Technische Universiteit Delft
Faculteit Elektrotechniek, Wiskunde en Informatica
Delft Institute of Applied Mathematics

Asymptotic distribution of the number of monochromatic arithmetic progressions in random colourings of the integers

(Nederlandse titel: Asymptotische verdeling van het aantal monochromatische rekenkundige progressies in willekeurige kleuringen van de gehele getallen)

Verslag ten behoeve van het
Delft Institute of Applied Mathematics
als onderdeel ter verkrijging

van de graad van

BACHELOR OF SCIENCE
in
TECHNISCHE WISKUNDE

door

ABDELLAH RISSALAH

Delft, Nederland
Juli 2019

BSc verslag TECHNISCHE WISKUNDE

Asymptotic distribution of the number of monochromatic arithmetic progressions in random colourings of the integers

(Nederlandse titel: Asymptotische verdeling van het aantal monochromatische rekenkundige progressies in random kleuringen van de gehele getallen)

ABDELLAH RISSALAH

Technische Universiteit Delft

Begeleider

Prof.dr. F.H.J. Redig

Overige commissieleden

Dr. D.C. Gijswijt

Dr. B. van den Dries

Juli, 2019

Delft

Abstract

In this thesis we research arithmetic progressions in random colourings of the integers. We ask ourselves how many arithmetic progressions are contained in zero density subsets of the integers? And what is the asymptotic distribution of the number of arithmetic progressions? Key motivation for this research are the famous results of Van der Waerden, Szemerédi and Green-Tao.

In chapter 1 we introduce the necessary mathematical background in the form of Ergodic theory. Some key concepts from Ergodic theory are studied and connected to random colourings of the integers. We spent a great deal of time defining various concepts such as zero density sets and the arithmetic progression counting summation. We are able to derive a law of rare events for zero density sets, with the use of moment generating functions.

Secondly, we look at different modes of convergence. Various modes of convergence like convergence in probability, distribution, almost sure convergence and total variation are properly defined. This is done to look at the occurrence of arithmetic progressions in zero density sets. We proof that almost surely, zero density sets, generated by independent colourings, contain infinitely many arithmetic progressions of any finite length. A fun application to a stochastic analogue of the prime numbers is presented.

In chapter 3 we introduce the Chen-Stein method for Poisson approximation. First we build the basic theory behind the method and show it's application for proving Poisson convergence of the length 1 arithmetic progression counting summation. We show that, by utilizing a very general theorem of Arratia, Goldstein & Gordon, that the number of finite length arithmetic progressions is Poisson distributed. This is the main result of this thesis. We end by giving another application of this theorem to obtain a similar result for the occurrence of very large progressions.

Lastly, we try to apply the transfer matrix method as an alternative method for showing Poisson convergence. These transfer matrices are introduced, because they provide explicit expressions and convenient ways of computation for the moment generating function of the arithmetic progression counting summation. We obtain a rigours result for arithmetic progressions of length 1, and a partial result for progressions of length 2. We emphasize that more research is needed on this topic.

Preface

In front of you is my Bachelor thesis 'Asymptotic distribution of the number of monochromatic arithmetic progressions in random colourings of the integers'. Written as a conclusion of my undergraduate study Applied Mathematics at the TU Delft.

Applied Mathematics at the TU Delft introduced me to numerous branches of mathematics. Favorites of mine include linear algebra, analysis and probability theory. Upon choosing my research subject I've kept these topics in mind.

The choice to research patterns in the integers was not very hard to make. During my study I have already read into some of the most famous works of Ben Green and Terence Tao. Their most famous proof for arithmetic progressions in the prime numbers inspired me to make this choice.

I'm very satisfied with my research subject. During this process I've learned a lot about what it means to be a mathematician. Getting stuck on hard problems and looking for various new ways of tackling these problems, turned out to be very fun and instructive.

I would like to thank my advisor Prof.dr. F.H.J Redig for his support and guidance during my Bachelor thesis. All your valuable insights, feedback and interesting discussions are very much appreciated.

Contents

Chapter 1

Introduction

Problems in number theory have always been a very fruitful branch of mathematics. Usually a lot of new mathematics arises when a previously unresolved theorem is proven. Very often proofs in number theory branch into different areas of mathematics. Problems involving subsets of integers frequently involve techniques from probability theory. In this thesis we are concerned with patterns in large subsets of the integers. Large integer subsets do contain many patterns. Most commonly studied patterns are those concerning arithmetical progressions in large integer subsets. Examples of this kind are the Van der Waerden theorem, Szemerédi and more recently Green-Tao.

An arithmetical progression (AP) starting at a with common difference b is a collection $\{a + bi : i = 1, 2, \ldots, k\} \subset \mathbb{N}$, for $k \in \mathbb{N}$. In this thesis we only consider progressions starting at $a = 0$. We will for look these progressions in so called random subsets of the integers. Random integer subsets are identified with random colourings of the integers. A random colouring, of $k+1$ different colours, is the realization of a sequence of random variables $\{\omega_i : i \in \mathbb{N}\}$, such that ω_i takes on values (or colours) in $\{0, 1, \ldots, k\}$. This type of random colouring partitions the integers into $k+1$ random disjoint subsets. Each of these disjoint subsets corresponds to a random integer subset $\{i : \omega_i = \alpha\}$, where α is one of the colours in $\{0, 1, \ldots, k\}$. As such is the identification of random integer subsets with random colourings.

We will mostly focus on random 2-colourings that are identically and independently distributed. Meaning that we will look at sequences $\{\omega_i : i \in \mathbb{N}\}$ with positive density, i.e with random variables ω_i such that $\mathbb{P}(\omega_i = 1) = p \in (0, 1]$. The Szemerédi theorem states that positive density subsets of the integers contain arbitrarily long arithmetic progressions.

Far less is known for integer subsets of zero density. Zero density sets, that are generated by independent colourings, correspond to sequences $\{\omega_i : i \in \mathbb{N}\}$ such that $\mathbb{P}(\omega_i = 1) = p_i \longrightarrow 0$. Zero density sets are of particular interest to us, because it is still unknown if they in general all contain infinitely many arbitrary length arithmetic progressions. And even less is know about the distribution of the number of progressions. Probably the most remarkable zero density set containing infinitely many arbitrary length arithmetic progressions are the prime numbers (Green-Tao).

In this thesis we will try to answer this question for zero density sets that are generated independently by random 2-colourings. The choice to restrict ourselves to these kinds of zero density sets is, because this creates a lot of additional structure by which we can obtain more precise results.

We will be interested in the number of arithmetic progressions, of a fixed length, far away from 0. Since the probability of observing arithmetic progressions converges to 0 (rare event) we intuitively expect the number of progressions to be approximately Poisson distributed, by the law of rare events. We will count the number of arithmetic progressions far away from 0 with the length k arithmetic progression counting summation.

$$X_n^{(k,\lambda)} = \sum_{i=n}^{r_n(\lambda)} \omega_i \omega_{2i} \cdots \omega_{ki}$$

Here $r_n(\lambda)$ is a natural number that ensures that $\sum_{i=n}^{r_n} p_i \cdots p_{ki} \longrightarrow \lambda > 0$. Most of the results that we obtain will concern the asymptotic distribution of $X_n^{(k,\lambda)}$. A straightforward approach utilizing moment generating functions will yield us a first result for arithmetic progressions of length 1. Secondly an application of the very general theorem of Arratia, Goldstein & Gordon will ensure us that the limiting distribution of $X_n^{(k,\lambda)}$ is that of a Poisson random variable. This theorem will also be used to obtain a similar result for the occurrence of very long arithmetic progressions in the interval $[1, N]$. In the last chapter we introduce a more direct method for computing the moment generating function of $X_n^{(k,\lambda)}$, for $k = 1, 2$. This method has the advantage that it produces the exact form of the moment generating function. It is inspired from the transfer matrix method that is often used in statistical mechanics and dynamical systems theory to compute partition functions.

All of these concepts and results have their mathematical basis in a branch of probability theory called Ergodic theory. This a very broad field and thus we will only limit to some elementary concepts relevant to the subject at hand. A firm understanding of arithmetic progressions, in random subsets of the integers, is rooted in the study of Ergodic theory.

1.1 Ergodic theory

Ergodic theory is a field of study that studies the long term average behaviour of dynamical systems with an invariant measure. The behaviour of these dynamical systems is accounted for by a measurable map $T : X \longrightarrow X$, where X is the collection of all states of the dynamical system. Various important Ergodic theorems include the Poincaré recurrence theorem, Furstenberg & Weiss topological multiple recurrence and Furstenberg multiple recurrence. All these theorems assert something about the long time average behaviour of orbits of T. To give this all a firm mathematical backbone we start with some of the basic definitions. Basic objects of Ergodic theory are dynamical systems, thus we will make this precise.

Definition 1.1.1. *A dynamical system is given by a probability space (X, Σ, μ) and a measure preserving transformation $T : X \longrightarrow X$ with respect to μ, i.e. $\mu(T^{-1}A) = \mu(A)\ \forall A \in \Sigma$. Moreover T is called Ergodic, if $\forall E \in \Sigma\ :\ T^{-1}(E) = E$, then either $\mu(E) = 0$ or $\mu(E) = 1$. This measure preserving dynamical system is denoted by the quadruple $\chi = (X, \Sigma, \mu, T)$.*

Remark. Note that if X is taken to be any finite set and Σ a σ-algebra generated by some subsets of X. Also letting μ be the normalised counting measure on X and taking $T : X \longrightarrow X$ to be a bijection. Then the quadruple $\chi = (X, \Sigma, \mu, T)$ is a measure preserving dynamical system. T is clearly a measurable map and $\forall A \in \Sigma$ we have that $\mu(T^{-1}A) = \frac{|T^{-1}A|}{|X|} = \frac{|A|}{|X|} = \mu(A)$.

At first glance it is not obvious why these dynamical systems are relevant to the subject at hand. The following example illustrates a connection between random colourings and Ergodic theory. In particular it shows how to perform this 'game' of colouring the integers.

Example 1.1.1 (Bernoulli shifts)**.** *We begin by defining a measure preserving system $\chi = (X, \Sigma, \mu, T)$. Let $X := \{0, 1\}^{\mathbb{N}}$, the space of all infinite sequences consisting of only zeros and ones. And let $\Sigma = \mathcal{B}$ be the Borel σ-algebra. Here we have that every cylinder subset of X is open. We note that this Borel σ-algebra is generated by the cylinder sets. To the define the probability measure we first restrict it for the cylinder sets. Let $C_{i_1 \dots i_k}$ be a some arbitrary cylinder set. Then $C_{i_1 \dots i_k} := \{\omega \in X : \omega_{i_1} = \alpha_{i_1}, \dots, \omega_{i_k} = \alpha_{i_k}\}$ for fixed $i \in \mathbb{N}$. And let the measure of $\mu^{\sim}(C_{i_1 \dots i_k}) = \prod_{j=1}^{k} p^{\alpha_i}(1-p)^{1-\alpha_i}$ for all $i \in \mathbb{N}$. By the Dynkin $\pi - \lambda$ theorem we can find an unique extension μ of $\mu^{\sim}$ on the whole Borel σ-algebra. As a consequence of this the random variables $\Omega_i : X \longrightarrow \{0, 1\} \quad \omega \longmapsto \omega_i$ are independent. Now that we have defined the probability space we are just left with finding a measurable and measure preserving map. The (left) shift, $T : X \longrightarrow X$ defined as $(\omega_l)_{l \in \mathbb{N}} \longmapsto (\omega_{l+1})_{l \in \mathbb{N}}$ suffices as such a map. Hence $\chi = (X, \Sigma, \mu, T)$ is a measure preserving system. This system will turn out to be handy for proving the Van der Waerden theorem.* [**1**]

The long term behaviour of a measure preserving systems is described by its orbits. It is important to know these orbits, because they give us the properties of the system. Essentially the orbits describe what the measure preserving map $T : X \longrightarrow X$ does to the points of X.

Definition 1.1.2. *Let $\chi = (X, \Sigma, \mu, T)$ be a measure preserving system and let $x \in X$. Then the orbit of x is the set $\{T^n(x) : n \in \mathbb{N}\}$.*

An interesting property of measure preserving systems is that for most points $x \in X$, the points in it's orbit happen infinitely often. This behaviour is known as recurrence. This recurrence behaviour leads to the question; do orbits return back to the same point? Formally, if $x \in A \subset X$ is a neighbourhood of x, can we find a $n \in \mathbb{N}$ such that $T^n(x) \in A$?

1.2 Recurrence theorems

The first result on recurrence was obtained by Poincaré. His recurrence states that if $T : X \longrightarrow X$ is a measure preserving transformation on probability space (X, Σ, μ) and if $A \subset X : \mu(A) > 0$, then for μ-almost every $x \in A$ we can find a $n \geq 1$ such that $T^{n(x)}(x) \in A$. This theorem was originally applied by Poincaré to Hamiltonian systems [**2**].

For us a generalization, due to Furstenberg and Weiss, for topological probability spaces, is more useful.

