NTNU
Norwegian University of
Science and Technology

Kingman's subadditive ergodic theorem and its application

Erling Brækhus Haugsand

Master of Science in Physics and Mathematics
Submission date: June 2017
Supervisor: Eugenia Malinnikova, IMF

Norwegian University of Science and Technology
Department of Mathematical Sciences

Preface

This thesis is the final submission of a master's degree program in applied physics and mathematics at the Norwegian University of Science and Technology. The work was carried out at the department of mathematical sciences, during the spring of 2017 under the supervision of Professor Eugenia Malinnikova.

First and foremost, I would like to thank Professor Eugenia Malinnikova, for her excellent guidance, support and patience throughout the time I have been under her supervision. Without her invaluable input, the work on this thesis would surely have been less rewarding. In addition, I would like to thank my lecturer Silvius Klein for introducing me to measure and ergodic theory.

Abstract

In this master thesis we study Kingman's subadditive ergodic theorem and its application. We prove Kingman's theorem based on a proof by Steel [1]. We also study two major applications of Kingman's theorem, convergence of products of random matrices and the asymptotic behaviour of the longest increasing subsequences of a random permutation.

Sammendrag

I denne masteroppgaven studerer vi Kingmans subadditiv ergodiske teorem og dens applikasjoner. Vi beviser Kingmans teorem basert på et bevis av Steel [1]. Vi ser også på to viktige konsekvenser av Kingmans teorem, konvergens for produkter av tilfeldige matriser og asymptotisk oppførsel av den lengste stigende undersekvensen av en tilfeldig permutasjon.

Contents

1. Introduction

In this paper we will study Kingman's subadditive ergodic theorem and its applications. The theorem covers sequences of functions defined on a measure space with a transformation satisfying certain criteria. These spaces are known as measure-preserving dynamical systems. Kingman's theorem is often viewed as a generalization of Birkhoff's ergodic theorem, which is one of the cornerstones of ergodic theory. One can also view Kingman's theorem as a random variable version of Fekete's subadditive lemma. The theorem is named after John Kingman, an English mathematician who proved the theorem in 1968 [2].

This paper is organized as follows. First we introduce subadditive sequences and Fetekes lemma, as a comparison to Kingman's theorem. We also consider an interesting example which follows from Fetekes lemma, i.e how to count the number of self-avoiding walks on a lattice. Then we state and prove Kingman's theorem. The proof presented is based on a paper by Steele [1]. Afterwards, as an introduction to Ligget's version of Kingman's theorem, we discuss stationary sequences. Stationary sequences are closely related to measure-preserving dynamical systems, and are needed when studying Ligget's version of Kingman's theorem and other applications. Subadditive stochastic processes were originally invented by Hammersley and Welsh [3] in order to deal with timedependent percolation processes: the study of connected clusters in a random graph. It turned out that subaddtivity had several more applications, and is viewed as one of the major achievements in ergodic theory in the second half of the 20th century.

In the next chapter we discuss the first major application of Kingman's theorem, products of random matrices. Given a set of non-singular matrices $\{A_1, A_2, \cdots, A_m\}$ and an associated probability vector $(p_1, p_2, \cdots, p_m)$ consider products of the type

$$\Pi_n = A_{i_1} A_{i_2} \cdots A_{i_n}$$

where $i_k = j$ with a probability of p_j. What can we say about Π_n as n approaches infinity? In 1960 Fustenberg and Kesten proved their well known theorem stating that

$$\lim_{n\to\infty} \frac{1}{n} \log\|\Pi_n\| \tag{1.1}$$

exists almost surely [4]. This was eight years prior to Kingman's theorem, and their proof was obviously independent of the subaddative theorem. It turns out however that Fustenberg-Kesten's theorem follows directly from Kingman's theorem, which is proven in chapter three.

Equation (1.1) is actually an example of a much studied concept in mathematics called Lyapunov exponents. These exponents are a quantitative number that measures the dependence on initial conditions, and will also be studied in chapter 3.

The last chapter is devoted to finding the asymptotic behaviour of the longest increasing subsequences of a random permutation. Given a permutation $\pi : i \to \pi(i)$ of order n we define an increasing subsequence of π as:

$$i_1 < i_2 < \cdots < i_k; \quad \pi(i_1) < \pi(i_2) < \cdots < \pi(i_k).$$

We let $L(\pi)$ be the longest such subsequence in a permutation π. In chapter 4 will we study l_n which is the average of $L(\pi)$ over all permutations of order n. It turns out that Kingman's theorem is applicable to find the limit of $l_n/\sqrt{n}$. The problem was first studied by Hammersley in the seventies [5]. By looking at each random transformation as random points in the plane, Hammersley constructed a Poisson point process of unit intensity and defined $l_{s,t}$ to be the longest increasing subset inside the box $[s,t) \times [s,t)$. By formulating the problem this way one is able to apply Kingman's theorem and show that $l_n \to c\sqrt{n}$ for a finite constant c. Much research has been devoted to finding the exact value of c, and it turns out that $c = 2$. This will not been shown in this paper, but a few bounds for c using different methods will be discussed.

This paper is based on previous work and already established theory. The goal was to get a deeper understanding for Kingman's subadditive ergodic theorem by studying the theorem, proving it and looking at several applications of it. Also a second goal was to make the text an easy to read paper for someone new to measure and ergodic theory. Much of the literature on the topic, especially on the longest increasing subsequences of a random permutations are quite complicated and requires much background in mathematics. Hopefully this paper can be a manageable introduction for someone just starting measure theory. Even though basically everything is done before, several examples and a few propositions are proved in a new way. Also, in addition to theorems and proofs, we have included a few reflections and thoughts on what we are proving. Hopefully this makes the paper a more interesting read, and can help the reader achieve a deeper understanding of mathematics in general. Mathematics is not just about theorems and proof, but a large part of it is reflections on what, why and how. Working on the longest increasing subsequence of a random permutation has been especially interesting. It combines several branches of mathematics and its level of difficulty is surprisingly high for such an easy to

understand-problem. This paper is the continuation of a shorter paper on Birkhoff's theorem, a project done the fall of 2016.

1.1. **Preliminaries.**

Definition 1.1. A *measure-preserving dynamical system* (MPDS) is a quartet $(\mathcal{X}, \mathcal{B}, \mu, T)$ where $(\mathcal{X}, \mathcal{B}, \mu)$ is a measure space and $T : \mathcal{X} \to \mathcal{X}$ is such that

(1) T is measurable: $E \in \mathcal{B} \Rightarrow T^{-1}E \in \mathcal{B}$
(2) μ is T-invariant: $\mu(T^{-1}E) = \mu(E)$ for all $E \in \mathcal{B}$

A *probability preserving transformation system* (PPT) is a MPDS where $\mathcal{X}$ is probability space.

Let $(\mathcal{X}, \mathcal{B}, \mu, T)$ be a MPDS.

Definition 1.2. A measurable set $E \in \mathcal{B}$ is called *T-invariant*, if $T^{-1}E = E$.

Definition 1.3. A MPDS $(\mathcal{X}, \mathcal{B}, \mu, T)$ is called *ergodic*, if every invariant set E satisfies $\mu(E) = 0$ or $\mu(X \setminus E) = 0$. We say μ is an ergodic measure.

Definition 1.4. A measurable function $f : \mathcal{X} \to \mathbb{R}$ is called *T-invariant* if $f \circ T = f$ a.e.

Definition 1.5. For all $n \in \mathbb{N}$, $T^{n+1} = T \circ T^n$, with $T^2 = T \circ T$ are called the *iterates* of T.

Proposition 1.1. *Let $(\mathcal{X}, \mathcal{B}, \mu, T)$ be a PPT. The following are equivalent*

(1) $(\mathcal{X}, \mathcal{B}, \mu, T)$ is ergodic.
(2) For all $A \in \mathcal{B}$ with $\mu(A) > 0$ we get $\mu(\cup_{n\geq 1} T^{-n}A) = 1$.
(3) For all $E \in \mathcal{B}$ such that $\mu(E \Delta T^{-1}E) = 0$ we have that $\mu(E) \in \{0, 1\}$.
(4) For any measurable function $f : \mathcal{X} \to \mathbb{R}$, if f is T-invariant, then $f =$ constant a.e.

