of positive measure has full measure, and so T is ergodic.

We recall the following fact about Lebesgue measure: if $P[A] > 0$ then for all $\varepsilon > 0$ there is an interval I such that

$$P[A \cap I] > (1 - \varepsilon) P[I].$$

Choose n so that $P[A \cap \tau_n] > 0$, and choose a level J of τ_n for which $P[A \cap J] > 0$. Choose an interval $I \subset J$ as above, so that

$$P[A \cap I] > (1 - \varepsilon) P[I].$$

For each $m > n$, J is subdivided into intervals which are levels of τ_m, all of which have the same length. If m is sufficiently large, the common length of these subintervals is far less than the length of I, and so at least one of such interval K has

$$P[A \cap K] > (1 - 2\varepsilon) P[K].$$

But then for every interval K' in τ_m we have

$$P\left[\tilde{A} \cap K'\right] > (1 - 2\varepsilon) P\left[K'\right]$$

and so

$$P\left[\tilde{A} \cap \tau_m\right] > (1 - 2\varepsilon) P[\tau_m].$$

So again, if m is sufficiently large, we have $P[\tau_m] > 1 - \varepsilon$. This would give

$$P\left[\tilde{A}\right] > (1 - 2\varepsilon)(1 - \varepsilon).$$

Since $\varepsilon > 0$ is arbitrary, this gives

$$P\left[\tilde{A}\right] = 1. \blacksquare$$

In general, any transformation that can be constructed in this manner is called a rank-1 transformation. Specifically, the transformation is constructed by building a sequence of towers, whose levels are intervals, where each tower is obtained from the previous one by subdividing it into columns which are concatenated, and by the possible addition of some more levels, as long as the sequence of measures of these towers is bounded. We note that the proofs of the previous two lemmas apply without change to all such constructions.

Now we return to the details of our constructed transformation T and see that we can arrange for the process (T, X) to be divergent.

Proposition 3.9 *In the construction of the process* (T, X) *as above, the sequence* $\{c_n\}_{n=0}^{\infty}$ *can be chosen so that* (T, X) *is divergent.*

Proof: We describe inductively how the numbers c_n will be chosen, so as to produce a divergent process (T, X). Suppose that we have already constructed $(n-1)^{\text{st}}$ tower, τ_{n-1}, on which we have defined the function X. In order to construct the n^{th} tower τ_n we only need to choose the number c_n, which specifies the number of columns into which τ_{n-1} is subdivided.

As before, we let h_{n-1} denote the height of τ_{n-1}. Recall by our construction that for $n > 1$, $h_{n-1} = 2c_{n-1}(h_{n-2} + c_{n-1})$. Let E_{n-1} denote the base of τ_{n-1}. Recall also that our construction makes X constant on each level of τ_{n-1} and that for each $\omega \in E_{n-1}$, the values of X along the orbit of ω up the tower gives the same number of $1's$ as $-1's$. That is, for each $\omega \in E_{n-1}$,

$$\sum_{i=0}^{h_{n-1}} X\left(T^i\omega\right) = 0.$$

In fact, c_n will be chosen with respect to another parameter L_n, which will be identified later. Once L_n is given, we will choose c_n so large that the vast majority of points $\omega \in \tau_n$ will have the property that either

1. All the points $\left\{T^i\omega\right\}_{i=0}^{L_n-1}$ lie in the bottom half of τ_n,

   or

2. All the points $\left\{T^i\omega\right\}_{i=0}^{L_n-1}$ lie in the top half of τ_n.

We let B_n^+ denote the subset of Ω satisfying condition 1, and B_n^- the subset of Ω satisfying condition 2.

We know that $h_n = 2c_n(h_{n-1} + 1)$, so if L_n has already been chosen, and since h_{n-1} is already known, we just need to choose c_n so large that

$$L_n << h_n = 2c_n(h_{n-1} + 1).$$

To be explicit, we can choose c_n so that

$$\frac{2L_n}{h_n} < \frac{1}{2^n}$$

which implies that

$$P\left(B_n^+ \cup B_n^-\right) < \frac{1}{2^n} P(\tau_n).$$

Therefore, we start by choosing L_n.

Our goal is to choose L_n so that for all $\omega \in B_n^+$,

$$\sum_{k=0}^{L_n-1} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$$

is hugely positive, while for all $\omega \in B_n^-$,

$$\sum_{k=0}^{L_n-1} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$$

is hugely negative.