Theorem 1.2.1 (Topological Multiple Recurrence, Furstenberg & Weiss)**.** *Let (X, μ) be compact metric space and $T : X \longrightarrow X$ be a continuous dynamical system on X with μ being T-invariant. Then for every $k \in \mathbb{N}$ and $\epsilon > 0$ there exists a $x \in X$ and $n \in \mathbb{N}$ such that $d\left(T^{in}(x), x\right) < \epsilon$ for all $i \in \{1, \cdots, k\}$. Moreover, if there is a $Z \subset X$ that is dense in X, then we can take $x \in Z$.* [**1**]

We will use this theorem to give a proof of the Van der Waerden theorem. We call x recurrent if x is in the closure of its orbit, i.e if $x \in \overline{\{T^n(x) : n \in \mathbb{N}\}}$. Topologically reformulated this means that there exists a sequence $n_k \to \infty$ such that $T^{n_k}x \longrightarrow x$. $x \in X$ is multiple recurrent if there exist multiple measure preserving transformations $T_i : X \longrightarrow X,\ i = 1, 2, \cdots, k$ such that x is recurrent for each of these maps. Van der Waerden is essentially a statement on the existence of a multiple recurrent point [**6**]. Too clarify, this theorem states,

Theorem 1.2.2 (Van der Waerden)**.** *If we colour the integers with m different colours, then we can always find arbitrarily long arithmetic progressions of at least one colour.*

Van der Waerden proved this theorem by using a combinatorial approach. Here we will present a dynamical proof using Furstenbergs theorem.

Proof. [Van der Waerden]. Let $A = \{a_1, a_2, \cdots, a_k\}$ be a collection of colours and let $(z_1, z_2, \cdots) \in A^{\mathbb{N}}$ be a given colouring of the natural numbers. Here z_i denotes the colour of number i. Now let the (left) shift map $T : A^{\mathbb{N}} \longrightarrow A^{\mathbb{N}}$ be given as $(z_1, z_2, z_3, \cdots) \longmapsto (z_2, z_3, z_4, \cdots)$. We also define a metric d on $A^{\mathbb{N}}$ as, $d(x, y) = \frac{1}{w}$, where $w := \inf\{n \in \mathbb{N} | x_n \neq y_n\}$. Note that for all $x, y \in A^{\mathbb{N}} : d(x, y) < 1$ iff $x_1 = y_1$. This implies that for every $x, y \in A^{\mathbb{N}}$ and $m, l \in \mathbb{N} : d\left(T^m(x), T^l(y)\right)$ iff $x_{m+1} = y_{m+1}$.
In particular we have that the arithmetic progression $m, m+n, \cdots, m+kn$ is of the same colour iff $x_m = x_{m+n} = \cdots = x_{m+kn}$, when $x \in A^{\mathbb{N}}$ is colouring of $\mathbb{N}$. By definition of d:

$$d\left(T^{m-1}(x), T^{m-1+in}(x)\right) = d\left(T^{m-1}(x), T^{in}\left(T^{m-1}(x)\right)\right) < 1, \quad i = 1, 2, \cdots, k$$

If we now take $X = \overline{\{T^m(x)\}_{m=0}^{\infty}}$, then X is a compact metric space and T is a continuous dynamical system on X. Also $Z = \{T^m(x)\}_{m=0}^{\infty}$ is dense in X. Furstenberg & Weiss theorem now gives, with $\epsilon = 1$, that $\exists m \in \mathbb{N}$ such that $d\left(T^m(x), T^{in}\left(T^m(x)\right)\right) < 1$, for $i = 1, 2, \cdots, k$. This implies that $x_{m+1} = x_{m+1+n} = \cdots = x_{m+1+kn}$ is a k-length arithmetic progression of the same colour. [**4**] □

The Van der Waerden theorem gives us the existence of arbitrarily long arithmetic progressions in at least one of the m partitions, that is large enough to contain them. But Van der Waerden does not specify those partitions which are large enough. For this we need to first define what we mean for a set to be large enough.

Definition 1.2.1. *For any subset A of $\mathbb{N}$ the (upper) density, $d(A)$, of A is defined as:*

$$d(A) := \limsup_{N \to \infty} \frac{|[1, N] \cap A|}{N}$$

Szemerédi's theorem gives us a sufficient condition, in contrast to Van der Waerden, as to which subsets of $\mathbb{N}$ that contain arithmetic progressions of arbitrary length.

Theorem 1.2.3 (Szemerédi)**.** *Let $A \subset \mathbb{N}$ such that $d(A) > 0$. Then A contains arbitrarily long arithmetical progressions.*

Just as for Van der Waerden's theorem we will use a result from Ergodic theory to prove Szemerédi's theorem. The proof will make use of the Bernoulli shifts as in Example 1.1.1. For clarity we defined $X = \{0,1\}^{\mathbb{N}}$ along with the (left) shift transform $T : X \longrightarrow X$. The following result from Furstenberg is used.

Theorem 1.2.4 (Furstenberg's Multiple Recurrence Theorem)**.** *Let $T : X \longrightarrow X$ be a measure preserving transformation with μ being T-invariant and $\mu(A) > 0$. Then, if $k \geq 3$, there exists a $N \in \mathbb{N}$ such that,*

$$\mu\left(A \cap T^{-1}(A) \cap \cdots \cap T^{-(k-1)N}(A)\right) > 0$$

Remark. *Note that the case $k = 2$ reduces to Poincaré's recurrence theorem.*

With this theorem Furstenberg proved Szemerédi's theorem. A complete proof of this theorem would stray away too much from the topic. For a rigorous proof we refer to [1]. Here we will suffice with a proof for the Bernoulli shifts (Example 1.1.1).

Proof. [Example 1.1.1] Recall that the cylinders $C_i = \{\omega \in X : \omega_i = 1\}$, with measure $\mu(C_i) = p \; \forall i \in \mathbb{N}$, generate the Borel σ-algebra. This means it is sufficient to show Furstenberg's Multiple Recurrence on all cylinder subsets. Let $C_0, C_1, \cdots, C_k$ be cylinder sets, then for a sufficiently large $n \in \mathbb{N}$, we have that the fixed coordinates of the cylinder $T^{-nl}(C_l)$ are distinct. Thus,

$$\mu\left(C_0 \cap T^{-n}(C_1) \cap \cdots \cap T^{-kn}(C_k)\right) = \mu(C_0)\mu(C_1)\cdots\mu(C_k) > 0$$

Hence Furstenberg's Multiple Recurrence Theorem holds for the Bernoulli shifts. □

Now we are ready to tackle Szemerédi's theorem with the help of Furstenberg's multiple recurrence.

Proof. [Szemerédi] Let $X = \{0,1\}^{\mathbb{N}}$ and define $T : X \longrightarrow X$ to be the left shift map. Let $(\omega_n) = (\mathbb{1}_A(n))$, with $\mathbb{1}_A(n)$ being the indicator function of subset $A \subset \{0,1\}$. Define $\mu_k = \frac{1}{k}\sum_{j=0}^{k-1} \delta_{T^j(\omega)}$, with $\delta(\omega)$ being the Dirac measure. Note that $\mu = \lim_k \mu_k$ is a T-invariant probability measure. As to why this holds true we refer to [3]. Next define $Y = \{(\delta_n) : \delta_1 = 1\}$. Y is compact in X, thus $\mu(Y) = \lim_k \mu_k(Y) = \frac{1}{k}|A \cap [1,k]| > 0$. Hence, with Furstenberg's multiple recurrence we can find a N such that $\mu\left(Y \cap T^{-N}(Y) \cap \cdots \cap T^{-(k-1)N}(Y)\right) > 0$. Specifically, there is a $z \in Y \cap T^{-N}(Y) \cap \cdots \cap T^{-(k-1)N}(Y)$. This means $\exists x \in \mathbb{N}$ such that $x, x+N, \cdots, x+(k-1)N \in A$. k is chosen arbitrary, thus A contains arbitrarily long arithmetical progressions. [**1**] □

Szemerédi gives us a 'class' of subsets that contain arbitrary long arithmetical progressions, namely those with positive density. But these are not all subsets of $\mathbb{N}$ containing arbitrary length arithmetical progressions. What can be said for the number of progressions in zero density sets? A famous conjecture of Erdös and Turán gives a sufficient condition.

Conjecture 1.2.1. [*Erdös-Turán*] *Let $A \subset \mathbb{N}$ such that*

$$\sum_{N \in A} \frac{1}{N} = \infty.$$

Then A contains infinitely many arithmetical progressions.

This conjecture remains unproven. Even for arithmetical progressions of length 3 we still don't know if this conjecture holds true.

1.3 Poisson convergence in zero density sets

Let $\{\omega_i : i \in \mathbb{N}\}$ be a sequence of independent random variables such that $\omega_i \sim$Ber(p), for $p > 0$. Then the set $\{i : \omega_i = 1\}$ satisfies the conditions of Szemerédi. And thus it contains infinitely many arithmetic progressions of any finite length. Now we want p to be dependent on i in such a way that the density of the set $\{i : \omega_i = 1\}$ is zero, but nevertheless do have infinitely many arithmetic progressions. Meaning that the black sites ($\omega_i = 1$) become increasingly more rare as we go further. By letting $p_i \longrightarrow 0$ as $i \to \infty$ we make $\{i : \omega_i = 1\}$ a zero density set. The question now becomes if it still contains arbitrary length arithmetical progressions and what is the asymptotic distribution of these progressions?

To obtain an answer for these questions we will first look at arithmetic progressions of length 1. Showing convergence in distribution will require the use of moment generating functions.

Definition 1.3.1. *Let X be a random variable, then the moment generating function of X is defined as,*

$$M_X(t) := \mathbb{E}\left[e^{tx}\right] = \int_{\mathbb{R}} e^{tx} dF_X(x)$$

for $t \in \mathbb{R}$.

Also will be used the following theorem,

Theorem 1.3.1. *Let X and Y be two random variables. If for all $t \in \mathbb{R} : M_X(t) = M_Y(t)$, then $X \overset{d}{=} Y$.*

With this theorem we are ready too find the limiting distribution.

Theorem 1.3.2. *Let $\{\omega_i : i \in \mathbb{N}\}$ be a sequence of independent random variables such that $\omega_i \sim$ Bernoulli(p), where $p > 0$. Set $\sum_{i=1}^{N} p = \lambda \Rightarrow p = \frac{\lambda}{N}$. If $X_N := \sum_{i=1}^{N} \omega_i$, then $X_N \xrightarrow{d} Po(\lambda)$.*

Proof. Theorem 1.3.1 implies that we only have too show convergence of moment generating functions. Hence we start by calculating the moment generating function of X_N.

$$M_{X_N}(t) = \mathbb{E}\left[e^{t\sum_{i=1}^{N}\omega_i}\right] = \prod_{i=1}^{N} \mathbb{E}\left[e^{t\omega_i}\right] = \left(\mathbb{E}\left[e^{t\omega_1}\right]\right)^N$$

The last two equality's hold, because the ω_i are identically and independently distributed. By definition of the expectation we get,

$$M_{X_N}(t) = \left(\mathbb{P}(\omega_1 = 0) + \mathbb{P}(\omega_1 = 1)e^t\right)^N = \left((1 - \frac{\lambda}{N}) + \frac{\lambda}{N}e^t\right)^N = \left(1 + \frac{\lambda(e^t - 1)}{N}\right)^N$$

Note that when $N \to \infty$ we obtain $\left(1 + \frac{\lambda(e^t-1)}{N}\right)^N \longrightarrow \mathrm{e}^{\lambda(e^t-1)}$, by definition of the exponential function. And this concludes the proof, because $\mathrm{e}^{\lambda(e^t-1)}$ is indeed the moment generating function of a random variable Y with $\mathcal{L}(Y) = Po(\lambda)$. □

This theorem is actually widely known as the law of rare events. It roughly states that when considering a large number N of independent events with small success probabilities $\frac{\lambda}{N}$, then the limiting distribution of the total number of successful events is Poisson distributed with parameter $\lambda = Np$.
In this thesis we are only interested in zero density sets. We want to know if we can derive a similar result, theorem 1.3.2, for zero density sets. The setup is as such. Consider a zero density set $A = \{i : \omega_i = 1\}$. Then each $\omega_i \sim \mathrm{Ber}(p_i)$, with $p_i \longrightarrow 0$ as $i \to \infty$. Just as before we suspect the asymptotic distribution to be that of a Poisson random variable with parameter λ. But how to choose λ? For the law of rare events to hold λ needs to approximately be the expected number of successful events, in a certain interval. Thus define $\lambda > 0$ and let $\sum_{i=n}^{r_n} p_i$, where $r_n = \inf\{k > n : \sum_{i=n}^{k} p_i \geq \lambda\}$. This way we ensure that $\sum_{i=n}^{r_n} p_i \longrightarrow \lambda$. It is not directly obvious that this should be the case. Hence this deserves to be its own lemma.

Lemma 1.3.1. *Assume $p_i \longrightarrow 0$ as $i \to \infty$. Let $\lambda \in \mathbb{R}_{\geq 0}$ and define $r_n(\lambda) = \inf\{k > n : \sum_{i=n}^{k} p_i p_{2i} \cdots p_{ki} \geq \lambda\}$. Then, $\lim_{n\to\infty} \sum_{i=n}^{r_n(\lambda)} p_i p_{2i} \cdots p_{ki} = \lambda$ for all $k \geq 1$.*

Proof. By definition we have that $\sum_{i=n}^{r_n(\lambda)} p_i p_{2i} \cdots p_{ki} \geq \lambda$ and $\sum_{i=n}^{r_n(\lambda)-1} p_i p_{2i} \cdots p_{ki} < \lambda$. Hence,

$$0 \leq \left| \sum_{i=n}^{r_n(\lambda)} p_i p_{2i} \cdots p_{ki} - \lambda \right| \leq \sum_{i=n}^{r_n(\lambda)} p_i p_{2i} \cdots p_{ki} - \sum_{i=n}^{r_n(\lambda)-1} p_i p_{2i} \cdots p_{ki} = p_{r_n(\lambda)} \longrightarrow 0 \quad \text{as } n \to \infty$$

In conclusion, $\lim_{n\to\infty} \sum_{i=n}^{r_n(\lambda)} p_i p_{2i} \cdots p_{ki} = \lambda$. □

Another thing underlining theorem 1.3.2 is that arithmetical progressions, no matter how far we go along, can always be observed. This is what we know from Szemerédi's theorem. For the zero density sets Szemerédi cannot be used, thus we cannot assume that $\sum_{i=n}^{N} \omega_i \omega_{2i} \cdots \omega_{ki} \xrightarrow{a.s.} \infty$ as $N \to \infty$. In the next chapter we extensively focus on this problem, but for now we will just assume this to be the case.