For the proof of (1) $\Leftrightarrow$ (4) see [6], for the rest of the proof see [7].

Poisson distribution and Poisson point process. Ergodic theory and probability theory are closely related, while studying one you often encounter the other. We thus need some background in probability theory, specifically the Poisson point process, which is a field in spatial statistics that models random points on the plane. This process originates from the Poisson distribution, for more details on that see [8].

Definition 1.6. A stochastic process $\{N(t), t \geq 0\}$ is said to be a *counting process* if $N(t)$ satisfies:

(1) $N(t) \geq 0$.
(2) $N(t)$ is integer valued.
(3) If $s < t$, then $N(s) \leq N(t)$.
(4) For $s < t, N(t) - N(s)$ equals the number of events that occur in the interval $(s, t]$.

Definition 1.7. A function f is said to be $o(h)$ is

$$\lim_{h \to 0} \frac{f(h)}{h} = 0.$$

Definition 1.8. A counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process with rate λ if the following holds:

(1) $N(0) = 0$.
(2) $\{N(t), t \geq 0\}$ has independent increments.
(3) $P(N(t+h) - N(t) = 1) = \lambda h + o(h)$.
(4) $P(N(t+h) - N(t) \geq 2) = o(h)$.

The reason the process is called a Poisson process is because the number of events that occurs in any interval of length t is Poisson distributed with mean λt.

Theorem 1.1. *If $\{N(t), t \geq 0\}$ is a Poisson process with rate $\lambda > 0$, then for all $s > 0, t > 0, N(t+s) - N(s)$ is a Poisson random variable with mean λt.*

See theorem 5.1 in [9] for a proof.

Now let $\{N(t), t \geq 0\}$ be a Poisson process with rate $\lambda > 0$ and let S_n be the arrival time of the nth event, that is

$$S_n = \inf\{t > 0, N(t) = n\}.$$

For $n > 1$, let T_n be the elapsed time between the $(n-1)$th and the nth event, that is $T_n = S_n - S_{n-1}$. The sequence $\{T_n, n \in \mathbb{N}\}$ is called the *sequence of interarrival times.*

Theorem 1.2. *$T_n, n = 1, 2, 3 \cdots$, are i.i.d exponential random variables with mean $1/\lambda$.*

See [9] for a proof.

Definition 1.9. A *Poisson point process* Π is a collection of points in $\mathbb{R}^2$ with intensity λ that is characterized by two properties:

(1) The number of points of Π in any bounded set B follows a Poisson distribution with mean $\lambda m(b)$ where m is the Lebesgue measure in $\mathbb{R}^2$.

(2) The numbers of points of Π in k disjoint sets form k independent random variables, for arbitrary k.

The concept can be defined on an abstract space, but we only need it for $\mathbb{R}^2$.

2. Kingman's subadditive ergodic theorem

Kingman's subadditive ergodic theorem is one of the most important theorems in ergodic theory. The theorem has much application in other areas of mathematics, we will study a couple of them later. The theorem is viewed as a generalization of Birkhoff's ergodic theorem, which is one the fundamental theorems in ergodic theory.

Theorem 2.1 (Birkhoff's pointwise ergodic theorem)**.** *Let $(\mathcal{X}, \mathcal{B}, \mu, T)$ be a PPT and let $f : \mathcal{X} \to \mathbb{R}$ be measurable function. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n-1} f(T^i x) = f^*(x) \text{ a.e.}$$

where $f^*(x)$ is an invariant function.

Unlike Birkhoff's theorem, where only one measurable function is considered, Kingman's theorem looks at limit laws for a sequence of functions. Specifically, we look at sequences of measurable functions $\{f_n\}_{n\in\mathbb{N}}$ on a PPT $(\mathcal{X}, \mathcal{B}, \mu, T)$ that satisfies

$$f_{n+m}(x) \leq f_n(x) + f_m(T^n x) \quad \text{for all } n, m \geq 1 \quad \text{a.e.} \tag{2.1}$$

Example 2.2. Let $(\mathcal{X}, \mathcal{B}, \mu, T)$ be MPDS and $A : \mathcal{X} \to GL_d$ be a measurable function. Here GL_d denotes the general linear group of degree d, which is the set of $d \times d$ invertible matrices. Define

$$A^{(n)}(x) = A(T^{n-1}x) \cdots A(Tx)A(x).$$

Then the sequence $\phi_n(x) = \log\|A^{(n)}(x)\|$ satisfies equation (2.1). Indeed, first of

$$A^{(n+m)} = A^{(n)}(T^m x)A^{(m)}(x).$$

Then since

$$\|B_1 B_2\| \leq \|B_1\|\|B_2\| \text{ for all } B_1, B_2 \in GL(d)$$

we get that

$$\phi_{n+m}(x) \leq \phi_m(x) + \phi_n(T^m x).$$

Since the functions we are studding are subadditive, we will look at this concept for motivation.

2.1. Subadditivity.

Subadditivity is an important concept not only in measure theory, but in several fields of mathematics. There are numerous examples of subadditive functions in various areas of mathematics, particularly norms and square roots.

Definition 2.1. A sequence $\{a_n\}_{n\in\mathbb{N}}$ is called subadditive if for all m, n we have that $a_{n+m} \leq a_m + a_n$.

To increase our understanding of Kingman's theorem, let us look at a similar result in $\mathbb{R}$ called Fekete's subadditive lemma.

Lemma 2.1. *For every subadditive sequence* $\{a\}_{i=0}^{\infty}$ *the limit*

$$\lim_{n\to\infty} \frac{a_n}{n}$$

exists and is equal to $\inf \frac{a_n}{n}$.

Proof. If $a_n = -\infty$ for some n, then for all $m > n$, $a_m = -\infty$ as well. Then both sides of equality are equal to $-\infty$, and the lemma holds. Now assume that $a_n > -\infty$ for all n, and let $L = \inf \frac{a_n}{n}$. Pick any $B > L$ and let $k \geq 1$ be such that

$$\frac{a_k}{k} < B.$$

For any $n > k$ apply the division algorithm, that is $n = p_n k + q_n$ for integers p_n, q_n where $0 \leq q_n < k$. Now apply the subadditivity repetitively to obtain

$$a_n = a_{p_n k + q_n} \leq a_{p_n k} + a_{q_n} \leq p_n a_k + a_{q_n}.$$

When dividing this by n we get

$$\frac{a_n}{n} \leq \frac{p_n k}{n}\frac{a_k}{k} + \frac{a_{q_n}}{n}.$$

When n goes to infinity $\frac{p_n k}{n}$ converges to 1 and $\frac{a_{q_n}}{n}$ converges to 0. So we have for all $B > L$

$$L \leq \limsup_{n\to\infty} \frac{a_n}{n} \leq \frac{a_k}{k} \leq B.$$

Hence we have that

$$L \leq \liminf_{n\to\infty} \frac{a_n}{n} \leq \limsup_{n\to\infty} \frac{a_n}{n} \leq B.$$

Now let B go to L and we obtain

$$L = \inf_n \frac{a_n}{n} = \lim_{n\to\infty} \frac{a_n}{n}.$$

□

This famous lemma helps us understand Kingman's theorem, as it can be viewed as a non-random version of Kingman's theorem.

As mentioned in the introduction, the theory of subadditivity arose while studying random flows in lattice. We will consider such an example now as an illustration of Fekete's lemma.

Example 2.3. A self-avoiding walk is a sequence on a lattice which never intersect itself. See figure 1 for an example on the square grid graph.

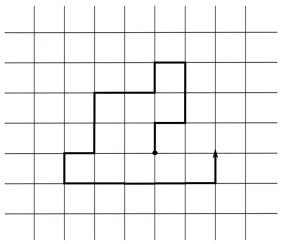

FIGURE 1. A self-avoiding walk of 18 steps.