To see how to do this, we first note that if ω were to lie in $(B_n^+ \cup B_n^-)$ and also in the base E_{n-1} of τ_{n-1}, and if L_n were a multiple, say M, of $h_{n-1}+1$, then the sum $\sum_{k=0}^{L_n-1} X\left(T^k\omega\right)\frac{1}{k}$ could be broken up into two convenient parts: the part coming from the extra level D_n inserted in the construction of τ_n and the rest, which would come from M whole columns of τ_{n-1}.

The contribution from the whole columns can be bounded above by the following estimate. The whole sum can be written as a sum of h_{n-1} alternating series, whose first terms are no greater than 1. So each of these alternating sums must be, in absolute value, no greater than 1. So the sum of all h_{n-1} of them must be, in absolute value, no greater than h_{n-1}. Note that this estimate is valid, no matter how large M is, and moreover, it is an estimate we can make before we choose c_n. On the other hand, the other part of this sum is a sum of the form

$$\pm\sum_{j=1}^{M}\frac{1}{j\left(h_{n-1}+1\right)}.$$

We know that the absolute value of this sum grows without bound as $M \to \infty$. In particular, if M is large enough, then the absolute value of this sum far exceeds the estimate that we just gave for the rest of the series, so that the whole sum $\sum_{k=0}^{L_n-1} X\left(T^k\omega\right)\frac{1}{k}$ would have to be very large in absolute value.

However, the sum that arises from the points $\sum_{k=0}^{L_n-1} X\left(T^k\omega\right)\frac{1}{k}$ (for $\omega \in B_n^+$, say) may have a more complicated description. If ω does not lie in E_{n-1} then the orbit $\left\{T^k\omega\right\}_{k=0}^{L_n-1}$ breaks into an initial segment that is part of a column of τ'_{n-1}, and then a long sequence of whole columns and then another part of a column. The sum $\sum_{k=0}^{L_n-1} X\left(T^k\omega\right)\frac{1}{k}$ could then be broken into corresponding parts, and we need to see that if L_n is chosen large enough, this sum will also be large.

So we choose L_n so that we'll see at least M_n such whole columns, and that will be the case if we take

$$L_n > \left(M_n+2\right)h_{n-1}.$$

And we observe that the contribution to the sum from the initial and final segments of columns of τ_{n-1} is bounded by $2h_{n-1}$, since all the summands are no greater than 1. So in this general case, our estimate of the portion of the sum that must be overwhelmed by the extra 1's in B_n^+ is really $3h_{n-1}$ instead of h_{n-1}. Note that in this case the extra 1's may not occur at multiples of $h_{n-1}+1$, but their contribution will be bounded below by

$$\sum_{j=1}^{M_n} \frac{1}{j\left(h_{n-1}+1\right)}.$$

A similar discussion applies to points in B_n^-.

So now we are able to specify our choices exactly.

We first choose M_n so that

$$\left|\sum_{j=1}^{M_n} \frac{1}{jh_{n-1}}\right| > n + 3h_{n-1},$$

and then choose L_n such that

$$L_n = \left(M_n + 2\right) h_{n-1},$$

and then choose c_{n-1} so that

$$\frac{2L_n}{h_n} < \frac{1}{2^n}.$$

With these choices we have that

$$\sum_{n=1}^{\infty} P[\tau_n \backslash (B_n^+ \cup B_n^-)] = \sum_{n=1}^{\infty} \frac{1}{2^n} < \infty.$$

So, by the Borel-Cantelli lemma, theorem 2.17, we know that

$$P[\lim_{n\to\infty} \sup\left(\tau_n \backslash (B_n^+ \cup B_n^-)\right)] = 0.$$

But for $a.e.\ \omega \in \Omega$ there is some N_1 so that for all $n \geq N_1$, $\omega \in \tau_n$ and so for $a.e.\ \omega$ there is some N_2 so that for all $n \geq N_2$ $\omega \notin \tau_n \backslash (B_n^+ \cup B_n^-)$.