In order to obtain results we throughout assume the following.

Assumption 1.3.1.

$$\sum_{i=1}^{\infty} p_i p_{2i} \cdots p_{ki} = \infty, \qquad \text{for any } k \in \mathbb{N}.$$

This assumption is partly motivated from sets like the set of prime numbers. The prime numbers are a zero density subset of the natural numbers. We can make a stochastic analogue of the prime numbers by taking the same setup as with A. If we want to imitate the prime numbers we choose $p_i = \frac{1}{\log(i)}$. This choice is by analogy of the prime number theorem. Observe that with this choice of p_i, assumption 1.3.1 holds true, because $\sum_{i=1}^{\infty} \frac{1}{\log(i)} = \infty$.

This assumption is also needed to show that $\sum_{i=n}^{N} \omega_i \omega_{2i} \cdots \omega_{ki} \xrightarrow{a.s.} \infty$. Because if the assumption 1.3.1 does not hold then,

$$\sum_{i=1}^{\infty} p_i \cdots p_{ki} < \infty \implies \exists M \in \mathbb{N} \text{ such that } \mathbb{P}\left(\sum_{i=M}^{\infty} \omega_i \cdots \omega_{ki} = 0\right) = 1.$$

This means that assumption 1.3.1 is indeed for good reason. In the following chapter we actually come to the conclusion that assumption 1.3.1 and $\sum_{i=1}^{N} \omega_i \cdots \omega_{ki} \xrightarrow{a.s.} \infty$ are equivalent.

Now we are ready to propose a law of rare events for zero density sets. Unlike in theorem 1.3.2 we now have probabilities $p_i \longrightarrow 0$ as $i \to \infty$. This highlights the important difference that p_i is now only dependent on the place i and not on total number of progression N.

Theorem 1.3.3. *Let $\{\omega_i : i \in \mathbb{N}\}$ be a sequence of independent random variables such that $\omega_i \sim$ Bernoulli(p_i). Assume that $p_i \longrightarrow 0$ as $i \to \infty$ and $\sum_{i=1}^{\infty} p_i = \infty$. Let $\lambda > 0$ and define $r_n(\lambda) = \inf\{k > n : \sum_{i=n}^{k} p_i \geq \lambda\}$. If $X_n^{(1)} := \sum_{i=n}^{r_n} \omega_i$, then $X_n^{(1)} \xrightarrow{d} Po(\lambda)$ as $n \to \infty$.*

Proof. We will show convergence in distribution of $X_n^{(1)}$ to a Poisson distribution by calculating the limit of the moment generating function of $X_n^{(1)}$. For convenience we calculate $\log(M_{X_n^{(1)}}(t))$. This yields,

$$\log\left(M_{X_n^{(1)}}(t)\right) = \log\left(\mathbb{E}\left[e^{t\sum_{i=n}^{r_n}\omega_i}\right]\right) = \log\left(\prod_{i=n}^{r_n}\mathbb{E}\left[e^{t\omega_i}\right]\right) = \sum_{i=n}^{r_n}\log\left(1 + p_i(e^t - 1)\right)$$

Equality holds here, because the ω_i are independent from one another. To show that this quantity converges to the desired result, we will find a suiting lower and upper bound. For this first note that by Taylor's theorem $x - \frac{1}{2}x^2 \leq \log(1+x) \leq x$ holds for all $x \in [0, \infty)$. Hence we derive,

$$\sum_{i=n}^{r_n} p_i(e^t - 1) - \frac{1}{2}p_i^2(e^t-1)^2 \leq \sum_{i=n}^{r_n}\log\left(1 + p_i(e^t - 1)\right) \leq \sum_{i=n}^{r_n} p_i(e^t - 1)$$
$$\implies 0 \leq \left|\sum_{i=n}^{r_n}\log\left(1 + p_i(e^t - 1)\right) - \sum_{i=n}^{r_n} p_i(e^t - 1)\right| \leq \frac{1}{2}(e^t-1)^2\sum_{i=n}^{r_n} p_i^2$$

Note that $\sum_{i=n}^{r_n} p_i^2 \leq \left(\sup_{n \leq i \leq r_n} p_i\right)\sum_{i=n}^{r_n} p_i \longrightarrow 0 \cdot \lambda = 0$, because by assumption $p_i \longrightarrow 0$ as $i \to \infty$. Because the difference between $\log\left(M_{X_n^{(1)}}(t)\right)$ and $\sum_{i=n}^{r_n} p_i(e^t - 1)$ becomes arbitrarily small, we can derive the law of $X_n^{(1)}$ from $\sum_{i=n}^{r_n} p_i(e^t - 1)$. Hence,

$$\lim_{n\to\infty}\log\left(M_{X_n^{(1)}}(t)\right) = \lim_{n\to\infty}\sum_{i=n}^{r_n} p_i(e^t - 1) = \lambda(e^t - 1)$$

Thus we have shown convergence in distribution of $X_n^{(1)}$, because if $M_{X_n^{(1)}}(t) \xrightarrow{n\to\infty} e^{\lambda(e^t-1)}$, then $X_n^{(1)} \xrightarrow{d} Po(\lambda)$. □

Even for this relatively simple case of arithmetic progressions of length 1 we needed to do a lot more work to get the desired result of convergence in distribution. Although this is to be expected as now $\{\omega_i : i \in \mathbb{N}\}$ are not identically distributed. An essential step in this proof is that each term of $X_n^{(1)}$ is independent from all other terms. This allowed us to move the product outside of the expectation. This step will fail if we start looking at longer progressions. Thus we can not take the same approach to prove similar results for longer progressions. Even for progressions of length 2 this step fails. The length 2 progression counting summation is defined as $X_n^{(2)} := \sum_{i=n}^{r_n} \omega_i\omega_{2i}$. Then terms like $\omega_n\omega_{2n}$ and $\omega_{2n}\omega_{4n}$ are clearly not independent.

Different methods of tackling these types of problems, will be studied in chapters 3 and 4. Here we use more abstract techniques such as the Chen-Stein method and the Transfer matrix method. The first one will give a us a very strong, but not very practical result. On the other hand the Transfer matrices will turn out to be very tangible.

Chapter 2

Infinitely many progressions in zero density sets

In this thesis we want to show weak convergence of the AP-k counting summation, $X_n^{(k,\lambda)}$. For convenience we just write $X_n^{(k)}$. As we have seen in the previous chapter, this requires assuming that zero density sets contain arbitrary length arithmetic progressions. To prove this assumption and to obtain results for the asymptotic behaviour of $X_n^{(k)}$ we need various concepts of limits of random variables. Familiar modes of convergence of random variables include weak convergence, convergence in probability and almost sure convergence. These will be discussed in this section. Another less familiar convergence is the total variation distance. The total variation distance measures the accuracy of a approximation. In our case we measure the total variation distance between the law of $X_n^{(k)}$ and a Poisson random variable with parameter $\lambda \longleftarrow \sum_{i=n}^{r_n} p_i \cdots p_{ki}$.

2.1 Modes of convergence

Consider $(X_N)_{N\in\mathbb{N}}$ a sequence of random variables on a probability space $(\Omega, \Sigma, \mathbb{P})$. We want to know the difference in probability between each X_N and a certain random variable X. If this difference in probabilities becomes negligible as N increases, we can speak of a limit.

Definition 2.1.1 (Convergence in probability). *A sequence of random variables $(X_N)_{N\in\mathbb{N}}$ converges to random variable X in probability iff for all $\epsilon > 0$,*

$$\lim_{N\to\infty} \mathbb{P}\left(\omega : |X_N(\omega) - X(\omega)| > \epsilon\right) = 0$$

The following example will illustrate this concept quite neatly.

Example 2.1.1. *Let $(X_N)_{N\in\mathbb{N}}$ be a sequence of independent random variables such that $X_N \sim$ Ber(p_N). Where p_N is a real valued sequence such that $p_N \longrightarrow 0$. We can show that $X_N \xrightarrow{\mathbb{P}} 0$. Let $\epsilon > 0$. We note by Markov's inequality that $\mathbb{P}\left(|X_N - 0|\right) \leq \frac{p_N}{\epsilon} \longrightarrow 0$ as $N \to \infty$. This implies that $\lim_{N\to\infty} \mathbb{P}\left(|X_N - 0|\right) = 0$, and thus $X_N \xrightarrow{\mathbb{P}} 0$.*

A stronger notion of convergence is almost sure convergence. This type of convergence is the same as pointwise convergence almost everywhere, as is know in analysis. We will use this mode of convergence to prove that zero density sets contain arbitrary length arithmetic progressions. But first a formal definition.

Definition 2.1.2 (Almost sure convergence)**.** *A sequence of random variables $(X_N)_{N\in\mathbb{N}}$ converges to a random variable X almost surely (or with probability 1) iff,*

$$\mathbb{P}\left(\omega : \lim_{N\to\infty} X_N(\omega) = X(\omega)\right) = 1$$

This definition is not always easy to work with. Luckily we can often use the following lemmas.

Lemma 2.1.1 (Borel-Cantelli)**.** *Let $(A_N)_{N\in\mathbb{N}}$ be a sequence of events in some σ-algebra Σ. Then if,*

$$(1)\quad \sum_{N=1}^{\infty} \mathbb{P}(A_N) < \infty \Longrightarrow \mathbb{P}\left(\limsup_{N\to\infty} A_N\right) = 0$$

$$(2)\quad \sum_{N=1}^{\infty} \mathbb{P}(A_N) = \infty \text{ and if } \{A_N : N \in \mathbb{N}\} \text{ are independent } \Longrightarrow \mathbb{P}\left(\limsup_{N\to\infty} A_N\right) = 1$$

Lemma 2.1.2. *Let $(X_N)_{N\in\mathbb{N}}$ be a sequence of random variables and let X be some other random variable. If $\sum_{N=1}^{\infty} \mathbb{P}(|X_N - X| > \epsilon)$ converges for every $\epsilon > 0$, then $X_N \xrightarrow{a.s.} X$.*

Using Lemma 2.1.2 we can illustrate an example of almost sure convergence.

Example 2.1.2. *Let $(X_N)_{N\in\mathbb{N}}$ be a sequence of independent random variables such that,*

$$\mathbb{P}(X_N = k) = \begin{cases} 1 - \frac{1}{N^2} & k = 0 \\ \frac{1}{N^2} & k = 1 \\ 0 & \text{else} \end{cases}$$

We will show that $X_N \xrightarrow{a.s.} 0$. Let $\epsilon > 0$. Observe that,

$$\sum_{N=1}^{\infty} \mathbb{P}(|X_N - 0| > \epsilon) = \sum_{N=1}^{\infty} \frac{1}{N^2} = \frac{\pi^2}{6} < \infty$$

From Lemma 2.1.2 we can now conclude that $X_N \xrightarrow{a.s.} 0$.

Almost sure convergence is a stronger notion of convergence than convergence in probability. Almost sure convergence implies convergence in probability, but the converse of this is not true. We have seen in Example 2.1.1 that X_N, defined as such, converges in probability. If we take $p_N = \frac{1}{N^2}$, then, as we have shown in the previous example, X_N also converges almost surely. But setting $p_N = \frac{1}{N}$ will not result almost sure convergence. Reason for this is that $\sum_{N=1}^{\infty} \frac{1}{N}$ diverges. Implying that, by Borel-Cantelli, $\mathbb{P}(X_N = 1 \text{ i.o}) = 1$. Thus X_N cannot converge almost surely to 0.

In the introduction we alluded to the fact that we will show that the asymptotic distribution of the AP-k counting summation $X_n^{(k)} = \sum_{i=n}^{r_n} \omega_i \cdots \omega_{ki}$, is that of a Poisson random variable. This is yet another mode of convergence. In general if the distribution of some sequence of random variables converges we say that this sequence converges weakly or converges in distribution.

Definition 2.1.3. *(Convergence in distribution) Let $(X_N)_{N\in\mathbb{N}}$ be a sequence of random variables with distribution functions $F_{X_N}(x)$, and let X be some other random variable with distribution function $F_X(x)$. Then $(X_N)_{N\in\mathbb{N}}$ converges in distribution to X if,*

$$\lim_{N\to\infty} F_{X_N}(x) = F_X(x) \quad \text{for all } x \text{ where } F_X \text{ is continuous}$$

It is important to note that convergence in distribution is equivalent to,

$$\lim_{N\to\infty} \mathbb{E}\left[f\left(X_N\right)\right] = \mathbb{E}\left[f\left(X\right)\right]$$

for all bounded and continuous functions $f : \mathbb{R} \longrightarrow \mathbb{R}$. This usually is referred to as weak convergence.

As the name suggest, weak convergence is an even weaker form of convergence. Convergence in probability implies convergence in distribution, but the converse of this is generally not true. This is best illustrated in the following example.