We will try to count the number $K(n)$ of all such self-avoiding walks with n steps on the square grid graph. We have that $K(1) = 4$ and $K(2) = 12$, see figure 2 and 3. We want to use Fekete's subadditive lemma to find the asymptotic behaviour of $K(n)$. The sequence is not subadditive, but we can show that $K(m+n) \leq K(n)K(m)$. Indeed, we can consider all walks of length $m+n$ as the concatenation of a self-avoiding walk of length m followed by one of length n. By sticking together all self-avoiding walks of length m with self-avoiding walks of length n we get all non-intersecting paths of length $m+n$, plus some intersecting ones. Now, since log is a strictly increasing function, $\log K(n)$ is a subadditive sequence. Thus we have that

$$\lim_{n\to\infty} \frac{\log K(n)}{n} = A,$$

for some A. Put differently we get

$$\lim_{n\to\infty} K(n)^{\frac{1}{n}} = \mu.$$

μ is known as the *connective constant*, since $K(n)$ depends on the particular lattice chosen for the walk so does μ. Let us try to give an estimate for μ on the square grid graph. First note that $2 \leq \mu \leq 3$, since there are always maximum three new directions a path can take each step. μ is also bounded below by two since one can always choose ether to go up or to the left on each step and never intersect the path later. Similarly, paths that only goes up and right, down and right, or down and left will never intersect each other. Thus, we can improve our lower bound, $K(n) \geq 4 \cdot 2^n - 4$. Similarly, $K(n) \leq 4 \cdot 3^{n-1}$, since the first step can go in four directions and after that there are maximum

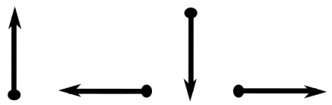

FIGURE 2. All self-avoiding walks of 1 step.

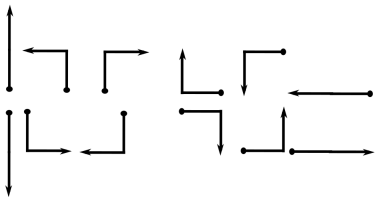

FIGURE 3. All self-avoiding walks of 2 steps.

three options each step. The exact value of μ is only known for the hexagonal lattice, where it is equal to

$$\sqrt{2+\sqrt{2}}$$

which was first proven in 2010 by Duminil-Copin and Smirnov [10]. For the square grid lattice μ is believed to approximately 2.63815853, a number taken from a paper by Jensen and Guttmann [11].

Example 2.4. Assume we have two sequences $(a_i)_{i\geq 0}$ and $(b_i)_{i\geq 0}$ taking values in $\{0, 1, \cdots, l-1\}$ for a $l > 0$. A typical problem is to find the longest common subsequence of these two sequences up to an integer n. Define $C_{n,l}$ to be the longest common subsequence up to n. That is

$$\begin{aligned} C_{n,l} = &\max\{K : a_{i_k} = b_{j_k} \text{ for } 1 \leq k \leq K, \text{ where} \\ &0 < i_1 < i_2 < \cdots < i_K \leq n \text{ and } 0 < j_1 < j_2 < \cdots < j_K \leq n\}. \end{aligned}$$

Let $c_{n,l}$ be the expected value of $C_{n,l}$, then $c_{n,l}$ is superadditive on n. That is

$$c_{n+m,l} \geq c_{n,l} + c_{m,l}.$$

This is because sequences of length $n+m$ can be broken into subsequences of length n and m, and the longest common subsequences from those sequences can be combined to from a common subsequence of the whole string. We therefore get that $-c_{n,l}$ is a subadditive sequence and by Fekete's lemma we get that

$$\lambda_l = \lim_{n\to\infty} \frac{c_{n,l}}{n}.$$

These constants are known as the Chvatal-Sankoff constants, which exact values are not known today. In 2010 Lueker showed that $0.788071 \leq \lambda_2 \leq 0.826280$ [12]. Kiwi, Loebl and Matousekb showed in 2004 [13] that the Chvatal-Sankoff constant grows inversely proportional to the square root of l, that is

$$\lim_{l\to\infty} \lambda_l \sqrt{l} = 2.$$

The longest common subsequence problem is closely related to the longest increasing subsequence problem: The longest increasing subsequnece of a permutation π is the same as the longest common subsequence between π and the numbers 1 to n in increasing order.

2.2. **Kingman's subadditive ergodic theorem.** Before we state the theorem we need a proposition.

Proposition 2.2. *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of functions on a MPDS $(\mathcal{X}, \mathcal{B}, \mu, T)$ which satisfies the following*

$$f_{n+m}(x) \leq f_n(x) + f_m(T^n x) \quad \text{a.e. for all } n, m \geq 1. \tag{2.2}$$

Assume we have a disjoint partition of the interval $[1, n)$

$$[1, n) = [1, n_1) \cup [n_1, n_1 + l_1) \cup [n_2, n_2 + l_2) \cup \cdots \cup [n_m, n_m + l_m)$$

where $1 \leq n_i, l_1 \leq n$ and $n_i + l_i = n_{i+1}$. Then we have that

$$f_n(x) \leq f_{n_1}(x) + f_{l_1}(T^{n_1}x) + \cdots + f_{l_i}(T^{n_i}x) + \cdots + f_{l_m}(T^{n_m}x).$$

Proof. The proof is just repetitive use of (2.2). Let us do the first few steps

$$\begin{aligned} f_n(x) &\leq f_{n_1}(x) + f_{n-n_1}(T^{n_1}x) \leq f_{n_1}(x) + f_{l_1}(T^{n_1}x) + f_{n-n_2}(T^{n_1+l_1}x) \\ &\leq f_{n_1}(x) + f_{l_1}(T^{n_1}x) + f_{l_2}(T^{n_2}x) + f_{n-n_3}(T^{n_3}x). \end{aligned}$$

The proposition follows by induction on n. □

Theorem 2.5 (Kingman's subadditive ergodic theorem)**.** *Let $(\mathcal{X}, \mathcal{B}, \mu, T)$ be a PPT. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions $f_n : \mathcal{X} \to \mathbb{R}$ such that $f_1 \in L_1(\mu)$ and*

$$f_{n+m}(x) \leq f_n(x) + f_m(T^n x) \quad \text{for all } n, m \geq 1 \quad \text{a.e.}$$

Then

$$\lim_{n\to\infty} \frac{f_n(x)}{n} = f(x) \geq -\infty \text{ a.e.}$$

where $f(x)$ is an invariant function.

Remark 2.1. Before we give a proof, note that Fekete's subadditive lemma follows directly from this theorem. To see this let $\{a_n\}$ be a subadditive sequence. Then pick any ergodic PPT $(\mathcal{X}, \mathcal{B}, \mu, T)$. Now define a sequence of constant functions $\{f_n\}_{n\in\mathbb{N}}$ on $\mathcal{X}$, defined by $f_n(x) = a_n$ for all $x \in \mathcal{X}$. Then clearly $f_{n+m}(x) \leq f_n(x) + f_m(T^n x) \quad$ for all $n, m \geq 1$ since a_n is a subadditive sequence. Also, since the space is a probability space, each function f_n is absolutely integrable. Thus, by theorem 2.5

$$\lim_{n\to\infty} \frac{a_n}{n} = \lim_{n\to\infty} \frac{f_n(x)}{n} = \inf_n \frac{f_n(x)}{n} = \inf_n \frac{a_n}{n} \text{ a.e.}$$

On the other hand we can use Fekete's lemma to show that

$$\lim_{n\to\infty} \frac{1}{n} \int_{\mathcal{X}} f_n(x) = \inf_n \frac{1}{n} \int_{\mathcal{X}} f_n(x).$$

Simply put $a_n = \int f_n(x)$ and apply the lemma.

Remark 2.2. We want to compare Kingman's theorem to Birkhoff's theorem to see why this is an even stronger theorem. First let us understand why Birkhoff's theorem is a special case of Kingman's theorem. Assume that we have a L_1 function $f : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X}$ is an ergodic space. Birkhoff's theorem says that

$$\sum_{i=0}^{n-1} \frac{f(T^i x)}{n}$$

converges almost surely. Now define

$$f_n(x) = \sum_{i=0}^{n-1} f(T^i x),$$

then clearly each f_n is an absolutely integrable function and

$$f_{n+m}(x) = f_n(x) + f_m(T^n x).$$

That is, we do not only have the subadditive property, but the sequence is also additive. Thus, from Kingman's theorem, $\frac{f_n}{n}$ converges almost surely. The reason we are interested in the subadditive property instead of the additive property is that there are far more interesting applications that satisfies this property, a few of them will be studied later. In our proof of Kingman's theorem we used Birkhoff's theorem. It is possible to prove it without using it [14], so Birkhoff's theorem can be viewed as a direct consequence of Kingman's theorem.