Therefore, for all $n \geq \max\{N_1, N_2\}$, $\omega \in B_n^+ \cup B_n^-$, and thus for such ω and n, we will have

$$\left|\sum_{k=0}^{L_n-1} X\left(T^k\omega\right)\frac{1}{k}\right| > n.$$

Therefore, we conclude that for *a.e.* ω,

$$\sum_{k=0}^{n} X\left(T^{k}\omega\right)\frac{1}{k}$$

diverges, or in other words, (T, X) is divergent. ■

The above argument shows that for *a.e.* ω

$$\limsup_{n\to\infty}\left|\sum_{k=0}^{n} X\left(T^{k}\omega\right)\frac{1}{k}\right| = \infty$$

so that for *a.e.* ω, either

$$\limsup_{n\to\infty}\sum_{k=0}^{n} X\left(T^{k}\omega\right)\frac{1}{k} = \infty \text{ or } \liminf_{n\to\infty}\sum_{k=0}^{n} X\left(T^{k}\omega\right)\frac{1}{k} = -\infty.$$

In fact, another application of summation by parts, Corollary 1.3, gives us that $\limsup_{n\to\infty}\sum_{k=0}^{n} X\left(T^{k}\omega\right)\frac{1}{k} = \infty$ then $T(\omega)$ has the same property. In other words, if

$$A = \left\{\omega\in\Omega : \limsup_{n\to\infty}\sum_{k=0}^{n} X\left(T^{k}\omega\right)\frac{1}{k} = \infty\right\},$$

then $T(A)\subset A$. The ergodicity of T then implies that $P(A) = 0$ or $P(A) = 1$. The same applies to the set

$$A' = \left\{\omega\in\Omega : \liminf_{n\to\infty}\sum_{k=0}^{n} X\left(T^{k}\omega\right)\frac{1}{k} = -\infty\right\}.$$

But in fact, for this process (T, X) we can see from the construction that both these sets have probability 1.

Proposition 3.10 *For the process (T, X) constructed above, we have for a.e. $\omega\in\Omega$,*

$$\limsup_{n\to\infty}\sum_{k=0}^{n} X\left(T^{k}\omega\right)\frac{1}{k} = \infty \text{ and } \liminf_{n\to\infty}\sum_{k=0}^{n} X\left(T^{k}\omega\right)\frac{1}{k} = -\infty.$$

Proof: It will suffice to prove that for *a.e.* $\omega \in \Omega$, and for infinitely many n, $\omega \in B_n^+$ and also, for infinitely many n, $\omega \in B_n^-$. The construction of T shows that for each n

$$P(B_{n+1}) \leq \frac{1}{2} P(B_n)$$

so that for each N

$$P\left(\bigcap_{n \geq N} B_n\right) = 0$$

and so

$$P\left(\bigcup_N \bigcap_{n \geq N} B_n\right) = 0$$

so that almost every ω is infinitely often not in B_n^+. A similar argument shows the same for B_n^-. But we showed already that almost every ω is for all sufficiently large n, in the union $B_n^+ \cup B_n^-$. ■

3.2 Weiss's Theorem

We now adapt the construction of (T, X) given in the previous section to prove the main theorem of this chapter, namely that every ergodic transformation T admits a function X taking only the values ± 1 with $E(X) = 0$ so that (T, X) is divergent.

Note that we might have conjectured, given the results of the previous chapter, that if T is the shift map on the sequence space $\Omega = \{1, -1\}^{\mathbb{Z}}$ with the independent $\left(\frac{1}{2}, \frac{1}{2}\right)$ product measure, then every other function $X : \Omega \to \{1, -1\}$ with $E(X) = 0$ would given a convergent process (T, X). A similar conjecture might have been made with respect to an irrational rotation on the unit circle. But this theorem says that no ergodic transformation T has the property that all such (T, X) are convergent.
To avoid trivial counterexamples, we will have to restrict our discussion to non-atomic probability spaces: every measurable set with positive measure properly contains sets with positive measure. In addition, our argument will depend on the following fundamental lemma.

Lemma 3.11 (Kakutani-Rokhlin) *Let T be an ergodic, measure preserving transformation on a non-atomic probability space $(\Omega, \mathcal{A}, P)$. Then:*

1. *For all $\varepsilon > 0$ and all $n \in \mathbb{N}$, there is a set $A \in \mathcal{A}$ such that the sets $\left\{T^i A\right\}_{i=0}^{n-1}$ are pairwise disjoint and $P\left[\bigcup_{i=0}^{n-1} T^i A\right] > 1 - \varepsilon$.*

2. *Given a finite measurable partition Q of Ω the tower may be chosen so that the conditional distribution of P on A equals the distribution of Q.*

A proof can be found in a book by Shields, item 6 in the references.