Example 2.1.3. *Define random variables X_N,*

$$X_N = \begin{cases} -X & N \bmod 2 \equiv 0 \\ X & N \bmod 2 \equiv 1 \end{cases}$$

Where we let X to be a random variable with distribution function $F_X : [-1,1] \longrightarrow \mathbb{R}$ defined as,

$$F_X(x) = \begin{cases} x+1 & x \in [-1,0) \\ 1-x & x \in [0,1] \\ 0 & \text{else} \end{cases}$$

Note that this distribution is a symmetric function about 0. We claim that $X_N \xrightarrow{d} X$. For odd values of N this is clear. For even values of N note that $F_{X_N} := \mathbb{P}\left(-X \leq x\right) = \mathbb{P}\left(X \geq -x\right) = \mathbb{P}\left(X \leq x\right) := F_X$. The last step is justified, because F_X is symmetric about 0. Hence $X_N \xrightarrow{d} X$. However we now claim that we do not have convergence in probability. For this first observe that for even N : $|X_N - X| = 2\,|X|$. And thus if we pick an $\epsilon \in [0,1]$. We get that $\mathbb{P}\left(|X_N - X| > \epsilon\right) = \mathbb{P}\left(2\,|X| > \epsilon\right) = \mathbb{P}\left(|X| > \frac{\epsilon}{2}\right) = \left(1 - \frac{\epsilon}{2}\right)^2 \neq 0$. And thus $\mathbb{P}\left(|X_N - X| > \epsilon\right) \not\longrightarrow 0$. We conclude that $X_N \overset{\mathbb{P}}{\not\to} X$.

The last mode convergence we need is the so called total variation distance. The total variation distance is a metric for how close two probability measures are. Formally we define this as such,

Definition 2.1.4 (Total variation distance)**.** *The total variation distance between two probability measures, $\mathbb{P}$ and $\mathbb{Q}$, on some σ-algebra Σ, we define as,*

$$d_{TV}\left(\mathbb{P}, \mathbb{Q}\right) := \sup_{A\in\Sigma} |\mathbb{P}(A) - \mathbb{Q}(A)|$$

This definition can be used to show that a sequence of random variables $\left(X_N\right)_{N\in\mathbb{N}}$ converges to a random variable variable X, in total variation. When, $\sup_{A\in\Sigma} |\mathbb{P}(X_N \in A) - \mathbb{Q}(X \in A)| \longrightarrow 0$ as $N \to \infty$, we precisely have convergence in total variation. This convergence mode is very useful, because it lets us know the difference between probability measures. The total variation distance is a convergence mode that is probabilistic in nature. Essentially it gives us the maximal distance between two probability measures.

Do note that convergence in total variation is a stronger notion of convergence than weak convergence. Meaning that if the total variation distance tends to 0 as $N \to \infty$, then we have also have convergence in distribution. This is of course very useful, and thus we will make extensively use of this fact in following chapters. The converse of this is not true. Example 2.1.4 illustrates this fact quite neatly. In chapter 3 we will discuss the Chen-Stein method and here we will come back to the total variation distance.

Example 2.1.4. *Define independent and identically distributed random variables X_N taking on values -1 and +1. By the central limit theorem we know that,*

$$S_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \xrightarrow{d} N(0,1)$$

However, $d_{TV}(S_N, Z) = 1$ for every $N \in \mathbb{N}$. Hence we do not have convergence in total variation distance. [**8**]

2.2 Zero density sets with infinitely many progressions

With these different modes of convergence we have all the right tools to show that zero-density sets contain infinitely many arbitrary long arithmetic progressions. We remind the reader of the setup. Consider a sequence $\{\omega_i : i \in \mathbb{N}\}$ of independent random variables with $\mathbb{P}(\omega_i = 1) = p_i$ and $\mathbb{P}(\omega_i = 0) = 1 - p_i$. We require that $p_i \longrightarrow 0$ as $i \to \infty$. With these random variables we can count the number of k length arithmetic progressions. Note that an arithmetic progression of length k precisely occurs when the numbers $i, 2i, \ldots, ki$ are all black sites, i.e precisely when $\omega_i = \omega_{2i} = \cdots = \omega_{ki} = 1$. This event co insides with the event $\omega_i\omega_{2i}\cdots\omega_{ki} = 1$. Because if the numbers $i, 2i, \ldots, ki$ would contain white sites than at least one of the ω_i equates to 0. Using this handy product we count the number of arithmetic progressions, in the interval $[n, N]$, by simply summing over this interval. The number of arithmetic progression of length k is thus given by $\sum_{i=n}^{N} \omega_i\omega_{2i}\cdots\omega_{ki}$. To say that a set of zero density contains infinitely many k length arithmetic progressions means that $\sum_{i=n}^{N} \omega_i\omega_{2i}\cdots\omega_{ki} \xrightarrow{a.s.} \infty$. In order to obtain this result we need assumption 1.3.1 to hold.

Theorem 2.2.1. *Let $\{\omega_i : i \in \mathbb{N}\}$ be a sequence of independent random variables such that $\omega_i \sim$ Bernoulli(p_i). Assume that $p_i \longrightarrow 0$ as $i \to \infty$ and $\sum_{i=1}^{\infty} p_i p_{2i} \cdots p_{ki} = \infty$, then $\forall n \in \mathbb{N}$: $\sum_{i=n}^{N} \omega_i\omega_{2i}\cdots\omega_{ki} \xrightarrow{a.s.} \infty$ as $N \to \infty$.*

Proof. Under these assumptions we can indeed show that $\sum_{i=n}^{N} \omega_i\omega_{2i}\cdots\omega_{ki} \xrightarrow{a.s.} \infty$ as $N \to \infty$. For this we show that $W_N := \sum_{i=1}^{N} \omega_i\omega_{2i}\cdots\omega_{ki} \xrightarrow{a.s.} \infty$ as $N \to \infty$. To see that this is sufficient note that, $\sum_{i=n}^{N} \omega_i\omega_{2i}\cdots\omega_{ki} \geq W_N - n$, because for each $i \in \mathbb{N}$: $\omega_i\omega_{2i}\cdots\omega_{ki} \in \{0,1\}$.
Almost sure convergence of W_N to infinity means that the product $\omega_i\omega_{2i}\cdots\omega_{ki}$ is equal to 1 infinitely often. Formally this means that $\mathbb{P}(\forall n \in \mathbb{N}\, \exists i \geq n : \omega_i\omega_{2i}\cdots\omega_{ki} = 1) = 1$. Or equivalently $\mathbb{P}(W_N > N) \longrightarrow 1$ as $N \to \infty$. We will prove the theorem with the so-called second moment bound.

Lemma 2.2.1 (Payley-Zygmund inequality)**.** *Let Z be a positive random variable such that $Var(Z) < \infty$. If $\theta \in [0, \infty]$, then*

$$\mathbb{P}(Z > \theta\mathbb{E}[Z]) \geq (1-\theta)^2 \frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]}$$

Proof. Note that we can write, $\mathbb{E}[Z] = \mathbb{E}\left[Z\mathbb{1}_{\{Z \leq \theta\mathbb{E}[Z]\}}\right] + \mathbb{E}\left[Z\mathbb{1}_{\{Z > \theta\mathbb{E}[Z]\}}\right]$. The first term is at most $\theta\mathbb{E}[Z]$ and for the second term we can use the Cauchy-Schwarz inequality. Hence,

$$\mathbb{E}[Z] \leq \theta\mathbb{E}[Z] + \sqrt{\mathbb{E}[Z^2]\,\mathbb{E}\left[\mathbb{1}^2_{\{Z>\theta\mathbb{E}[Z]\}}\right]} = \theta\mathbb{E}[Z] + \sqrt{\mathbb{E}[Z^2]\,\mathbb{P}(Z > \theta\mathbb{E}[Z])}$$

By squaring and rearranging the terms in this inequality we get the Payley-Zygmund inequality. □

To conclude almost sure convergence of W_N from Payley-Zygmund we need the following to hold,

$$(1) \quad (\delta_N)_N := \left(\frac{\mathbb{E}\left[W_N\right]^2}{\mathbb{E}\left[W_N^2\right]}\right)_N \longrightarrow 1 \quad \text{as} \quad N \to \infty$$

$$(2) \quad (\sigma_N)_N := (\mathbb{E}\left[W_N\right])_N \longrightarrow \infty \quad \text{as} \quad N \to \infty$$

Indeed, if (1) and (2) are satisfied, then for all $\theta \in [0,1] : \mathbb{P}\left(W_N > \theta\sigma_N\right) \geq (1-\theta)^2 \delta_N$. Now let $\epsilon \in (0,1)$ arbitrary. Then $\exists N_0 \in \mathbb{N}$ such that $\forall N \geq N_0 : \mathbb{P}\left(W_N > \theta\sigma_N\right) \geq (1-\theta)^2 (1-\epsilon)$. This in turn implies that $\mathbb{P}\left(\limsup_{K\to\infty} W_K > \theta\sigma_N\right) \geq (1-\theta)^2 (1-\epsilon)$. This holds because W_N is increasing and thus for all $N \in \mathbb{N} : \limsup_{K\to\infty} W_K \geq W_N$. Note that $\sigma_N \longrightarrow \infty$ when $N \to \infty$. Hence we can conclude that, $\mathbb{P}\left(\limsup_{K\to\infty} W_K = \infty\right) \geq (1-\theta)^2 (1-\epsilon)$ for all θ and ϵ in $(0,1)$. Now if we let $(\theta,\epsilon) \longrightarrow (0,0)$, it follows that $\mathbb{P}\left(\limsup_{K\to\infty} W_K = \infty\right) = 1$. This means that $\mathbb{P}\left(W_N > N\right) \longrightarrow 1$ as $N \to \infty$ and thus we can infer almost sure convergence of W_N to ∞.

It remains to show that (1) and (2) are indeed satisfied. The easiest way too prove this is by straightforward calculation. We will begin with (2) as this is quickly proven. Linearity of the expectation yields,

$$(\sigma_N)_N = \sum_{i=1}^{N} p_i p_{2i} \cdots p_{ki} \longrightarrow \infty \quad \text{as} \quad N \to \infty$$

For (2) we have to do a bit more work. Direct calculation gives,

$$(\delta_N)_N = \frac{\mathbb{E}\left[W_N\right]^2}{\mathbb{E}\left[W_N^2\right]} = \frac{\sum_{j=1}^{N}\sum_{i=1}^{N} p_i \cdots p_{ki} p_j \cdots p_{kj}}{\sum_{j=1}^{N}\sum_{i=1}^{N} \mathbb{E}\left[\omega_i \cdots \omega_{ki}\omega_j \cdots \omega_{kj}\right]}$$

We want to find a lower bound for δ_N, this means we need to find an upper bound for the denominator. We start by noticing that if $\{i,\ldots,ki\} \cap \{j,\ldots,kj\} = \emptyset$, then $\mathbb{E}\left[\omega_i \cdots \omega_{ki}\omega_j \cdots \omega_{kj}\right] = p_i \cdots p_{ki} p_j \cdots p_{kj}$. And if $\{i,\ldots,ki\} \cap \{j,\ldots,kj\} \neq \emptyset$, then $\mathbb{E}\left[\omega_i \cdots \omega_{ki}\omega_j \cdots \omega_{kj}\right] \leq p_i \cdots p_{ki}$. Secondly observe that when $\{i,\ldots,ki\} \cap \{j,\ldots,kj\} \neq \emptyset$ that this occurs when $j = \left(\left(\frac{ai}{b}\right)_{a=1}^{k}\right)_{b=1}^{k}$, with b fixed for each $1 \leq a \leq k$. This gives a number of k^2 overlaps, but we counted too many. As for fixed b and $1 \leq a \leq k$ we get the overlap associated with $j = i$. Also for each overlapping $j = \frac{ai}{b}$, the overlap $i = \frac{bj}{a}$ should also be counted. This doubles the number of overlaps. Hence there are $2(k^2 - k + 1)$ overlaps. This yields,

$$\mathbb{E}\left[W_N^2\right] \leq \sum_{i,j \in D}\sum p_i \cdots p_{ki} p_j \cdots p_{kj} + 2(k^2 - k + 1)\sum_{i=1}^{N} p_i \cdots p_{ki}$$

with $D = \{i, j = 1, \ldots, N : \{i,\ldots,ki\} \cap \{j,\ldots,kj\} = \emptyset\}$. Removing the restrictions on the first summation will only make it larger. Hence

$$\mathbb{E}\left[W_N^2\right] \leq \sum_{j=1}^{N}\sum_{i=1}^{N} p_i \cdots p_{ki} p_j \cdots p_{kj} + C(k)\sum_{i=1}^{N} p_i \cdots p_{ki}$$

By dividing the numerator and denominator, of δ_N, by $\sum_{j=1}^{N}\sum_{i=1}^{N} p_i \cdots p_{ki} p_j \cdots p_{kj}$, we get,

$$(\delta_N)_N \geq \frac{1}{1 + C(k)\frac{\sum\limits_{i=1}^{N} p_i \cdots p_{ki}}{\sum\limits_{j=1}^{N}\sum\limits_{i=1}^{N} p_i \cdots p_{ki} p_j \cdots p_{kj}}} = \frac{1}{1 + C(k)\frac{1}{\sum\limits_{i=1}^{N} p_i \cdots p_{ki}}}$$

Assumption 1.3.1 implies that $(\delta_N)_N$ converges to 1 as $N \to \infty$. Hence (2) is satisfied and thus we have proven that $\sum\limits_{i=n}^{N} \omega_i \omega_{2i} \cdots \omega_{ki} \xrightarrow{a.s.} \infty$ as $N \to \infty$. □

A fun application of theorem 2.2.1 is for the stochastic caricature of the prime numbers. We can obtain a stochastic analogue of Green-Tao theorem. This incredibly deep theorem states that the prime numbers contain infinitely many arbitrary finite length arithmetic progressions. With the following example we illustrate that Green-Tao's theorem also holds for the stochastic prime numbers. This example is inspired by an older thesis of Rik Versendaal [**4**].