Proof. This proof is based on a paper by Steel from 1989 [1]. The main part of the proof is under on the assumption that $f_m(x) \leq 0$ for all x.

This is however not part of the assumption, so we need to define a new function:

$$f'_m(x) = f_m(x) - \sum_{i=0}^{m-1} f_1(T^i x). \tag{2.3}$$

These functions are less than zero for every x and it also satisfies the subadditivity property. Indeed,

$$\begin{aligned} f'_{n+m}(x) &= f_{n+m}(x) - \sum_{i=0}^{n+m-1} f_1(T^i x) \\ &\leq f_n(x) + f_m(T^n x) - \sum_{i=0}^{n-1} f_1(T^i x) - \sum_{i=n}^{n+m-1} f_1(T^i x) \\ &\leq f_n(x) - \sum_{i=0}^{n-1} f_1(T^i x) + f_m(T^n x) - \sum_{i=0}^{m-1} f_1(T^{n+i} x) \\ &\leq f'_n(x) + f'_m(T^n x). \end{aligned}$$

We want to show almost surely convergence for

$$\frac{1}{m} f_m.$$

If we can show that both

$$\frac{1}{m} f'_m \text{ and } \frac{1}{m} \sum_{i=1}^{m-1} f_1(T^i x)$$

converges almost surely, we get by equation (2.3) that $\frac{1}{m} f_m$ converges almost surely. The sum is a Birkhoff sum, so Birkhoff's ergodic theorem provides convergence for this term. We are therefore left to show almost sure convergence for $\frac{1}{m} f'_m(x)$. For simplicity we will write $f_m(x)$ and assume that $f_m(x) \leq 0$.

Our goal is to show that

$$\limsup_{n\to\infty} \frac{f_n(x)}{n} \leq \liminf_{n\to\infty} \frac{f_n(x)}{n}. \tag{2.4}$$

Now define

$$f(x) = \liminf_{n\to\infty} \frac{f_n(x)}{n}$$

and note that this function is T-invariant. Indeed, we have that

$$\frac{f_{n+1}(x)}{n} \leq \frac{f_1(x)}{n} + \frac{f_n(Tx)}{n},$$

taking limit inferior we see that $f(x) \leq f(Tx)$. Thus, this tells us that $\{x : f(x) > \alpha\} \subset T^{-1}\{x : f(x) > \alpha\}$. T is a measure preserving transformation, so the sets can differ by at most a set of measure zero. That is, $f(x) = f(Tx)$ almost surely.

Let $\epsilon > 0$ and $0 < M < \infty$, and define

$$G_M(x) = \max\{-M, f(x)\}.$$

The goal of this proof is to try to bound $\limsup f_n$ by the function $G_M(x)$. Since $G_M(x)$ is itself bounded by $f(x)$ we therefore get equation (2.4) by showing this.

Pick a $0 < N < \infty$ and consider the set

$$B(N, M) = \{x : f_l(x) > l(G_M(x) + \epsilon) \text{ for all } 1 \leq l \leq N\}$$

and its compliment $A(N, M) = B(N, M)^c$. Let $x \in \mathcal{X}$ and $n \geq N$. We want to decompose the integer set $[1, n)$ into a union of three classes of intervals. First let $k = 1$. Then each step take the least integer k in $[1, n)$ which is not in an interval already considered and look at $T^k x$. Remember that if $T^k x \in A(N, M)$, then there is an $l \leq N$ so that $f_l(T^k z) \leq l(G_M(T^k x) + \epsilon) = l(G_M(x) + \epsilon)$. We have three different cases

- If $T^k x \in A(N, M)$ and $k + l \leq n$ we use the interval $[k, k + l)$.
- If $T^k x \in A(N, M)$ and $k + l > n$ we use $[k, k + 1)$.
- If $T^k x \in B(N, M)$ we also take $[k, k + 1)$.

So for any $x \in \mathcal{X}$ we have a decomposition of $[1, n)$ into a set of u intervals on the form $[\tau_i, \tau_i + l_i)$ where $f_{l_i}(T^{\tau_i} x) \leq l_i(G_M(x) + \epsilon)$ with $1 \leq l_i \leq N$, a set of v singletons $[\sigma_i, \sigma_i + 1)$ where $T^{\sigma_i} x \in B(N, M)$ and another set of w singletons $[\rho_i, \rho_i + 1)$ which are all in $(n - N, n)$. By proposition 2.2 we can bound $f_n(x)$

$$f_n(x) \leq \sum_{i=1}^{u} f_{l_i}(T^{\tau_i} x) + \sum_{i=1}^{v} f_1(T^{\sigma_i} x) + \sum_{i=1}^{w} f_1(T^{\rho_i} x).$$

Remember that we assumed $f_m \leq 0$ so we can bound f_n even further by removing the last two sums,

$$f_n(x) \leq \sum_{i=1}^{u} f_{l_i}(T^{\tau_i} x).$$

We also know that $f_{l_i}(T^{\tau_i} x) \leq l_i(G_M(x) + \epsilon)$ hence we get that

$$f_n(x) \leq (G_M(x) + \epsilon) \sum_{i=1}^{u} l_i \leq G_M(x) \sum_{i=1}^{u} l_i + n\epsilon.$$

Thus, we get

$$\limsup_n \frac{1}{n} f_n(x) \leq \limsup_n \left(G_M(x) \frac{1}{n} \sum_{i=1}^{u} l_i \right) + \epsilon.$$

Remember that since $G_M(x) \leq 0$, the previous equation can be reformulated to

$$\limsup_n \frac{1}{n} f_n(x) \leq G_M(x) \liminf_n \frac{1}{n} \sum_{i=1}^{u} l_i + \epsilon.$$

We therefore want a lower bound on $\sum_{i=1}^{u} l_i$. By the construction of the intervals we have

$$\sum_{i=1}^{u} l_i \geq n - \sum_{k=1}^{n} 1_{B(N,M)}(T^k x) - N.$$

Thus,

$$\frac{1}{n} \sum_{i=1}^{u} l_i \geq 1 - \frac{1}{n} \sum_{k=1}^{n} 1_{B(N,M)}(T^k x) - \frac{N}{n}.$$

Note that the sum is a Birkhoff sum, so we can apply Birkhoff's theorem:

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{u} l_i \geq 1 - \mu(B(N,M) \text{ a.e.}$$

Putting everything together we get

$$\limsup_{n\to\infty} \frac{f_n(x)}{n} \leq G_M(1 - \mu(B(N,M))) + \epsilon \text{ a.e.}$$

We claim that $1_{B(N,M)} \to 0$ a.e. as $N \to \infty$. To see this let us rewrite $B(N,M)$

$$B(N,M) = \cap_{l=1}^{N} \{x : f_l(x) > l(G_M(x) + \epsilon)\}.$$

Note that $B(1,M) \supset B(2,M) \supset \cdots \supset B(N,M) \cdots$ so we want to show that

$$\cap_{N\in\mathbb{N}} B(N,M)) = \emptyset.$$

If $f(x) > -M$ then $x \in \cap_{N\in\mathbb{N}} B(N,M)$ would mean that

$$\frac{1}{l} f_l(x) > f(x) + \epsilon \text{ for all } l.$$

and that would mean that $f(x) > f(x) + \epsilon$ which is clearly a contradiction. If $f(x) < -M$ then $x \in \cap_{N\in\mathbb{N}} B(N, M)$ would mean that

$$\frac{1}{l} f_l(x) > -M + \epsilon \text{ for all } l$$

which would mean that $f(x) > -M + \epsilon$, also a contradiction.

We have shown that

$$\limsup_{n\to\infty} \frac{f_n(x)}{n} \leq G_M(1 - \mu(B(N, M)) + \epsilon \text{ a.e.}$$

Then for a fixed M and letting $N \to \infty$

$$\limsup_{n\to\infty} \frac{f_n(x)}{n} \leq G_M(x) + \epsilon \text{ a.e.}$$

This holds for all $M \geq 0$ and $\epsilon > 0$, we get that

$$\limsup_{n\to\infty} \frac{f_n(x)}{n} \leq \liminf_{n\to\infty} \frac{f_n(x)}{n},$$

which is exactly what we wanted to show. □

2.3. **Stationary sequences.** The way we formulated Kingman's theorem was a bit different from how Kingman himself did it originally [2]. He based everything on stochastic processes and much of the applications of his theorem is therefore in this format.