As before, the collection of sets $\tau = \{T^i A\}_{i=0}^{n-1}$ given by the lemma is called a (Kakutani-Rokhlin)-tower for T. It's height is n and the tower τ may be referred to as an (n, ε) −tower with base A. We use the same symbol τ to denote $\bigcup_{i=0}^{n-1} T^i A$. Repeated application of this lemma will provide towers for T on which we may imitate the construction of the divergent process which was done in the previous section.

Proof: (Weiss's theorem) Fix an ergodic measure preserving transformation T on a non-atomic probability space $(\Omega, \mathcal{A}, P)$. Our construction of the r.v. X making (T, X) divergent will be governed by several parameters. We will defer the exact specification of these parameters until later in the argument.

The general plan of proof will be as follows. We will construct a sequence of towers τ_n such that for each n :

1. τ_n is a $(2h_n, \varepsilon_n)$ tower with base A_n, where h_n is even and $\lim_{n\to\infty} \varepsilon_n = 0$.
2. Dividing each τ_{n+1}−orbit into its top and bottom halves, we have that each τ_n−orbit appears as an unbroken sequence in each half- τ_{n+1}−orbit. (So in particular, $\tau_n \subset \tau_{n+1}$).
3. For each half-τ_{n+1}−orbit, the consecutive τ_n−orbits are separated by at least 10 levels, and by not more than $2h_n$ levels.
4. In each τ_{n+1}-orbit, we see the same number of τ_n−orbits in the top half as in the bottom half.

Repeated applications of the Kakutani-Rokhlin lemma and the ergodic theorem allow one to construct such towers, but we do not include the explicit details here.

With such a sequence of towers, we construct X inductively as follows. Define X on τ_1 by setting $X = (-1)^i$ on $T^i A_1$. Having defined X on τ_n we fix $\omega \in A_{n+1}$ and define X on the τ_{n+1}−orbit of ω by a two-step procedure.

Step 1: Fill in the values of X on the complement of τ_n by a single alternating sequence of 1 and −1.

Step 2: Then in each gap between consecutive τ_n−orbits in the bottom half of the τ_{n+1}−orbit of ω, change one −1 to a 1. Similarly, in each gap in between consecutive τ_n−orbits in the bottom half of the τ_{n+1}−orbit of ω, change one 1 to a −1.

Since $\lim_{n\to\infty} \varepsilon_n = 0$, this defines X *a.e.* We show that the parameters $\{(h_n, \varepsilon_n)\}_{n=1}^{\infty}$ can be chosen so that this construction gives a divergent process (T, X).

The argument that (T, X) is divergent is almost the same as the argument we gave for the rank-1 process of the previous section. We will arrange that for each n, there is some number L_n with

$$h_{n-1} << L_n << h_n$$

and for most points ω in the bottom half of τ_n,

$$\sum_{k=1}^{L_n} X\left(T^k\omega\right)\frac{1}{k}$$

is large positive, while for most points ω in the top half of τ_n

$$\sum_{k=1}^{L_n} X\left(T^k\omega\right)\frac{1}{k}$$

is large negative. And if we arrange this, then the same arguments we gave for the previous example will show that $\sum_{k=1}^{L_n} X\left(T^k\omega\right)\frac{1}{k}$ diverges *a.e.*, in the strong sense of proposition 3.10.

The estimate to show that $\sum_{k=1}^{L_n} X\left(T^k\omega\right)\frac{1}{k}$ is large for most ω in the bottom of τ_n will go as follows. For most such ω, the L_n-orbit of ω lies entirely in the bottom half of τ_n, and contains a sequence of τ_{n-1} orbits, where the gaps between them are never greater than h_{n-1}. There may also be a fraction of a gap and a τ_{n-1}-orbit at the beginning and at the end of this L_n-orbit. We estimate the sum

$$\sum_{k=1}^{L_n} X\left(T^k\omega\right)\frac{1}{k}$$

by partitioning the index set $\{k\}_{k=1}^{L_n}$ into four subsets $\{S_i\}_{i=1}^{4}$.