Example 2.2.1 (Stochastic analogue of Green-Tao)**.** *Let $\{\omega_i : i \in \mathbb{N}\}$ be a sequence of independent Bernoulli random variables such that $\mathbb{P}(\omega_1 = 1) = 0, \mathbb{P}(\omega_2 = 1) = 1$ and $\mathbb{P}(\omega_i = 1) = \frac{1}{\log(i)}$. We will show that the collection $\mathcal{P} = \{i : \omega_i = 1\}$ of randomly generated primes contain infinitely many k length arithmetic progressions. Theorem 2.2.1 holds true, thus we can simply apply this theorem. To get the desired result we need to check if the assumptions of theorem 2.2.1 hold in this case. Clearly $p_i = \frac{1}{\log(i)} \longrightarrow 0$ as $i \to \infty$. It remains to prove that assumption 1.3.1 checks out. We need to show that*

$$\sum_{i=3}^{\infty} \frac{1}{\log(i)} \frac{1}{\log(2i)} \cdots \frac{1}{\log(ki)} = \infty$$

Begin by noticing that $\log(i) \leq \log(2i) \leq \ldots \leq \log(ki)$*, and thus*

$$\sum_{i=3}^{\infty} \frac{1}{\log(i)} \frac{1}{\log(2i)} \cdots \frac{1}{\log(ki)} \geq \sum_{i=3}^{\infty} \frac{1}{\log(ki)^k}$$

We know that a logarithm of n grows slower that any positive power of n, i.e $\log(n) \leq n^r$ *for all $r > 0$. Hence,* $\log(ki) \leq (ki)^{\frac{1}{k}}$*. Implying that*

$$\sum_{i=3}^{\infty} \frac{1}{\log(ki)^k} \geq \sum_{i=3}^{\infty} \frac{1}{(ki)^{\frac{1}{k}k}} = \sum_{i=3}^{\infty} \frac{1}{ki} = \infty$$

We conclude that $\sum_{i=3}^{\infty} \frac{1}{\log(i)} \frac{1}{\log(2i)} \cdots \frac{1}{\log(ki)} = \infty$*. The requirements of theorem 2.2.1 check out. We get that* $\sum_{i=1}^{N} \omega_i \omega_{2i} \cdots \omega_{ki} \xrightarrow{a.s.} \infty$*. Meaning that with probability 1 the collection $\mathcal{P}$ contains infinitely many arithmetic progressions of any finite length. This is the desired result.*

Chapter 3

Poisson convergence: Chen-Stein method

In the following section we examine the Chen-Stein method. The Chen-Stein method is a mathematical tool used for Poisson approximation. The main focus of the method is to bound the total variation distance between the law of some sum $W = \sum_{i=1}^{N} X_i$, of (usually dependent) Bernoulli random variables, and a Poisson distribution with mean $\lambda = \mathbb{E}[W]$.

This method is very effective for finding bounds on the total variation distance. We try to apply this our AP-k counting summation, as the approach we took for theorem 1.3.3 seems to fail for arithmetic progressions of length 2 or greater. The terms of $X_n^{(k)}$ are dependent, even though it is generated by independent Bernoulli random variables ω_i, with $\omega_i \sim$ Ber(p_i). We wanted $p_i \longrightarrow 0$ as $i \to \infty$, because this way we generate a zero density set $A = \{i : \omega_i = 1\}$. $p_i \longrightarrow 0$ means that the random variables ω_i are becoming increasingly more rare events. From experience we know that when rare events are observed, they distribute themselves along a Poisson distribution. We suggest the same must hold true for $X_n^{(k)}$.

3.1 Chen-Stein method

The Chen-Stein method is a modification of Stein's method for normal approximation. Chen modified his method for Poisson approximation. The method is described as such [**3**]. Let W and Z be a random variables where we try to approximate $\mathcal{L}(W)$ with the target distribution of Z. Let $\mathcal{C}$ and $\mathcal{D}$ be classes of real valued functions defined on some space Ω. In this process of approximation we write $\mathbb{E}h(W) - \mathbb{E}h(Z) = \mathbb{E}Lf_h(W)$, where $h \in \mathcal{C}$ is a test function, L is a linear operator from $\mathcal{C}$ to $\mathcal{D}$ and $f_h \in \mathcal{C}$ is a solution to,

$$Lf = h - \mathbb{E}h(Z). \tag{3.1}$$

L is usually referred to as the Stein operator and equation (3.1) as the Stein equation. The error bound we are trying to find is given by $\mathbb{E}Lf_h$. We can bound $\mathbb{E}Lf_h$ by finding the solution f_h. This solution clearly depends on the operator L. The Stein operator L is such that, $\mathcal{L}(W) = \mathcal{L}(Z)$ iff the Stein identity $\mathbb{E}Lf(W) = 0$, holds for a sufficiently large class of functions f. L completely depends on $\mathcal{L}(Z)$. For normal approximation Stein used,

$$Lf(w) = f^{'}(w) - wf(w) \quad , w \in \mathbb{R}.$$

Chen modified the method to be used for Poisson approximation and defined,

$$Lf(w) = \lambda f(w+1) - wf(w) \quad , w \in \mathbb{Z}_+.$$

Where $\lambda > 0$ is the expectation of the Poisson random variable Z. The operator L, that Chen defined, is made by showing that $\mathbb{E}[\lambda f(Z) - Zf(Z)] = 0$, for every bounded, real valued functions f.

Lemma 3.1.1. *If $Z \sim Po(\lambda)$, then $\mathbb{E}[Lf(Z)] = 0$ for all bounded, real valued functions f.*

Proof. Let $Z \sim Po(\lambda)$ and let f be bounded and real valued, then

$$\begin{aligned}
\mathbb{E}[\lambda f(Z+1)] &= \sum_{w=0}^{\infty} \lambda f(w+1) e^{-\lambda} \frac{\lambda^w}{w!} \\
&= \sum_{w=0}^{\infty} e^{-\lambda} \frac{\lambda^{w+1}}{(w+1)!}(w+1) f(w+1) \\
&= \mathbb{E}[Zf(Z)]
\end{aligned}$$

We conclude that $\mathbb{E}[Lf(Z)] = \mathbb{E}[\lambda f(Z+1) - Zf(Z)] = 0$. □

In the case of our AP-k counting summation finding error bounds for the total variation distance is of interest. To get some acquaintance with the Chen-Stein method we apply it to the AP-1 counting summation $X_n^{(1)}$. Again we give a proof of theorem 1.3.3, but this time we make use of Chen-Stein. In fact we can even show that $d_{TV}\left(\mathcal{L}(X_n^{(1)}), Po(\lambda)\right) \longrightarrow 0$ as $n \to \infty$. Which implies $X_n^{(1)} \xrightarrow{d} Z$, where $Z \sim Po(\lambda)$.

Theorem 3.1.1. *Let $\{\omega_i : i \in \mathbb{N}\}$ be a sequence of independent random variables such that $\omega_i \sim Bernoulli(p_i)$. Assume that $p_i \longrightarrow 0$ as $i \to \infty$ and $\sum_{i=1}^{\infty} p_i = \infty$. Let $\lambda > 0$ and define $r_n(\lambda) = \inf\{k > n : \sum_{i=n}^{k} p_i \geq \lambda\}$. If $X_n^{(1)} := \sum_{i=n}^{r_n} \omega_i$, then $d_{TV}\left(\mathcal{L}(X_n^{(1)}), Po(\lambda)\right) \longrightarrow 0$ as $n \to \infty$.*

Before we attempt to give a proof we look at following lemma.

Lemma 3.1.2. *If f_h is a solution of*

$$Lf(w) = \lambda f(w+1) - wf(w) = h - \mathbb{E}h(Z), \tag{3.2}$$

with $|h| = 1$ and $Z \sim Po(\lambda)$. Then,

$$\|\Delta f_h\|_\infty \leq \frac{1 - e^{-\lambda}}{\lambda} \leq \left(1 \wedge \frac{1}{\lambda}\right),$$

where $\Delta f(w) = f(w+1) - f(w)$.

The proof of this lemma is a bit technical and thus we refer to [**3**] for the full details. We will just use this result to prove theorem 3.1.1.

Proof. [Theorem 3.1.1.] Define $W^{(i)} = X_n^{(1)} - \omega_i$. Then for any bounded real valued f,

$$\begin{aligned}
\mathbb{E}\left[\lambda f(X_n^{(1)}+1) - X_n^{(1)} f(X_n^{(1)})\right] &= \sum_{i=n}^{r_n} \mathbb{E}\left[p_i f(X_n^{(1)}+1) - \omega_i f(X_n^{(1)})\right] \\
&= \sum_{i=n}^{r_n} p_i \mathbb{E}\left[\lambda f(X_n^{(1)}+1) - f(W^{(i)}+1)\right] \\
&= \sum_{i=n}^{r_n} p_i \mathbb{E}\left[\omega_i \Delta f(W^{(i)}+1)\right] \\
&= \sum_{i=n}^{r_n} p_i^2 \mathbb{E}\left[\Delta f(W^{(i)}+1)\right]
\end{aligned}$$

Let $f = f_h$ be a bounded solution of (3.2) with $h = \mathbb{1}_A$, $A \subset \mathbb{Z}_+$. We get that,

$$\begin{aligned}
d_{TV}\left(\mathcal{L}(X_n^{(1)}), Po(\lambda)\right) &= \sup_{A \subset \mathbb{Z}_+} \left|\mathbb{P}(X_n^{(1)} \in A) - \mathbb{P}(Z \in A)\right| \\
&\leq \|\Delta f_h\|_\infty \sum_{i=n}^{r_n} p_i^2 \\
&\leq \left(1 \wedge \frac{1}{\lambda}\right) \sup_{n \leq i \leq r_n} (p_i) \cdot \sum_{i=n}^{r_n} p_i \longrightarrow \left(1 \wedge \frac{1}{\lambda}\right) \cdot 0 \cdot \lambda = 0
\end{aligned}$$

Hence, $X_n^{(1)} \xrightarrow{d} Po(\lambda)$. ☐

3.2 Poisson convergence

If we start looking at the limiting behaviour of $X_n^{(k)} = \sum_{i=n}^{r_n} \omega_i\omega_{2i}\cdots\omega_{ki}$ for $k \geq 2$, it becomes a lot harder to calculate the moment generating function. The sequence of random variables $\Omega_i^{(k)} := (\omega_i\omega_{2i}\cdots\omega_{ki})_{i=n}^{r_n}$ are not independent, thus it is difficult to show what the asymptotic distribution is. Luckily there is an important theorem, due to the Chen-Stein method, which bounds the total variation [**3**].

Theorem 3.2.1 (Arratia, Goldstein & Gordon)**.** *Let $\{X_\alpha : \alpha \in J\}$ be Bernoulli random variables with success probabilities $p_\alpha, \alpha \in J$. Let $W = \sum_{\alpha\in J} X_\alpha$ and $\lambda = \mathbb{E}W = \sum_{\alpha\in J} p_\alpha$. Then, for any collection of sets $B_\alpha \subset J, \alpha \in J$,*

$$d_{TV}\left(\mathcal{L}(W), Po(\lambda)\right) \leq \left(1 \wedge \frac{1}{\lambda}\right)(b_1 + b_2) + \left(1 \wedge \frac{1.4}{\sqrt{\lambda}}\right) b_3$$

where

$$b_1 = \sum_{\alpha\in J}\sum_{\beta\in B_\alpha} p_\alpha p_\beta, \qquad b_2 = \sum_{\alpha\in J}\sum_{\beta\in B_\alpha\backslash\{\alpha\}} \mathbb{E}\left(X_\alpha X_\beta\right),$$

$$b_3 = \sum_{\alpha\in J} \left|\mathbb{E}\left(X_\alpha | X_\beta, \beta \notin B_\alpha\right) - p_\alpha\right|.$$

Note that if we choose X_α independent of $\{X_\beta : \beta \notin B_\alpha\}$ for every $\alpha \in J$, then $b_3 = 0$. Let us now look at the case of arithmetic progressions of length 2, i.e. $\Omega_i^{(2)} = (\omega_i\omega_{2i})_{i=n}^{r_n}$. By first considering this relatively simpler case, as opposed to the general case $\Omega_i^{(k)}$, we get some valuable insights and ideas for the general case. Theorem 3.2.1 gives us enough tools to prove the following theorem.

Theorem 3.2.2. *Let $\{\omega_i : i \in \mathbb{N}\}$ be a sequence of independent random variables such that $\omega_i \sim$ Bernoulli(p_i). Assume that $p_i \longrightarrow 0$ as $i \to \infty$ and $\sum_{i=1}^{\infty} p_i p_{2i} = \infty$. Let $\lambda > 0$ and define $r_n(\lambda) = \inf\{k > n : \sum_{i=n}^{k} p_i p_{2i} \geq \lambda\}$. If $X_n^{(2)} := \sum_{i=n}^{r_n} \omega_i\omega_{2i}$, then $d_{TV}(\mathcal{L}(X_n^{(2)}), Po(\lambda)) \longrightarrow 0$ as $n \to \infty$.*

Proof. In order too apply theorem 3.2.1 we start by defining $\{X_\alpha : \alpha \in J\}$ and the collection of sets $B_\alpha \subset J$. Clearly the X_α are the Bernoulli random variables $\Omega_i^{(2)}$. Thus let $\{X_\alpha : \alpha \in J\} = \{\omega_i\omega_{2i} : i = n, n+1, \ldots, r_n\}$, where J is taken to be $\{\{i, 2i\} : i = n, n+1, \ldots, r_n\}$. For B_α we have to think a bit more. It is best to choose B_α such that $b_3 = 0$. Then we must have for every $\alpha \in J$ and for every $\beta \notin B_\alpha$ that X_α is independent of X_β. Meaning that for every $\alpha \in J : X_\alpha \perp \{X_\beta : \beta \notin B_\alpha\}$.