Definition 2.2. Let $\{X_n\}$ be a sequence of random variables defined on a space $\mathcal{X}$. The sequence is said to be a *stationary sequence* if for every k and each m, $(X_0, \cdots, X_m)$ and $(X_k, \cdots, X_{k+m})$ have the same distribution.

See chapter 4 in [15] for details on joint probability distribution.

Example 2.6. If $X_0, X_1, \cdots$ are i.i.d then the sequence is stationary.

Example 2.7. Let $(\mathcal{X}, \mathcal{B}, \mu, T)$ be a MPDS and let $f : \mathcal{X} \to \mathbb{R}$ be a measurable function. Define $X_n(x) = f(T^n x)$, we claim that this sequence is a stationary sequence. Let $B \in \mathbb{R}^{n+1}$ and $A = \{x : (X_0(x), \cdots, X_n(x)) \in B\}$. Then for a k

$$\begin{aligned} P((X_k, \cdots, X_{k+n}) \in B) &= P(T^k x \in A) = P(x \in A) \\ &= P((X_0, \cdots, X_n) \in B). \end{aligned}$$

We will refer to the sequence in example 2.7 as the stationary sequence generated by T.

Remark 2.3. It turns out that stationary sequences are closely related to the study of measure-preserving dynamical systems. As we saw in example 2.7, it is possible to construct a stationary sequence with a MPDS and a random variable. Vica versa, if we begin with a stationary sequence, this leads in a natural way to a MDPS. Let $\{X_n\}$ be a random process on $\mathcal{X}$ and let $(\Omega, \mathcal{B}, \mu)$ be the associated probability space, where $\Omega = \mathcal{X}^{\mathbb{N}}$ is the sample space. Note that $X_n(\omega) = \omega_n$. Let $T : \Omega \to \Omega$ be the left shift operator:

$$T(\omega_1, \omega_2, \omega_3, \cdots) = (\omega_2, \omega_3, \cdots).$$

The process $\{X_n\}$ is said to be stationary if T is invariant under μ:

$$\mu(T^{-1}A) = \mu(A) \quad \text{for every } A \in \mathcal{B}.$$

This definition is equivalent to the standard definition of stationary sequences. See Chapter 16 in [16] for more details.

Definition 2.3. We call the transformation discussed in remark 2.3 the *associated transformation* to the stationary sequence.

Since the study of stationary sequences and measure preserving dynamical systems coincides it is natural to reuse a definition.

Definition 2.4. A stationary sequence is said to be ergodic if its associated transformation T is an ergodic transformation.

Definition 2.5. Let $\{X_n\}$ be a sequence of random variables. Define

$$\tau_n = \sigma(X_n, X_n + 1, \cdots), \quad \tau = \cap_n \tau_n.$$

Then τ is a σ-algebra, known as the *tail-algebra* of $\{X_n\}$. If $E \in \tau$ then E is called a *tail event*.

Proposition 2.3. *Every shift-invariant set is a tail event.*

For a proof see [17] page 31. For more details on stationary sequences and their associated dynamical systems see [18].

Example 2.8. Let $X_1, X_2, , \cdots$ be i.i.d. Then the process is ergodic. By proposition 2.3 we have that every shift-invariant set is a tail event. Since $X_1, X_2, \cdots$ are all independent Kolmogorov's zero-one-law states that every tail event occur with a probability of either 0 or 1. Thus we have that for every invariant set A, $\mu(A) = 0$ or 1.

Till now, we have looked at stationary sequences index by one variable. In some applications this is not general enough, which is why we want to look at sequences index by two indices. Let us begin with an example for motivation.

Example 2.9. Example 2.4 can be viewed as a stationary sequence index by two indices. Assume we have two ergodic stationary sequences $\{X_n\}$ and $\{Y_n\}$ taking values in $\{0, 1, \cdots l-1\}$ for a $l > 0$. Define $L_{m,n}$ to be the longest common subsequence between m and n. That is

$$\begin{aligned} L_{m,n} = &\max\{K : X_{i_k} = Y_{j_k} \text{ for } 1 \le k \le K, \text{ where} \\ &m < i_1 < i_2 < \cdots < i_K \le n \text{ and } m < j_1 < j_2 < \cdots < j_K \le n\}. \end{aligned}$$

$-L_{m,n}$ is actually a superadditive sequence. Indeed, we have that

$$L_{0,n} \ge L_{0,m} + L_{m,n}.$$

2.4. **Ligget's version.** Previously we proved Kingman's theorem in a form that resembles Birkhoff's ergodic theorem. The original version of Kingman's theorem was in a different form however. Later we will see an improved version of Kingman's theorem known as Ligget's version which covers more applications and the longest common subsequence problem in particular. Ligget's version is a extension of the original version, so we will state this now.

Theorem 2.10. *Let $(\mathcal{X}, \mathcal{B}, \mu, T)$ be a PPT. If there exists a family of random variables index by two parameters, $\{X_{m,n}, 0 \le m < n < \infty\}$ where each $X_{m,n}$ is integrable with respect to μ and the family satisfy*

- $X_{0,n} \le X_{0,m} + X_{m,n}$
- $X_{m+1,n+1} = X_{n,m} \circ T$

Then

$$\lim_{n\to\infty} \frac{X_{0,n}}{n} = Y \text{ a.e.}$$

where $Y \in [-\infty, \infty)$ is T-invariant and a constant if $\mathcal{X}$ is ergodic.

Remark 2.4. This is indeed the same version as seen earlier. First assume theorem 2.5 and want to show theorem 2.10. We have a family of random variables $\{X_{m,n}, 0 \le m < n < \infty\}$ that satisfies the conditions in theorem 2.10. Then define

$$\begin{aligned} f_m(x) &= X_{0,m} \\ f_{n-m}(T^m x) &= X_{m,n}. \end{aligned}$$

Then we get that

$$f_n(x) \le f_m(x) + f_{n-m}(T^m x)$$

is the same as

$$X_{0,n} \le X_{0,m} + X_{m,n}.$$

Thus we have by theorem 2.5 that

$$\lim_{n\to\infty} \frac{X_{0,n}}{n} = \lim_{n\to\infty} \frac{f_n(x)}{n} = \inf_n \frac{1}{n} \int_{\mathcal{X}} f_n(x) := Y \text{ a.e.}$$

Now we assume theorem 2.10 and we want to show theorem 2.5. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of functions that satisfies the conditions in theorem 2.5. Then define

$$X_{m,n} = f_{n-m}(T^m x).$$

Again we get that

$$f_n(x) \leq f_m(x) + f_{n-m}(T^m x)$$

is the same as

$$X_{0,n} \leq X_{0,m} + X_{m,n}.$$

We also have that

$$\begin{aligned} X_{m+1,n+1} &= f_{n+1-(m+1)}(T^{m+1}x) = f_{n-m}(T^{m+1}x) \\ &= (f_{n-m}(T^m) \circ T)(x) = X_{n,m} \circ T. \end{aligned}$$

Thus the conditions in theorem 2.10 are satisfied and we get that

$$\lim_{n\to\infty} \frac{f_n(x)}{n} = \lim_{n\to\infty} \frac{X_{0,n}}{n} = Y \text{ a.e.}$$

As mentioned at the beginning of the section, some applications are not covered by this theorem. In 1985 Ligget [19] discovered an improved version, which we will need later when studying the longest increasing subsequence problem.

Theorem 2.11. *Suppose $X_{m,n}$ is a collection of random variables indexed by integers satisfying $0 \leq m < n$ and assume:*

(1) $X_{0,n} \leq X_{0,m} + X_{m,n}$
(2) $\{X_{nk,(n+1)k}, n \geq 1\}$ is a stationary sequence for each k.
(3) The distribution of $\{X_{m,m+k}, k \geq 1\}$ does not depend on m,
(4) For each n, $\mathbb{E}\left|X_{0,1}\right| < \infty$ and $\mathbb{E}X_{0,n} \geq cn$ where $c > -\infty$.