First there is the set S_1 corresponding to the fragments of gaps and $\tau_{n-1}-$orbits at the beginning and end. There can be at most $4h_{n-1}$ such terms, so

$$\left|\sum_{k\in S_1} X\left(T^k\omega\right)\frac{1}{k}\right| \leq \sum_{k=1}^{4h_1}\frac{1}{k}.$$

Second, there is the set corresponding to $\tau_{n-1}-$orbits in this L_n-orbit. In each of these we see an equal number of 1's and -1's, so as before, we can write

$$\sum_{k\in S_2} X\left(T^k\omega\right)\frac{1}{k}$$

as a sum of h_{n-1} alternating series, each of which consists of one term taken from each of the $\tau_{n-1}-$orbits. Each of these alternating series has a sum which, in absolute value cannot exceed 1 so

$$\left|\sum_{k\in S_2} X\left(T^k\omega\right)\frac{1}{k}\right| \le h_{n-1}.$$

Note that this estimate is valid, no matter how many $\tau_{n-1}-$orbits appear in this L_n-orbit.

The third subset corresponds to the alternating filler sequence that appears between the gaps, excluding the extra 1's. Again, since this is an alternating sum of terms of absolute value less than 1, we get

$$\left|\sum_{k\in S_3} X\left(T^k\omega\right)\frac{1}{k}\right| \le 1.$$

Finally the set S_4 corresponds to the extra pairs of 1 that appear in every gap. If we chose L_n large enough compared to h_{n-1}, and h_n large enough compared to L_n, then, since the gaps between $\tau_{n-1}-$orbits are bounded by h_{n-1}, there will be some large number (say M_n) $\tau_{n-1}-$orbits in this L_n-orbits and so the contribution of the extra ones in those gaps will be bounded below by a sum of terms from the harmonic series, selected along an arithmetic sequence:

$$\sum_{k\in S_4} X\left(T^k\omega\right)\frac{1}{k} \ge \sum_{j=1}^{M_n}\frac{1}{4h_{n-1}+j\left(3h_{n-1}\right)}$$

But once h_{n-1} is known, the above three estimates are known, and so M_n can be chosen so that the above harmonic sum far exceeds the sum of these estimates. This tells us how to choose L_n and h_n so as to support the rest of the argument, and the proof is complete. ■

3.3 An Additional Result

Lastly, using a simpler construction, we prove that for all measure preserving and ergodic T there is an r.v. X such that (X, T) is convergent.

Theorem 3.12 *For every ergodic measure preserving transformation T on a probability space (Ω, A, P), there is a r.v. $X : \Omega \to \{1, -1\}$ with $E(X) = 0$, such that $\sum_{k=1}^{\infty} X\left(T^k \omega\right) \frac{1}{k}$ converges a.e.*

Proof: We fix a sequence of towers τ_n for T with base A_n, even height h_n and measure $> 1 - \varepsilon_n$, where $\varepsilon_n \to 0$, such that

1. For each n, $\tau_n \subset \tau_{n+1}$.
2. If we set τ_n' equal to the set of levels of τ_n that are at least n^2 levels from the top and from the bottom, then

$$\sum_{n=1}^{\infty} P\left[\tau_n \backslash \tau_n'\right] < \infty.$$

   Note that these conditions guarantee that for $a.e.\ \omega \in \Omega$, there exists some N so that for all $n \geq N$, $\omega \in \tau_n'$. When $\omega \in A_n$, we refer to the set $\left\{T^i \omega\right\}_{i=0}^{h_n - 1}$ as an $n-$block. We refer to the set $\left\{T^i \omega\right\}_{i=0}^{h_n - 1} \backslash \tau_{n-1}$ as $n-$filler.

3. Each $(n-1)-$block in τ_n is less than h_{n-1} levels from its neighboring $(n-1)-$blocks (or from the end of the n block, if it is the first or last $(n-1)-$block in its $n-$block).

We define X inductively as follows:

For $\omega \in A_1$ and $0 \leq i \leq h_1 - 1$ we set $X\left(T^i \omega\right) = (-1)^i$.

Having defined X on τ_{n-1} we extend it to τ_n by requiring that for each $\omega \in A_n$, the sequence of values $\left\{X\left(T^i \omega\right)\right\}_i$ for $0 \leq i \leq h_n - 1$, restricted to the $n-$filler, forms an alternating sequence of $\{1, -1\}$, starting with 1.