An arbitrary Bernoulli random variable in the sum $W = \sum_{\alpha\in J} X_\alpha$ is of the form $\omega_i\omega_{2i}$. This term is dependent on the terms $\omega_{\frac{i}{2}}\omega_i$ and $\omega_{2i}\omega_{4i}$. Hence we take

$$B_\alpha = \{\{i/2, i\}, \{i, 2i\}, \{2i, 4i\} : i = n, n+1, \ldots, r_n\}.$$

With this choice for B_α we get that $b_3 = \sum_{\alpha\in J} \left|\mathbb{E}\left(X_\alpha | X_\beta, \beta \notin B_\alpha\right) - p_\alpha\right| = \sum_{\alpha\in J} \left|\mathbb{E}\left(X_\alpha\right) - p_\alpha\right| = 0$.

Now it remains to prove that b_1 and b_2 go to zero for large values of n. We show this by direct calculation.

$$\begin{aligned} b_1 = \sum_{\alpha \in J} \sum_{\beta \in B_\alpha} p_\alpha p_\beta &= \sum_{i=n}^{r_n} p_i p_{2i} p_{\frac{i}{2}} p_i + p_i p_{2i} p_i p_{2i} + p_i p_{2i} p_{2i} p_{4i} \\ &\leq \sum_{i=n}^{r_n} p_i p_{2i} p_{\frac{i}{2}} + p_i p_{2i} p_i + p_i p_{2i} p_{2i} \\ &\leq \lambda \sup_{n \leq i \leq r_n} \left(p_{\frac{i}{2}} + p_i + p_{2i} \right) \longrightarrow 0 \quad \text{as} \quad n \to \infty \end{aligned}$$

For b_2 we get,

$$\begin{aligned} b_2 = \sum_{\alpha \in J} \sum_{\beta \in B_\alpha \setminus \{\alpha\}} \mathbb{E}\left[X_\alpha X_\beta\right] &= \sum_{i=n}^{r_n} \mathbb{E}[\omega_i \omega_{2i} \omega_{\frac{i}{2}} \omega_i] + \mathbb{E}[\omega_i \omega_{2i} \omega_{2i} \omega_{4i}] \\ &= \sum_{i=n}^{r_n} \mathbb{E}[\omega_i \omega_{2i} \omega_{\frac{i}{2}}] + \mathbb{E}[\omega_i \omega_{2i} \omega_{4i}] \\ &= \sum_{i=n}^{r_n} p_i p_{2i} p_{\frac{i}{2}} + p_i p_{2i} p_{4i} \\ &\leq \lambda \sup_{n \leq i \leq r_n} \left(p_{\frac{i}{2}} + p_{4i} \right) \longrightarrow 0 \quad \text{as} \quad n \to \infty \end{aligned}$$

We see that b_2 also goes to 0. Putting these results together gives,

$$\begin{aligned} d_{TV}\left(\mathcal{L}(W), Po(\lambda)\right) &\leq \left(1 \wedge \frac{1}{\lambda}\right)(b_1 + b_2) + \left(1 \wedge \frac{1.4}{\sqrt{\lambda}}\right) b_3 \\ &= \left(1 \wedge \frac{1}{\lambda}\right)(b_1 + b_2) \\ &\leq \left(1 \wedge \frac{1}{\lambda}\right) \lambda \sup_{n \leq i \leq r_n} \left(2p_{\frac{i}{2}} + p_i + p_{2i} + p_{4i} \right) \longrightarrow 0 \quad \text{as} \quad n \to \infty \end{aligned}$$

Hence we conclude $d(\mathcal{L}(X_n^{(2)}), Po(\lambda)) \longrightarrow 0$ as $n \to \infty$. And thus $X_n^{(2)} \xrightarrow{d} Po(\lambda)$. □

Now consider the general case of arithmetic progressions of length k, i.e. $\Omega_i^{(k)} = (\omega_i\omega_{2i}\cdots\omega_{ki})_{i=n}^{r_n}$. Just as in the case of $\Omega_i^{(2)}$, we want to show that the limiting distribution is a Poisson distribution with parameter $\lambda = \lim_{n\to\infty} \sum_{i=n}^{r_n} p_ip_{2i}\cdots p_{ki}$. We can essentially follow the line of the proof of theorem 3.2.2. Again we use theorem 3.2.1 to show that $X_n^{(k)} \xrightarrow{d} Po(\lambda)$ as $n \to \infty$.

Theorem 3.2.3. *Let $\{\omega_i : i \in \mathbb{N}\}$ be a sequence of independent random variables such that $\omega_i \sim$ Bernoulli(p_i). Assume that $p_i \longrightarrow 0$ as $i \to \infty$ and $\sum_{i=1}^{\infty} p_ip_{2i}\cdots p_{ki} = \infty$. Let $\lambda > 0$ and define $r_n(\lambda) = \inf\{k > n : \sum_{i=n}^{k} p_ip_{2i}\cdots p_{ki} \geq \lambda\}$. If $X_n^{(k)} := \sum_{i=n}^{r_n} \omega_i\omega_{2i}\cdots\omega_{ki}$, then $d_{TV}(\mathcal{L}(X_n^{(k)}), Po(\lambda)) \longrightarrow 0$ as $n \to \infty$.*

Proof. Just as before we want too apply theorem 3.2.1. Again we need to define $\{X_\alpha : \alpha \in J\}$ and the collection of sets $B_\alpha \subset J$. Let X_α be the Bernoulli random variables $\Omega_i^{(k)}$. Thus $\{X_\alpha : \alpha \in J\} = \{\omega_i\omega_{2i}\cdots\omega_{ki} : i = n, n+1, \dots, r_n\}$, where J is taken to be $\{\{i, 2i, \dots, ki\} : i = n, \dots, r_n\}$.

In order to find the right definition for B_α we have to consider all terms that are dependent on an arbitrary term inside the sum $W = \sum_{\alpha\in J} X_\alpha = \sum_{i=n}^{r_n} \omega_i\omega_{2i}\cdots\omega_{ki}$. A random term in W is of the form $\omega_i\omega_{2i}\cdots\omega_{ki}$. This term is dependent on $\omega_j\omega_{2j}\cdots\omega_{kj}$ for $j = \left(\left(\frac{ai}{b}\right)_{a=1}^{k}\right)_{b=1}^{k}$, with b fixed for each $1 \leq a \leq k$. Thus define,

$$B_\alpha = \left\{ \{j, 2j, \cdots, kj\} |\; j = \left(\left(\frac{ai}{b}\right)_{a=1}^{k}\right)_{b=1}^{k} : i = n, \dots, r_n \right\}$$

With this choice of B_α it follows that $b_3 = \sum_{\alpha\in J} |\mathbb{E}\left(X_\alpha | X_\beta, \beta \notin B_\alpha\right) - p_\alpha| = \sum_{\alpha\in J} |\mathbb{E}\left(X_\alpha\right) - p_\alpha| = 0$.

Now it remains to prove that b_1 and b_2 go to zero for large values of n. Direct calculation yields,

$$\begin{aligned} b_1 = \sum_{\alpha\in J}\sum_{\beta\in B_\alpha} p_\alpha p_\beta &= \sum_{i=n}^{r_n} p_ip_{2i}\cdots p_{ki}\left(\sum_{b=1}^{k}\sum_{a=1}^{k} p_jp_{2j}\cdots p_{kj}\big|_{j=\frac{ai}{b}}\right) \\ &\leq \sum_{i=n}^{r_n} p_ip_{2i}\cdots p_{ki}\left(\sum_{b=1}^{k}\sum_{a=1}^{k} p_{\frac{ai}{b}}\right) \\ &\leq \lambda \sup_{n\leq i\leq r_n}\left(\sum_{b=1}^{k}\sum_{a=1}^{k} p_{\frac{ai}{b}}\right) \longrightarrow 0 \quad \text{as} \quad n \to \infty \end{aligned}$$

Thus b_1 converges to 0 as $n \to \infty$. For b_2 we get,

$$b_2 = \sum_{\alpha\in J}\sum_{\beta\in B_\alpha\setminus\{\alpha\}} \mathbb{E}\left[X_\alpha X_\beta\right] = \sum_{i=n}^{r_n}\left(\sum_{\substack{b=1\\|\{a,b\}|=2}}^{k}\sum_{a=1}^{k} \mathbb{E}[\omega_i\cdots\omega_{ki}\omega_j\cdots\omega_{kj}]\big|_{j=\frac{ai}{b}}\right)$$

Note that, if $|\{a, b\}| = 2$, for each fixed b and $1 \leq a \leq k$ there is at least one $j \in \left\{\frac{ai}{b}, \frac{2ai}{b}, \dots, \frac{kai}{b}\right\}$ such that $\omega_i\omega_{2i}\cdots\omega_{ki} \neq \omega_i\omega_{2i}\cdots\omega_{ki}\omega_j$. The first such j were this holds we denote by Δ^i. To emphasize that this j depends on a and b, we write $\Delta^i(a, b)$. Hence we can now bound b_2,

$$b_2 \leq \sum_{i=n}^{r_n} \left(\sum_{\substack{b=1 \\ |\{a,b\}|=2}}^{k} \sum_{a=1}^{k} \mathbb{E}[\omega_i \omega_{2i} \cdots \omega_{ki} \omega_j]\big|_{j=\Delta^i(a,b)} \right) \leq \sum_{i=n}^{r_n} p_i p_{2i} \cdots p_{ki} \left(\sum_{\substack{b=1 \\ |\{a,b\}|=2}}^{k} \sum_{a=1}^{k} p_{\Delta^i(a,b)} \right)$$

$$\leq \lambda \sup_{n \leq i \leq r_n} \left(\sum_{\substack{b=1 \\ |\{a,b\}|=2}}^{k} \sum_{a=1}^{k} p_{\Delta^i(a,b)} \right) \longrightarrow 0 \quad \text{as} \quad n \to \infty$$

Now that we have derived that both b_1 and b_2 goes to 0 as $n \to \infty$, we can see that,

$$\begin{aligned} d_{TV}\left(\mathcal{L}(W), Po(\lambda)\right) &\leq \left(1 \wedge \frac{1}{\lambda}\right)(b_1 + b_2) + \left(1 \wedge \frac{1.4}{\sqrt{\lambda}}\right) b_3 \\ &= \left(1 \wedge \frac{1}{\lambda}\right)(b_1 + b_2) \\ &\leq \left(1 \wedge \frac{1}{\lambda}\right) \lambda \sup_{n \leq i \leq r_n} \left(\sum_{b=1}^{k} \sum_{a=1}^{k} p_{\frac{ai}{b}} + \sum_{\substack{b=1 \\ |\{a,b\}|=2}}^{k} \sum_{a=1}^{k} p_{\Delta^i(a,b)} \right) \longrightarrow 0 \quad \text{as} \quad n \to \infty \end{aligned}$$

Hence we conclude $d_{TV}(\mathcal{L}(X_n^{(k)}), Po(\lambda)) \longrightarrow 0$ as $n \to \infty$. And thus $X_n^{(k)} \xrightarrow{d} Po(\lambda)$. □

3.3 Large arithmetic progressions

Theorem 3.2.1 is a very general theorem and applicable to a broad number of problems. In this section we want to show that this theorem can be used to prove that the occurrence of very long arithmetic progressions is approximately Poisson distributed. A similar result is obtained by Kifer in [**5**].

The setup is as such. Consider a sequence of independent and identically distributed Bernoulli random variables $\{\omega_i : i \in \mathbb{N}\}$. We let $\mathbb{P}(\omega_i = 1) = p \in (0, 1]$. From Szemerédi's theorem we already know that the subset $A = \{i : \omega_i = 1\} \subset \mathbb{N}$ contains infinitely many arithmetic progressions of any finite length.

We would like to know the distribution of the number of 'large' arithmetic progressions in A. By 'large' we refer to arithmetic progressions $\omega_i \omega_{2i} \cdots \omega_{\varphi(N)i}$, where $\varphi(N)$ is an increasing function that diverges as $N \to \infty$. The occurrence of large progressions are rare events. This leads us to think that the law of rare events might be at play here.

Counting the number of large arithmetic progressions is done with $X_N = \sum_{i=1}^{N} \omega_i \omega_{2i} \cdots \omega_{\varphi(N)i}$. We want to show Poisson convergence for X_N, i.e show that $X_N \xrightarrow{d} Po(\lambda)$. First we need to define what the parameter λ should be. The law of rare events tells us that this is approximately the expected number of successful events. Hence we define,

$$\mathbb{E}[X_N] = \mathbb{E}\left[\sum_{i=1}^{N} \omega_i \cdots \omega_{\varphi(N)i}\right] = \sum_{i=1}^{N} \mathbb{E}\left[\omega_i \cdots \omega_{\varphi(N)i}\right] = Np^{\varphi(N)} \longrightarrow \lambda \text{ as } N \to \infty.$$

From this we can derive that $\varphi(N) \uparrow \log_p\left(\frac{\lambda}{N}\right)$ as $N \to \infty$. Note that this means that $\varphi(N)$ diverges as $N \to \infty$, because $p \in (0, 1)$.

We will now prove that $X_N \xrightarrow{d} Po(\lambda)$.