Then

(1) $\lim_{n\to\infty} \frac{\mathbb{E}X_{0,n}}{n} = \inf_m \frac{\mathbb{E}X_{0,m}}{m} \equiv \gamma$
(2) $\lim_{n\to\infty} \frac{X_{0,n}}{n} = X$ exists a.e. and in L^1, so $\mathbb{E}X = \gamma$.
(3) If all the stationary sequences in (2) are ergodic then $X = \gamma$ a.e.

For a proof see Ligget's own paper on the theorem [19].

Example 2.12 (First passage percolation)**.** First passage perlocation is a well-known problem in mathematics, and is the reason for several tools in mathematics, including the subadditive ergodic theorem [20]. Hammersley first studied percolation theory in 1957 [21], and it was because of first passage percolation he introduced subadditive stochastic processes a few years later [3]. Consider $\mathbb{Z}^d$ for a $d > 0$ as a graph with connecting edges $e = (x, y)$ for $x, y \in \mathbb{Z}^d$ if $\left|x - y\right| = 1$.

Add an independent non-negative random variable $\tau(e)$ for each edge in the graph, which represent the required time it takes to travel in either direction. The collection $\tau(e)$ is assumed to be independent, identically distributed with common distribution F. If $x, y \in \mathbb{Z}^d$ then a path from x to y is a sequence $x = x_1, x_2, \cdots, x_n = y$ such that $\left|x_m - x_{m-1}\right| = 1$. For a path we define the travel time to be $\tau(x_1, x_2) + \tau(x_2, x_3) + \cdots + \tau(x_{n-1}, x_n)$. For two vertices x, y we define the passage time from x to y as

$$t(x, y) = \inf\{\tau(e) | e \text{ is a path from } x \text{ to } y\}.$$

Now define

$$X_{m,n} = t(mu, nu)$$

where $u = (1, 0, \cdots, 0)$. Durret [22] shows that $X_{m,n}$ satisfies the conditions of theorem 2.11 including the ergodic property. Therefore, we know that

$$\frac{X_{0,n}}{n} \to \gamma(F)$$

for a constant $\gamma(F)$ depending on the distribution F. The value of $\gamma(F)$ is not known today, but some results are proved. For example under certain conditions $\gamma(F)$ varies continuously in F with respect to weak convergence [23]. See the book *50 years of first passage percolation* by Antonio Auffinger, Michael Damron and Jack Hanson for a comprehensive work on the topic [20].

3. Products of random matrices

We will now look at the first major consequence of Kingman's theorem, products of random matrices. Given a set of non-singular matrices $\{A_1, A_2, \cdots, A_m\}$ and an associated probability vector $(p_1, p_2, \cdots, p_m)$ consider products of the type

$$\Pi_n = A_{i_1} A_{i_2} \cdots A_{i_n}$$

where $i_k = j$ with a probability of p_j. What can we say about Π_n as $n \to \infty$?

Example 3.1. Consider the matrices

$$A_0 = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma^{-1} \end{pmatrix} \quad A_1 = \begin{pmatrix} \sigma^{-1} & 0 \\ 0 & \sigma \end{pmatrix}, \quad \text{where } \sigma > 1.$$

with probability p_0 and p_1 respectively. We see that $A_0 A_1 = I = A_1 A_0$. For a given n and Π_n, say that m_0 of the terms are A_0 and $m_1 = n - m_0$ of the terms are A_1. If $m_0 > m_1$, then all m_1 A_1's would cancel with m_1 A_0's and we are left with $m_0 - m_1$ A_0's. Thus Π_n becomes

$$\Pi_n = A_0^{m_0 - m_1} = \begin{pmatrix} \sigma^{m_0 - m_1} & 0 \\ 0 & \sigma^{-(m_0 - m_1)} \end{pmatrix}$$

If $m_1 > m_0$ then the opposite happens and we are left with $m_1 - m_0$ A_1's and Π_n becomes

$$\Pi_n = A_1^{m_1 - m_0} = \begin{pmatrix} \sigma^{-(m_1 - m_0)} & 0 \\ 0 & \sigma^{m_1 - m_0} \end{pmatrix},$$

If $m_0 = m_1$ then $\Pi_n = I$. Assume $p_0 > p_1$ then

$$\Pi_n \to \begin{pmatrix} \sigma^{n(p_0 - p_1)} & 0 \\ 0 & \sigma^{n(p_1 - p_0)} \end{pmatrix} \quad \text{as } n \to \infty.$$

In general Π_n dose not always convergence, and it is hard to extract any vital information. It is therefore interesting to consider

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \log \|\Pi_n\| \tag{3.1}$$

instead. This is actually an example of a general concept in mathematics called Lyapunov exponent. Fustenberg and Kesten proved in 1960 [4] that equation (3.1) exists. This was eight years prior to Kingman presented his theorem. It turns out that their theorem is an easy consequence of the subadditive ergodic theorem. The goal of this section is to prove this, and we need a more formal setup in order to do this.

3.1. **The Lyapunov exponent.** Say we have a space $\mathcal{X}$ and a time-dependent transformation on it such as a differential equation

$$x' = f(t, x).$$

Consider an initial value x_0 and a point close $x_0 + \epsilon$ for a small $\epsilon > 0$. Then $f(t, x)$ and $f(t, x + \epsilon)$ will have a separation depending on t, see figure 4. We are interested in studying the asymptotic behaviour of the separation when considering two initially close points. Specifically, we are interested in studying the limit

$$\lim_{\substack{t\to\infty \\ \epsilon\to 0}} \frac{1}{n} \log\left(\frac{\Delta(f(t, x_0), f(t, x_0 + \epsilon))}{\Delta(x_0, x_0 + \epsilon)}\right).$$

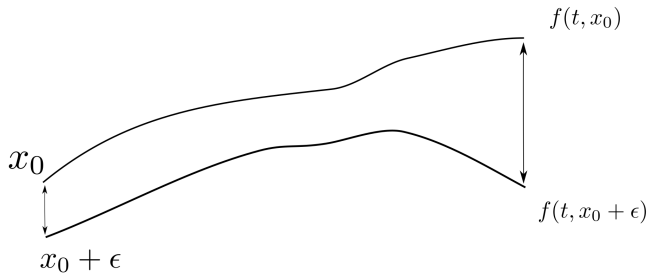

FIGURE 4. Separation of two initial close points using a differential equation $x' = f(t, x)$.

Consider instead a discrete transformation $T : \mathcal{X} \to \mathcal{X}$. For a point $x \in \mathcal{X}$, we say *the orbit of* x is the set $\{T^n x\}_{n\geq 0}$. If we consider two points in this space, x_0 and y_0, each of them will generate an orbit in $\mathcal{X}$. Let us use one of the orbits as a reference orbit, then the separation of the other orbits can also be viewed as a function of n. That is, for a $x_0 \in \mathcal{X}$, consider the point $x_0 + \delta x$ for a small δx. Say $\Delta(x, y)$ is a measure of distance between points in $\mathcal{X}$. Then the initial distance is $\Delta(x_0, x_0 + \delta x)$, and the distance after n steps is $\Delta(T^n x_0, T^n(x_0 + \delta x))$. We are interested in the quantity

$$\lim_{\substack{n\to\infty \\ \delta x\to 0}} \frac{1}{n} \log\left(\frac{\Delta(T^n x_0, T^n(x_0 + \delta x))}{\Delta(x_0, x_0 + \delta x)}\right).$$

Lyapunov exponents are an important topic in the theory of differential equations. Say we have a differentiable function $f : \mathbb{R} \to \mathbb{R}$, look at the iteration

$$x_{n+1} = f(x_n).$$

Pick two points x_0 and $y_0 = x_0 + \epsilon$ so that y_0 is close to x_0, that is $\big|\epsilon\big|$ is small. The distance between the points initially, at $n = 0$, is $\big|\epsilon\big|$.For the first iteration we get

$$x_1 = f(x_0) \quad \text{and} \quad y_1 = f(y_0) = f(x_0 + \epsilon).$$

The distance at $n = 1$ is

$$\big|f(x_0 + \epsilon) - f(x_0)\big|.$$

The increase in logarithm distance between 0 and 1 is therefore

$$\log(\big|f(x_0 + \epsilon) - f(x_0)\big|) - \log(\epsilon)$$

We want study the different for arbitrary close points, so letting ϵ approach zero this distance becomes:

$$\begin{aligned}\lim_{\epsilon\to 0} \log(\big|f(x_0 + \epsilon) - f(x_0)\big|) - \log(\big|\epsilon\big|) &= \log\left(\lim_{\epsilon\to 0}\left|\frac{f(x_0 + \epsilon) - f(x_0)}{\epsilon}\right|\right) \\ &= \log(\big|f'(x_0)\big|).\end{aligned}$$