Note that, since the h_n are even, the number of terms in the $n-$filler is even, and so the alternating sequence of values of X along the $n-$filler begins with 1 and ends with -1. Moreover, the function X on an $n-$block takes the same number of 1's as -1's.

We claim that (T, X) is convergent. That is, for almost every $\omega \in \Omega$

$$\sum_{k=1}^{\infty} X\left(T^k \omega\right)\left(\frac{1}{k}\right)$$

converges.

Fix $\omega \in \Omega$ such that for all sufficiently large n, $\omega \in \tau_n'$. Again, this can be done for $a.e.\ \omega$. We wish to show that the series

$$\sum_{k=1}^{\infty} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$$

is Cauchy. That is, we fix $\varepsilon > 0$ and argue that for all sufficiently large $M < M'$, the sums

$$\left|\sum_{k=1}^{M} X\left(T^k\omega\right)\left(\frac{1}{k}\right) - \sum_{k=1}^{M'} X\left(T^k\omega\right)\left(\frac{1}{k}\right)\right| < \epsilon.$$

Choose N so that for all $n \geq N$, $\omega \in \tau_n'$, and so that $\sum_{j=N+1}^{\infty} \frac{1}{j^2} < \varepsilon$.

For large M (or M'), this sum

$$\sum_{k=1}^{M} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$$

can be broken into the following pieces:

We let $S_1 = \left\{k \in [1, M] : k \text{ is less than the first } k' \text{ when } T^{k'}\omega \in A_N\right\}$

We let $S_2 = \left\{ \begin{array}{c} k \in [1, M] : k \text{ is between the first } k' \text{ such that } T^{k'}\omega \in A_N, \text{ and the last } k' \text{ such that} \\ T^{k'}\omega \text{ is on the top level of an } N-\text{block and } T^{k'}\omega \in \tau_N. \end{array} \right\}$

In other words, S_2 is the set of integers in $[1, M]$ where the orbit if ω is in complete $N-$blocks.

We let $S_3 = \left\{k \in [1, M] : T^k\omega \in \tau_N \text{ after the last full } N-\text{block}\right\}$

Finally, for $j \in \{N+1, N+2, N+3, ...\}$, we let $F_j = \left\{k \in [1, M] : T^k\omega \text{ is in the } j-\text{filler}\right\}$.

Then we obtain the following results:

1. As long as M is greater than the time of first entry into A_N, then

$$\sum_{k \in S_1} X\left(T^k\omega\right)\frac{1}{k}$$

   is a fixed contribution to the sum, i.e. the same for all $M' \geq M$.

2. The sum

$$\sum_{k\in S_2} X\left(T^k\omega\right)\frac{1}{k}$$

is the sum of h_N convergent alternating series, each of which is obtained by selecting one term from each whole $N-$block in the orbit segment $\left\{T^k\omega\right\}_{k=1}^{M}$. This is the case no matter how large M is. For any $M' > M$, the change in the sum of each of these alternating series would be less than $\frac{1}{M}$ in absolute value, so the change of their contribution collectively would be less than $\frac{h_N}{M}$. Thus, if M is sufficiently large, this change will be less than ε.

3. The sum

$$\sum_{k\in S_3} X\left(T^k\omega\right)\frac{1}{k}$$

is in absolute value less than

$$\sum_{k=1}^{h_N}\frac{1}{L+k}$$

where L is the last $k \in [1, M]$ with $T^k\omega \in A_N$. So if M is sufficiently large, this quantity is less than ε.

4. For each $j \geq N$, the contribution

$$\sum_{k\in F_j} X\left(T^k\omega\right)\frac{1}{k}$$

is an alternating series which in absolute value can't be bigger than the absolute value of its first term, and by the choice of N, the absolute value of the first term is at most $\frac{1}{j^2}$. So the contribution from all these sums is less than

$$\sum_{j=N}^{\infty}\frac{1}{j^2}$$

which is less than ε, by our choice of N.

Thus, the two sums

$$\sum_{k=1}^{M} X\left(T^k\omega\right)\left(\frac{1}{k}\right) \text{ and } \sum_{k=1}^{M'} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$$

differ by less than a multiple of ε, when M and M' are sufficiently large. Since ε was arbitrary, the series $\sum_{k=1}^{\infty} X\left(T^k\omega\right)\left(\frac{1}{k}\right)$ converges. ■

4 Conclusion and Further Questions

Random series is an exciting topic to study. It is an amalgamation of probability, analysis, and topology, in which each question leads to hundreds of others. In both the independent process case and the process induced by an ergodic and measure preserving transformation case, we only see the remarkable characteristic of *a.s.* convergence or *a.s.* divergence.