Theorem 3.3.1. *Let $\{\omega_i : i \in \mathbb{N}\}$ be a sequence of independent random variables such that $\omega_i \sim$ Bernoulli(p). Define $\lambda > 0$ and let $\sum_{i=1}^{N} \mathbb{E}\left[\omega_i \cdots \omega_{\varphi(N)i}\right] \longrightarrow \lambda$ as $N \to \infty$. If $X_N := \sum_{i=1}^{N} \omega_i \cdots \omega_{\varphi(N)i}$, then $d_{TV}(\mathcal{L}(X_N), Po(\lambda)) \longrightarrow 0$ as $N \to \infty$*

Proof. We take the same approach as in the previous section, and try too apply the powerful theorem of Arratia, Goldstein & Gordon.

Just as before we need to know how to correctly define $\{X_\alpha : \alpha \in J\}$ and the collection of sets $B_\alpha \in J$. Logically is to define $\{X_\alpha : \alpha \in J\} := \{\omega_i \omega_{2i} \cdots \omega_{\varphi(N)i} : i = 1, 2, \ldots, N\}$, where we let $J = \{\{i, 2i, \ldots, \varphi(N)i\} : i = 1, 2, \ldots, N\}$, as such. For B_α we use the same logic as in the previous sections. Thus let ,

$$B_\alpha := \left\{ \{j, 2j, \ldots, \varphi(N)j\} |\ j = \left(\left(\frac{ai}{b} \right)_{a=1}^{\varphi(N)} \right)_{b=1}^{\varphi(N)} : i = 1, \ldots, N \right\}$$

Note that by defining B_α like this X_α is completely independent of $\{X_\beta : \beta \notin B_\alpha\}$ for every $\alpha \in J$. Hence, $b_3 = 0$.

We are left to prove that b_1 and b_2 become arbitrarily small with increasing N. For b_1 we get,

$$b_1 = \sum_{\alpha \in J} \sum_{\beta \in B_\alpha} p_\alpha p_\beta \leq \sum_{i=1}^{N} p_i \cdots p_{\varphi(N)i} \left(\sum_{b=1}^{\varphi(N)} \sum_{a=1}^{\varphi(N)} p_j \cdots p_{\varphi(N)j} \big|_{j=\frac{ai}{b}} \right)$$

Each of the ω_i are identically distributed implying that $p_i = p_{2i} = \cdots = p_{\varphi(N)i} = p \in (0, 1)$. This gives,

$$b_1 = Np^{\varphi(N)} \left(\sum_{b=1}^{\varphi(N)} \sum_{a=1}^{\varphi(N)} p^{\varphi(N)} \right) = Np^{\varphi(N)} \varphi^2(N) p^{\varphi(N)}$$

We know that $Np^{\varphi(N)}$ converges to λ. In fact, $Np^{\varphi(N)}$ increases ,with increasing N, to λ. Thus, $Np^{\varphi(N)} \leq \lambda$, $p^{\varphi(N)} \leq \frac{\lambda}{N}$ and $\varphi^2(N) \leq \log_p^2\left(\frac{\lambda}{N}\right)$ for $N \geq 1$. Hence,

$$b_1 \leq \lambda \log_p^2 \left(\frac{\lambda}{N} \right) \frac{\lambda}{N}$$

Both $\log_p^2\left(\frac{\lambda}{N}\right)$, note $p \in (0, 1)$, and N converges (or diverges) to ∞. But N is of higher order, thus

$$\lambda^2 \lim_{N \to \infty} \frac{\log_p^2\left(\frac{\lambda}{N}\right)}{N} = 0$$

Hence b_1 converges to 0 as $N \to \infty$.

For b_2 we get,

$$b_2 = \sum_{\alpha \in J} \sum_{\beta \in B_\alpha \setminus \{\alpha\}} \mathbb{E}\left[X_\alpha X_\beta\right] = \sum_{i=1}^{N} \left(\sum_{\substack{b=1 \\ |\{a,b\}|=2}}^{\varphi(N)} \sum_{a=1}^{\varphi(N)} \mathbb{E}[\omega_i \cdots \omega_{ki} \omega_j \cdots \omega_{\varphi(N)j}] \big|_{j=\frac{ai}{b}} \right)$$

Observe that this expression is symmetric in i and j. Meaning that we can assume without loss of generality, that $i < j$. Implying that $a > b$.

Note that for $j = 2i$ we have a minimum amount of non overlapping ω_j. When $j = 2i$ the total number of non overlapping ω_j is $\frac{\varphi(N)}{2} - 1$. Hence,

$$b_2 = 2\sum_{i=1}^{N}\left(\sum_{\substack{b=1 \\ a>b}}^{\varphi(N)}\sum_{a=1}^{\varphi(N)}\mathbb{E}[\omega_i\cdots\omega_{ki}\omega_j\cdots\omega_{\varphi(N)j}]\big|_{j=\frac{ai}{b}}\right) \leq 2N\varphi^2(N)p^{\varphi(N)}p^{\frac{\varphi(N)}{2}-1}$$

Remember that $Np^{\varphi(N)} \leq \lambda$ and $\varphi^2(N) \leq \log_p^2\left(\frac{\lambda}{N}\right)$ for $N \geq 1$. We derive,

$$b_2 \leq \frac{2\lambda}{p}\log_p^2\left(\frac{\lambda}{N}\right)\sqrt{\frac{\lambda}{N}} = \frac{2\lambda^{\frac{3}{2}}}{p}\frac{\log_p^2\left(\frac{\lambda}{N}\right)}{\sqrt{N}}$$

Both $\log_p^2\left(\frac{\lambda}{N}\right)$, note $p \in (0,1)$, and $\sqrt{N}$ converges (or diverges) to ∞. But $\sqrt{N}$ is of higher order, thus

$$\frac{2\lambda^{\frac{3}{2}}}{p}\lim_{N\to\infty}\frac{\log_p^2\left(\frac{\lambda}{N}\right)}{\sqrt{N}} = 0$$

Now that we have derived that both b_1 and b_2 goes to 0 as $N \to \infty$, we can see that,

$$\begin{aligned} d_{TV}\left(\mathcal{L}(W), Po(\lambda)\right) &\leq \left(1 \wedge \frac{1}{\lambda}\right)(b_1 + b_2) + \left(1 \wedge \frac{1.4}{\sqrt{\lambda}}\right)b_3 \\ &= \left(1 \wedge \frac{1}{\lambda}\right)(b_1 + b_2) \\ &\leq \left(1 \wedge \frac{1}{\lambda}\right)\left(\lambda^2\frac{\log_p^2\left(\frac{\lambda}{N}\right)}{N} + \frac{2\lambda^{\frac{3}{2}}}{p}\frac{\log_p^2\left(\frac{\lambda}{N}\right)}{\sqrt{N}}\right) \longrightarrow 0 \quad \text{as } N \to \infty \end{aligned}$$

Hence we conclude $d_{TV}(\mathcal{L}(X_N), Po(\lambda)) \longrightarrow 0$ as $N \to \infty$. And thus $X_N \xrightarrow{d} Po(\lambda)$. □

Chapter 4

Poisson convergence: Transfer matrix method

This chapter we will introduce a more direct approach for showing Poisson convergence, by explicitly calculating the moment generating function of $X_n^{(k)}$, for $k = 1, 2$.

For arithmetic progressions of length 2 this will be carried out with a transfer matrix technique. This method is inspired from statistical mechanics. In statistical mechanics the transfer matrix method is used to derive a more simple form of the partition function. The partition function is defined as $Z(s) = \sum_{s_i} \exp(-\beta H(s_1, s_2, \ldots))$, where s_i are certain states of a (physical) system that are accessible and H is the Hamiltonian [**10**].

The precise meaning of this partition function is not of importance to us. The method of rewriting the partition function into a simple vector matrix vector multiplication, is what is of importance for us. This is done by writing, $Z(s) = \underline{v} \cdot \left(\prod_{i=1}^{N} M_i\right) \cdot \underline{w}$. Here $\underline{v}, \underline{w}$ are vectors of dimension d and M_i are transfer matrices of dimension d x d. This rewriting will turn out to be a very useful tool for computing the moment generating function of $X_n^{(2)}$.

The advantage of the transfer matrix method over the Chen-Stein method, is that this is a direct approach. It gives us an exact expression for the moment generating function, which is convenient for computation and controlled approximations.

4.1 Direct approach for length 1 progressions

The goal is to show that $X_n^{(k)} \xrightarrow{d} Po(\lambda)$, by direct calculation of the moment generating function. We start by looking at arithmetic progressions of length 1. By first looking at length 1 arithmetic progressions we can draw some inspiration and ideas to generalize for length k length arithmetic progressions.

Theorem 4.1.1. *Let $(\omega_i)_{i \in \mathbb{N}}$ be a sequence of independent random variables such that $\omega_i \sim$ Bernoulli(p_i). Assume that $p_i \longrightarrow 0$ as $i \to \infty$ and $\forall n \in \mathbb{N} : \sum_{i=n}^{\infty} p_i = \infty$. Let $\lambda > 0$ and define $r_n(\lambda) = \inf\{k > n : \sum_{i=n}^{k} p_i \geq \lambda\}$. If $X_n^{(1)} := \sum_{i=n}^{r_n} \omega_i$, then $X_n^{(1)} \xrightarrow{d} Po(\lambda)$ as $n \to \infty$.*

Proof. This time we will take a different approach too proving this. Direct calculation of the moment generating function of $X_n^{(1)}$ gives,

$$M_{X_n^{(1)}}(t) = \mathbb{E}\left[e^{t \sum_{i=n}^{r_n} \omega_i}\right] = \prod_{i=n}^{r_n} \mathbb{E}\left[e^{t\omega_i}\right] = \prod_{i=n}^{r_n} \left(1 + p_i(e^t - 1)\right)$$

We want to take a general approach to evaluate this product, because we want to generalize this for k-length arithmetical progressions. Carefully arranging the different terms yields,

$$\prod_{i=n}^{r_n}\big(1+p_i(e^t-1)\big) = 1+\sum_{i=n}^{r_n} p_i(e^t-1)+\frac{1}{2!}\sum_{\substack{i=n\\|\{i,j\}|=2}}^{r_n}\sum_{j=n}^{r_n} p_ip_j(e^t-1)^2+\frac{1}{3!}\sum_{\substack{i=n\\|\{i,j,k\}|=3}}^{r_n}\sum_{j=n}^{r_n}\sum_{k=n}^{r_n} p_ip_jp_k(e^t-1)^3$$
$$+\cdots+\frac{1}{M!}\sum_{\substack{i_1=n\\|\{i_1,i_2,\cdots,i_M\}|=M}}^{r_n}\sum_{i_2=n}^{r_n}\cdots\sum_{i_M=n}^{r_n} p_{i_1}p_{i_2}\cdots p_{i_M}(e^t-1)^M+\cdots+p_np_{n+1}\cdots p_{r_n}(e^t-1)^{r_n-n+1}$$

If the restrictions on the summations were to be removed, this would result the power series of $e^{\sum_{i=n}^{r_n} p_i(e^t-1)}$. Note that an arbitrary term,

$$\frac{1}{M!}\sum_{\substack{i_1=n\\|\{i_1,i_2,\cdots,i_M\}|=M}}^{r_n}\sum_{i_2=n}^{r_n}\cdots\sum_{i_M=n}^{r_n} p_{i_1}p_{i_2}\cdots p_{i_M}(e^t-1)^M = \frac{1}{M!}\left(\sum_{i=n}^{r_n} p_i(e^t-1)\right)^M - \frac{(e^t-1)^M}{M!}\sum_{i=2}^{M}\binom{M}{i} o_M^{(M-i)}(n)$$

Where $o_M^{(M-i)}(n)$ is a sum consisting of products of an i number of p_i. At first glance this might not be obvious. In order too get a better understanding of these $o_M^{(M-i)}(n)$ terms we calculate a few terms of this sum, up to the $n+2$ factor. Note that the upcoming part of this proof is not essential and can be skipped. The purpose of this is just to demonstrate how to find the $o_M^{(M-i)}(n)$ terms.

$$\prod_{i=n}^{n+2}\big(1+p_i(e^t-1)\big) = 1+\sum_{i=n}^{n+2} p_i(e^t-1)+\frac{1}{2!}\sum_{\substack{i=n\\|\{i,j\}|=2}}^{n+2}\sum_{j=n}^{n+2} p_ip_j(e^t-1)^2+\frac{1}{3!}\sum_{\substack{i=n\\|\{i,j,k\}|=3}}^{n+2}\sum_{j=n}^{n+2}\sum_{k=n}^{n+2} p_ip_jp_k(e^t-1)^3 \tag{*}$$

Now we set this equal to the same terms without restrictions. This will obviously give to many terms. Thus we also subtract some correction terms and denote these as the $o_M^{(M-i)}(n)$. We get,

$$\prod_{i=n}^{n+2}\big(1+p_i(e^t-1)\big) \stackrel{set}{=\!=} 1+\sum_{i=n}^{n+2} p_i(e^t-1)+\frac{1}{2!}\left(\sum_{i=n}^{n+2} p_i(e^t-1)\right)^2-\frac{(e^t-1)^2}{2!}o_2(n)$$
$$+\frac{1}{3!}\left(\sum_{i=n}^{n+2} p_i(e^t-1)\right)^3-\frac{(e^t-1)^3}{3!}\left(\binom{3}{2}o_3'(n)+\binom{3}{3}o_3(n)\right)$$