Then the next iteration we do it the same way:

$$x_2 = f(x_1) = f^{(2)}(x_0) \quad \text{and} \quad y_2 = f^{(2)}(x_0 + \epsilon).$$

Here is $f^{(2)} = f \circ f$. Now as ϵ approaches zero the logarithm distance becomes

$$\lim_{\epsilon\to 0} \log(\big|f^{(2)}(x_0 + \epsilon) - f^{(2)}(x_0)\big|) - \log(\big|\epsilon\big|) = \log(\big|f^{(2)'}(x_0)\big|).$$

Using induction we can show that the logarithm distance after n steps is

$$\log(\big|f^{(n)'}(x_0)\big|)$$

where $f^{(n)}$ is defined by $f^{(n)} = f^{(n-1)} \circ f$. The Lyapunov exponent is the limit of average logarithm distance of close points, so for this system it becomes

$$\lambda(f, x_0) = \lim_{n\to\infty} \frac{1}{n} \log(\big|f^{(n)'}(x_0)\big|).$$

Often the calculation of $f^{(n)'}(x_0)$ is complicated, but using the chain rule and induction this becomes

$$f^{(n)'}(x_0) = f'(x_0)f'(x_1)\cdots f'(x_{n-1}).$$

Thus, the Lyapunov exponent can be written as

$$\begin{aligned}\lambda(f, x_0) &= \lim_{n\to\infty} \frac{1}{n} \log(\Pi_{i=0}^{n-1}\big|f'(x_i)\big|) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \log(\big|f'(x_i)\big|).\end{aligned}$$

What information does the Lyapunov exponent give us about the system? It is a quantitative number that measure the dependence on initial conditions. It measures the exponential rate at which error grows. Note that the sum of $\log(\left|f'(x_i)\right|)$ over an orbit can in many situations be calculated by Birkhoff's theorem. Indeed,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\log(\left|f'(x_i)\right|).$$

is a Birkhoff sum under the right conditions.

Example 3.2. Consider the logistic map given by

$$f(x) = 4x(1-x)$$

where f is limited to the unit interval, $f : [0,1] \to [0,1]$. The Lyapunov exponent for this map is $\log(2)$ [24], which can be calculated using Birkhoff's theorem. Indeed, the the logistic map is a measure-preserving transformation under the measure with probability density

$$\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}.$$

see [25]. Thus,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n}\log(\left|f'(x_i)\right|) = \int_0^1 \frac{\log(\left|4(1-2x)\right|)}{\pi\sqrt{x(1-x)}} = \log(2).$$

That means that if the initial error is ϵ the error after n steps is $2^n\epsilon$.

As we saw by the previous example, Birkhoff's theorem can be useful to calculated the Lyapunov exponent when we are in $\mathbb{R}$. In higher dimensions however this theorem is often insufficient. Instead, Kingman's theorem is useful. In higher dimension we have to consider the Jacobian determinant matrix, and as we saw in example 2.2 matrices can be viewed as a subadditive function.

Definition 3.1. The maximal Lyapunov exponent of a metric space $(\mathcal{X}, d)$ with a continuous transformation $T : \mathcal{X} \to \mathcal{X}$ is

$$\lambda = \lim_{n\to\infty}\frac{1}{n}\limsup_{d(x,y)\to 0}\log\frac{d(T^n x, T^n y)}{d(x,y)}.$$

Remark 3.1. Let us try to use Fekete's Subadditive Lemma to show that this limit exists. Let

$$\alpha_n = \limsup_{d(x,y)\to 0}\log\frac{d(T^n x, T^n y)}{d(x,y)}.$$

We want to show that the sequence $\{\alpha_n\}$ is a subadditive sequence. Consider

$$a_{n+m} = \limsup_{d(x,y)\to 0} \log \frac{d(T^{n+m}x, T^{n+m}y)}{d(x,y)}$$

We can rewrite the argument of the log-function.

$$\frac{d(T^{n+m}x, T^{n+m}y)}{d(x,y)} = \frac{d(T^{n+m}x, T^{n+m}y)}{d(T^n x, T^n y)} \frac{d(T^n x, T^n y)}{d(x,y)}.$$

Thus,

$$\begin{aligned} a_{n+m} &\leq \limsup_{d(x,y)\to 0} \log \frac{d(T^{n+m}x, T^{n+m}y)}{d(T^n x, T^n y)} + \limsup_{d(x,y)\to 0} \log \frac{d(T^n x, T^n y)}{d(x,y)} \\ &\leq a_m + a_n. \end{aligned}$$

For the first term we have used that T is a continuous function. That is, letting $x \mapsto T^n x$ and $y \mapsto T^n y$ we get that $d(x,y) \to 0$ implies that $d(T^n x, T^n y) \to 0$ since T is continuous. We therefore get by Fekete's Subadditive Lemma that

$$\lambda = \inf_n \frac{\alpha_n}{n}.$$

Example 3.3. Let us turn back to the classical example of the torus translation to illustrate the latest definition. Here $Tx = x + \alpha \mod (1)$, thus $T^n x = x + n\alpha \mod (1)$ and we get that $d(T^n x, T^n y) = d(x,y)$ and the maximal Lyapunov exponent of the system is 0.

Example 3.4. Now let us look at a nontrivial example, the doubling map. Here $Tx = 2x \mod (1)$, thus $T^n x \leq 2^n x \mod (1)$ and we get that $d(T^n x, T^n y) \leq 2^n d(x,y)$. We want to show that $d(T^n x, T^n y) = 2^n d(x,y) \mod (0.5)$. The metric here is defined by

$$d(x,y) = \begin{cases} \big|x-y\big| & \text{if } \big|x-y\big| \leq 0.5 \\ 1 - \big|x-y\big| & \text{if } \big|x-y\big| \geq 0.5 \end{cases}$$

This metric can also be written as $d(x,y) = \min\{\big|x-y\big|, 1-\big|x-y\big|\}$. Thus, we get $d(2x,2y) = \min\{2\big|x-y\big|, 1-2\big|x-y\big|\} \leq 2d(x,y)$. We therefore have to find a pair x, y such that $d(T^n x, T^n y) = 2^n d(x,y)$ for a given n. If we pick x, y such that $x \leq \frac{1}{2^n}$ and $y \leq \frac{1}{2^n}$ then we have that $d(T^n x, T^n y) = 2^n d(x,y)$. Which gives a Lyapunov exponent equal to $\log(2)$.

3.2. **Linear cocycles.** Before we can move on, we need to define the concept of linear cocycles. This is an example of a MPDS, which will be useful for us when studying matrices over measure spaces.

Definition 3.2. GL_d denotes the general linear group of degree d, which is the set of $d \times d$ invertible matrices.

Definition 3.3. Given a pair (f, A) of measurable maps on a measure space $(\mathcal{X}, \mathcal{B}, \mu)$ such that $f : \mathcal{X} \to \mathcal{X}$ and $A : \mathcal{X} \to GL_d$. The *linear cocycle* defined by A over f is the transformation

$$F : \mathcal{X} \times \mathbb{R}^d \to \mathcal{X} \times \mathbb{R}^d, \quad (x, v) \to (f(x), A(x)v).$$

Note that

$$F^n(x, v) = (f^n(x), A^{(n)}(x)v)$$

where

$$A^{(n)}(x) = A(f^{n-1}(x))A(f^{n-2}(x)) \cdots A(f(x))A(x).$$

One way to view a cocycle is as a new dynamical system, the cross product $\mathcal{X} \times \mathbb{R}^d$ with the map F.