Another direction we could take that would lead to various other puzzling questions is to determine the distribution of the series we considered. In the independent case, it seems that very little is known. We showed *a.s.* convergence: but where are the values of the sum likely to be? Byron Schmuland of the University of Alberta further examined the properties of the random harmonic series, and showed some of the properties that the random variable that the series converges to:

$$X(\omega) = \sum_{k=1}^{\infty} \epsilon_k(\omega) \left(\frac{1}{k}\right).$$

In particular, the probability density function of this random variable evaluated at 2 or at -2 takes a value differing from $\frac{1}{8}$ by less than 10^{-42}. Schmuland's paper, item 5 in the references, shows why this probability is so close to, but not exactly, $\frac{1}{8}$.

We have seen that every ergodic measure preserving transformation (T, Ω, P) admits a measurable function $X : \Omega \to \{\pm 1\}$ with $EX = 0$ such that (T, X). A natural follow-up question would be to ask which behavior is "typical". This question can be made precise as follows.

Recall that a subset of a topological space Z is called first category if it is contained in the countable union of closed sets with empty interior. Sets which are not first category are called second category. A set is first category precisely when its complement contains a countable intersection of open, dense sets. A theorem of Baire asserts that in a complete metric space, a countable intersection of open, dense sets must be dense. Such a set (a dense G_δ in a complete metric space Z) may be viewed as a *l*arge subset of Z in a topological sense, and a property of points in Z viewed as *t*ypical if the set of points having that property contains a dense G_δ.

The set Z of all measurable $X : \Omega \to \{\pm 1\}$ with $E(X) = 0$ is a closed subset of $L^1(\Omega, P)$ and so is a complete metric space with respect to the L^1 metric: $d(X_1, X_2) = \int_\Omega |X_1 - X_2| \, dP$. This space Z is partitioned into two subsets, $Z_c = \{X \in Z : (T, X) \text{ is convergent}\}$ and $Z_d = \{X \in Z : (T, X) \text{ is divergent}\}$. By the Baire category theorem, it can't be the case that both of these are first category in Z; at least one of these sets must be second category and conceivably both could be. We could investigate which of the remaining three possibilities could occur.

Finally, we have restricted our attention, for the most part to sums of the form $\sum_{k=1}^{\infty} a_k \left(\frac{1}{k}\right)$, where $a_k \in \{1, -1\}$, but it would be natural to investigate series in which the coefficients are more genereral complex sequences, where the decreasing sequence $\left\{\frac{1}{k}\right\}$ is replaced by some other sequence $\{b_k\}_{k=1}^{\infty}$ of positive terms where $\sum b_k$ diverges, or where we could also allow doubly infinite series:

$$\sum_{k=-\infty}^{\infty} a_k b_k.$$

Each of these variants would lead to significantly different problems.

References

[1] Folland, G. (1999). *Real analysis: Modern techniques and their applications.* (2nd ed.). New York, NY: John Wiley & Sons, Inc.

[2] Halmos, P. (1956). *Lectures on ergodic theory.* New York, NY: Chelsea Pub Co.

[3] Loève, M. (1978). *Probability theory I.* (4th ed., Vol. I). Ann Arbor, MI: Springer-Verlag New York, Inc.

[4] Munkres, J. (2000). *Topology.* (2nd ed.). Upper Saddle River, NJ: Prentice Hall, Inc.

[5] Schmuland, B. (2003). *Random harmonic series.* Mathematics, University of Alberta, Alberta, Canada.
Retrieved from http://www.stat.ualberta.ca/people/schmu/preprints/rhs.pdf

[6] Shields, P. (1973). *Theory of bernoulli shifts.* Chicago: University of Chicago.

[7] Thouvenot, J.P., Weiss, B. (2012). *Limit laws for ergodic processes.* Stochastics and Dynamics (Vol. XII).

[8] Walters, P. (1982). *An introduction to ergodic theory.* Ann Arbor, Michigan : Springer-Verlag New York, Inc.