Now we want to write this in terms of (*). Expanding every summation yields,

$$= 1 + \sum_{i=n}^{n+2} p_i(e^t - 1) + \frac{1}{2!} \sum_{\substack{i=n \\ |\{i,j\}|=2}}^{n+2} \sum_{j=n}^{n+2} p_i p_j (e^t - 1)^2 + \frac{(e^t-1)^2}{2!} \left(p_n^2 + p_{n+1}^2 + p_{n+2}^2\right) - \frac{(e^t-1)^2}{2!} o_2(n)$$

$$+ \frac{1}{3!} \sum_{\substack{i=n \\ |\{i,j,k\}|=3}}^{n+2} \sum_{j=n}^{n+2} \sum_{k=n}^{n+2} p_i p_j p_k (e^t - 1)^3 + \frac{(e^t-1)^3}{3!} \left(\binom{3}{2} (p_n^2 p_{n+1} + p_n^2 p_{n+2} + p_{n+1}^2 p_n + p_{n+1}^2 p_{n+2}\right.$$

$$\left. + p_{n+2}^2 p_n + p_{n+2}^2 p_{n+1}) + \binom{3}{3} (p_n^3 + p_{n+1}^3 + p_{n+2}^3) \right) - \frac{(e^t-1)^3}{3!} \left(\binom{3}{2} o_3'(n) + \binom{3}{3} o_3(n) \right)$$

By comparing this expression to (*) we can now find explicitly what the correction terms are. Thus,

$$\begin{aligned} o_2(n) &= p_n^2 + p_{n+1}^2 + p_{n+2}^2 \\ o_3'(n) &= p_n^2 p_{n+1} + p_n^2 p_{n+2} + p_{n+1}^2 p_n + p_{n+1}^2 p_{n+2} + p_{n+2}^2 p_n + p_{n+2}^2 p_{n+1} \\ o_3(n) &= p_n^3 + p_{n+1}^3 + p_{n+2}^3 \end{aligned}$$

From this we can conclude that every $o_M^{(M-i)}(n)$ is bounded by $o_2(n)$, because every $p_i \in [0,1]$. Now that this part is concluded, we continue with the proof. From now on we will use some more dense notation. Note that,

$$I := \prod_{i=n}^{r_n} \left(1 + p_i(e^t - 1)\right) = \sum_{M=0}^{r_n - n + 1} \frac{1}{M!} A_M(n)$$

$$\text{with,} \quad A_M(n) = \sum_{\substack{i_1=n \\ |\{i_1,i_2,\cdots,i_M\}|=M}}^{r_n} \sum_{i_2=n}^{r_n} \cdots \sum_{i_M=n}^{r_n} p_{i_1} p_{i_2} \cdots p_{i_M} (e^t - 1)^M$$

It also holds that,

$$\prod_{i=n}^{r_n} \left(1 + p_i(e^t - 1)\right) = \sum_{M=0}^{r_n - n + 1} \frac{1}{M!} \left(A_M'(n) - B_M(n)\right)$$

$$\text{with,} \quad A_M'(n) = \left(\sum_{i=n}^{r_n} p_i(e^t - 1) \right)^M$$

$$B_M(n) = A_M'(n) - A_M(n) \leq o_2(n) \sum_{i=0}^{M} \binom{M}{i} (e^t - 1)^M = o_2(n) 2^M (e^t - 1)^M$$

Now define $II := \sum_{M=0}^{r_n - n + 1} A_M'(n)$. Note that,

$$|I - II| = \sum_{M=0}^{r_n - n + 1} \frac{1}{M!} B_M(n) \leq \sum_{M=0}^{r_n - n + 1} \frac{1}{M!} 2^M (e^t - 1)^M o_2(n) = \sum_{M=0}^{r_n - n + 1} \frac{(2(e^t-1))^M}{M!} \sum_{i=n}^{r_n} p_i^2$$

$$\leq \sum_{M=0}^{\infty} \frac{(2(e^t-1))^M}{M!} \sup_{n \leq i \leq r_n} (p_i) \sum_{i=n}^{r_n} p_i \xrightarrow{n \to \infty} e^{2(e^t - 1)} \cdot 0 \cdot \lambda = 0$$

This means that I and II have the same asymptotic behaviour. Hence we can deduce the limiting distribution of $X_n^{(1)}$ from II. Thus,

$$\lim_{n\to\infty} II = \lim_{n\to\infty} \sum_{M=0}^{r_n-n+1} \frac{1}{M!}\left(\sum_{i=n}^{r_n} p_i(e^t-1)\right)^M = \exp \lim_{n\to\infty} \sum_{i=n}^{r_n} p_i(e^t-1) = \mathrm{e}^{\lambda(e^t-1)}$$

We can conclude that $X_n^{(1)} \xrightarrow{d} Po(\lambda)$ as $n \to \infty$. □

4.2 Transfer matrix application for length 2 arithmetic progressions

In this section we introduce the transfer matrix method to be used for the length 2 arithmetic progressions counting summation $X_n^{(2)}$. We emphasize that we only present this method as a way of computation, and do not provide a rigorous proof. Further research is needed for this.

The most clear manner of explaining the transfer matrix method is by direct calculation, as this approach is explicit in nature. Before we start calculating the moment generating function of $X_n^{(2)}$, we look more closely at $X_n^{(2)}$. $X_n^{(2)} = \sum_{i=n}^{r_n} \omega_i\omega_{2i}$ contains terms that have long range dependencies. For example take terms like $\omega_n\omega_{2n}$ and $\omega_{2n}\omega_{4n}$. The distance between these progressions will only increase as $n \to \infty$. Hence we say that these terms are long range dependent.

A consequence of these dependencies is that we can not write,

$$\mathbb{E}\left[e^{t\sum_{i=n}^{r_n}\omega_i\omega_{2i}}\right] = \prod_{i=n}^{r_n} \mathbb{E}\left[e^{t\omega_i\omega_{2i}}\right]$$

What we can do is the following. Observe that $X_n^{(2)}$ admits a decomposition into mutually independent layers as such

$$\begin{aligned}
&\omega_1\omega_2 + \omega_2\omega_4 + \omega_4\omega_8 + \cdots \\
&\omega_3\omega_6 + \omega_6\omega_{12} + \omega_{12}\omega_{24} + \cdots \\
&\vdots \\
&\omega_{2k+1}\omega_{2(2k+1)} + \omega_{2(2k+1)}\omega_{2^2(2k+1)} + \omega_{2^2(2k+1)}\omega_{2^3(2k+1)} + \cdots \\
&\vdots
\end{aligned}$$

Now define the collection $\mathcal{O} = \{m \in \{n, \dots, r_n\} : m \bmod 2 = 1\}$. Note that by this decomposition we can write,

$$X_n^{(2)} = \sum_{i=n}^{r_n} \omega_i\omega_{2i} = \sum_{k\in\mathcal{O}} \sum_{j=0}^{N} \sigma(k,j)\sigma(k,j+1).$$

Where $\sigma(k,i) = \omega_{k2^j}$, with $0 \le j \le \left\lfloor \log_2\left(\frac{r_n}{k}\right)\right\rfloor$. For convenience we denote $N = N(k) = \left\lfloor \log_2\left(\frac{r_n}{k}\right)\right\rfloor$.

Utilizing this alternate form of $X_n^{(2)}$ we derive,

$$\mathbb{E}\left[e^{t\sum_{i=n}^{r_n}\omega_i\omega_{2i}}\right] = \mathbb{E}\left[e^{t\sum_{k\in\mathcal{O}}\sum_{j=0}^{N}\sigma(k,j)\sigma(k,j+1)}\right] = \prod_{k\in\mathcal{O}}\mathbb{E}\left[e^{t\sum_{j=0}^{N}\sigma(k,j)\sigma(k,j+1)}\right].$$

By mutual independence of the layers the first summation over k is factored outside the expectation. What remains to be calculated is the expectation $\mathbb{E}\left[e^{t\sum_{j=0}^{N}\sigma(k,j)\sigma(k,j+1)}\right]$. Direct calculation yields,

$$\mathbb{E}\left[e^{t\sum_{j=0}^{N}\sigma(k,j)\sigma(k,j+1)}\right] = \sum_{\sigma_0\cdots\sigma_{N+1}}\prod_{j=0}^{N}e^{t\sigma(k,j)\sigma(k,j+1)}\prod_{j=0}^{N+1}p(k,j)^{\sigma(k,j)}(1-p(k,j))^{1-\sigma(k,j)}.$$

Here $p(k,j) = \mathbb{E}[\sigma(k,j)]$.

The next and most important step is the introduction of transfer matrices. We write this complicated expression into a more simple form. Note that,

$$\mathbb{E}\left[e^{t\sum_{j=0}^{N}\sigma(k,j)\sigma(k,j+1)}\right] = \begin{pmatrix}1 & 1\end{pmatrix}\left(\prod_{j=0}^{N}M^{(k,j)}_{\sigma_{(k,j)}\sigma_{(k,j+1)}}\right)\begin{pmatrix}1-p_{N+1}\\ p_{N+1}\end{pmatrix}$$

Where transfer matrix $M^{(k,j)}_{\sigma_{(k,j)}\sigma_{(k,j+1)}} = p(k,j)^{\sigma(k,j)}\left(1-p(k,j)\right)^{1-\sigma(k,j)}e^{t\sigma(k,j)\sigma(k,j+1)}$. The matrix looks like,

$$M^{(k,j)}_{\sigma_{(k,j)}\sigma_{(k,j+1)}} = \begin{pmatrix}1-p(k,j) & 1-p(k,j)\\ p(k,j) & p(k,j)e^t\end{pmatrix}$$

This method of rewriting is called the transfer matrix method.

We see that we are left with the immense task of calculating this matrix product. Unfortunately this is about the extend of the research. After this point no concrete results were obtained. Although we are quite convinced that this approach should give the desired Poisson convergence. We conjecture that $\sum_{j=0}^{N}\sigma(k,j)\sigma(k,j+1)\xrightarrow{d}Po(\mu)$, with $\mu\longleftarrow\mathbb{E}\left[\sum_{j=0}^{N}\sigma(k,j)\sigma(k,j+1)\right]$. Implying that $X_n^{(2)} = \sum_{k\in\mathcal{O}}\sum_{j=0}^{N}\sigma(k,j)\sigma(k,j+1)\xrightarrow{d}Po(\lambda)$. We think this implication should hold true, because the distribution of a sum of independent Poisson random variables is also Poisson.

We would like to leave the reader with some key observations for further research. An important observation is that the matrix $M^{(k,j)}_{\sigma_{(k,j)}\sigma_{(k,j+1)}}$ can be written as the sum of 2 matrices A and B.

$$M^{(k,j)}_{\sigma_{(k,j)}\sigma_{(k,j+1)}} = A + p(k,j)B = \begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix} + p(k,j)\begin{pmatrix}-1 & -1\\ 1 & e^t\end{pmatrix}$$

Note that $A^k = A$ for every $k \geq 1$. The matrices A and B unfortunately do not commute. Meaning that we need to take the order of operation into consideration. Carefully arranging of terms gives,

$$\prod_{j=0}^{N} M^{(k,j)}_{\sigma_{(k,j)}\sigma_{(k,j+1)}} = A^N + \sum_{j=0}^{N-1} p(k,j)ABA + p(k,N)AB + \sum_{j=1}^{N-2} p(k,j)p(k,j+1)AB^2A$$
$$+ p(k,0)p(k,1)B^2A^{N-1} + p(k,N-1)p(k,N)AB^2 + o(p(k,j)^3)$$

Where $o(p(k,j)^3)$ are the higher order terms.

Observe that the first order terms $\sum_{j=0}^{N-1} p(k,j)ABA$ equate to 0, because $ABA = 0$. Also note that $AB^2A = \begin{pmatrix} e^t - 1 & e^t - 1 \\ 0 & 0 \end{pmatrix}$, which will give the important factor of,

$$\begin{pmatrix} 1 & 1 \end{pmatrix} AB^2A \begin{pmatrix} 1 - p_{N+1} \\ p_{N+1} \end{pmatrix} = (e^t - 1)$$

What remains to be proven is that the remaining terms become arbitrary small as $n \to \infty$. Important insight could be taken from the direct approach for length 1 progressions. Unfortunately a rigorous proof of this is not obtained. More research on this topic is needed.

Bibliography

[1] A. Arbieto, C. Matheus, C.G. Moreira. (2009). *The remarkable effectiveness of ergodic theory in number theory.* Ensaios Mathemáticos.

[2] H. Furstenberg. (1981). *Recurrence in Ergodic Theory and Combinatorial Number Theory.* Princeton, New Jersey: Princeton University Press.

[3] L.H.Y Chen, A. Röllin. (2013). *Approximating dependent rare events.* National university of Singapore.

[4] R. Versendaal. (2014). *Arithmetische progressies in random kleuringen van de gehele getallen.* TU Delft, Nederland.

[5] Y. Kifer. (2011). *Nonconventional Poisson limit theorems.* Institute of mathematics, Jerusalem, Israel.

[6] J. Moreira. (2011). *A topological proof of Van der Waerden's Theorem.* Retrieved from: `https://joelmoreira.wordpress.com/2011/01/30/a-topological-proof-of-van-der-waerdens-theorem/`

[7] P. Dey. (2013). *Stein-Chen method for Poisson approximation.* Retrieved from: `https://faculty.math.illinois.edu/ psdey/414CourseNotes.pdf`

[8] D. Rosenberg. (2007). *Stein's method and applications.* Retrieved from: `https://statweb.stanford.edu/ souravc/Lecture2.pdf`

[9] N. Ross. (2011). *Fundamentals of Stein's method.* University of California, Berkeley.

[10] D.V. Schroeder. (2000). *Thermal Physics.* United States: Addison Wesley Longman.

[11] W.M. Ruszel. (2018). *Lecture notes on Advanced probability*