Example 3.5. Linear cocycles is an easy way to represent products of random matrices. Say we have a set of matrices $\{B_1, B_2, \cdots, B_k\}$ of degree d with associated probability vector $(p_1, p_2, \cdots, p_k)$. Let $\mathcal{X} = \{B_1, B_2, \cdots, B_k\}$ and define $(\mathcal{X}, \mathcal{B}, \mu)$ to be the discrete measure space such that $\mu(B_i) = p_i$ for $1 \leq i \leq k$. Let $\mathcal{Y} = \mathcal{X}^{\mathbb{Z}}$ be the product space over $\mathcal{X}$. Define

$$f : \mathcal{Y} \to \mathcal{Y} \quad f(y_1, y_2, y_3, \cdots) = (y_2, y_3, \cdots)$$
$$A : \mathcal{Y} \to GL_d \quad A(y_1, y_2, \cdots) = y_1$$

Then (f, A) is a linear cocycle over the product space $\mathcal{Y}$. Thus, we get that

$$\Pi_n = B_{i_1} B_{i_2} \cdots B_{i_n} = A^{(n)}(y)$$

for some $y \in \mathcal{Y}$.

Remark 3.2. Why do we study such systems? Consider the initial value problem

$$g'' + Vg = 0 \quad g(0) = a, g'(0) = b.$$

where $g : [0, \infty) \to \mathbb{R}$ and $V(t) = V(t + 1)$ is a periodic function with period 1. Assume g_0 is the solution of the IVP with $a = 1, b = 0$ and g_1 is the solution of the IVP with $a = 0, b = 1$. Then

$$g(t) = ag_0(t) + bg_1(t)$$

is a solution of the original IVP. We can write the solution on matrix form

$$\begin{bmatrix} g(1) \\ g'(1) \end{bmatrix} = \begin{bmatrix} ag_0(1) + bg_1(1) \\ ag_0'(1) + bg_1'(1) \end{bmatrix} = \begin{bmatrix} g_0(1) & g_1(1) \\ g_0'(1) & g_1'(1) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

and define the matrix

$$A_V = \begin{bmatrix} g_0(1) & g_1(1) \\ g_0'(1) & g_1'(1) \end{bmatrix}.$$

We call A_V the solution matrix of the IVP. That is, we can view the solution of the IVP as a matrix transformation:

$$\begin{bmatrix} g(1) \\ g'(1) \end{bmatrix} = A_V \begin{bmatrix} a \\ b \end{bmatrix}.$$

After n steps the solution is therefore

$$\begin{bmatrix} g(n) \\ g'(n) \end{bmatrix} = A_V^n \begin{bmatrix} a \\ b \end{bmatrix}.$$

Also note that A is completely determined by the function V. We can view this system as a cocycle. Indeed, this is a cocycle as in example 3.5 where $\mathcal{X}$ is the one point set containing the solution matrix A_V.

Example 3.6. Let us consider an easy example to illustrate this. Say that $V = c^2$ for some $c \in \mathbb{R}$. Then we have the solutions $g_0(t) = \cos(ct)$ and $g_1(t) = \frac{1}{c}\sin(ct)$ and we get the solution matrix

$$A_V = \begin{bmatrix} \cos(c) & \frac{1}{c}\sin(c) \\ -c\sin(c) & \cos(c) \end{bmatrix}.$$

A_V is diagonalizable. Indeed, the eigenvalues are

$$\lambda_{0,1} = \cos(c) \pm i\big|\sin(c)\big|.$$

This becomes

$$\lambda_{0,1} = e^{\pm ic}.$$

The eigenvectors are

$$v_0 = \begin{bmatrix} 1 \\ ic \end{bmatrix} \quad v_1 = \begin{bmatrix} 1 \\ -ic \end{bmatrix}.$$

The matrix A_v can therefore be written as

$$\begin{aligned} A_V &= PDP^{-1} \\ &= \frac{1}{2}\begin{bmatrix} 1 & 1 \\ ic & -ic \end{bmatrix}\begin{bmatrix} e^{ic} & 0 \\ 0 & e^{-ic} \end{bmatrix}\begin{bmatrix} 1 & \frac{-i}{c} \\ 1 & \frac{i}{c} \end{bmatrix}. \end{aligned}$$

We can use this to calculate $A^{(n)}$ after n steps:

$$\begin{aligned} A_V^{(n)} &= PD^nP^{-1}. \\ &= \frac{1}{2}\begin{bmatrix} 1 & 1 \\ ic & -ic \end{bmatrix}\begin{bmatrix} e^{nic} & 0 \\ 0 & e^{-nic} \end{bmatrix}\begin{bmatrix} 1 & \frac{-i}{c} \\ 1 & \frac{i}{c} \end{bmatrix} \\ &= \begin{bmatrix} \cos(nc) & \frac{1}{c}\sin(nc) \\ -c\sin(nc) & \cos(nc) \end{bmatrix}. \end{aligned}$$

Remark 3.3. In physical applications of this equation, the function $V(t)$ is not periodic, but consists of several different periodic functions. That is, we can have several different V_i based on a probably distribution.

$$V(t) = \begin{cases} V_1(t) & \text{with probability } p_1 \\ V_2(t) & \text{with probability } p_1 \\ \vdots & \\ V_n(t) & \text{with probability } p_n \end{cases}$$

where each V_i is a periodic function. It is important to note that $V(t)$ is restricted to one function V_i on each interval, that is

$$V(t)|_{[k,k+1]} = V_i(t) \quad \text{for some } i.$$

For example in quantum mechanics the potential can vary with time. Since we have a probability distribution we can not exactly say what happens after a certain time, so we try to figure out what happens on average. We get the system

$$\begin{bmatrix} f(0) \\ f'(0) \end{bmatrix} \to_{A_0} \begin{bmatrix} f(1) \\ f'(1) \end{bmatrix} \to_{A_1} \begin{bmatrix} f(2) \\ f'(2) \end{bmatrix} \to \cdots$$

Where each A_i is the solution matrix for the corresponding V_i for that IVP. The solution after N steps is

$$A^{(n)}\begin{bmatrix} a \\ b \end{bmatrix} = A_N \cdots A_1 A_0 \begin{bmatrix} a \\ b \end{bmatrix}.$$

So it is very natural to look at this as a cocycle similar to the one in example 3.5. Let $\mathcal{X}$ the set of all the solution matrices corresponding to all V_i's. Let $(\mathcal{X}, \mathcal{B}, \mu)$ be the discrete measure space such that $\mu(A_i) = p_i$. Then as before let $\mathcal{Y} = \mathcal{X}^{\mathbb{Z}}$ be the product space over $\mathcal{X}$ and define

$$\begin{aligned} &f : \mathcal{Y} \to \mathcal{Y} \quad f(y_1, y_2, y_3, \cdots) = (y_2, y_3, \cdots) \\ &A : \mathcal{Y} \to GL_d \quad A(y_1, y_2, \cdots) = y_1 \end{aligned}$$

Then (f, A) is a cocycle over $\mathcal{Y}$.

Example 3.7. Let us expand our previous example a bit. Given two functions $V_0 = c^2$ and $V_1 = -c^2$ for a $c \in \mathbb{R}$ with corresponding probabilities p_0 and p_1. Then as before for V_0 we get the matrix

$$A_{V_0} = \begin{bmatrix} \cos(c) & \frac{1}{c}\sin(c) \\ -c\sin(c) & \cos(c) \end{bmatrix}$$

and for V_1 we get the matrix

$$A_{V_1} = \begin{bmatrix} \cosh(c) & \frac{1}{c}\sinh(c) \\ c\sinh(c) & \cosh(c) \end{bmatrix}.$$

3.3. **Lyapunov exponent for linear cocycles.**

Definition 3.4. For a linear cocycle (f, A) on a dynamical system $(\mathcal{X}, \mathcal{B}, \mu)$ we define the Lyapunov exponent to be

$$\lambda(x) = \lim_{n\to\infty} \frac{1}{n} \log\|A^{(n)}(x)\|.$$

Remark 3.4. Let us turn back to the IVP to see why we are interested in this quantity. Remember that the Lyapunov exponent measures the dependence on the initial conditions. Let

$$v_0 = \begin{bmatrix} f(0) \\ f'(0) \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

and consider a small distribution from this

$$v_0 + \epsilon z$$

for a $\epsilon > 0$ and a vector z such that $\|z\| = 1$. Say we have a sequence of functions $\{V_i\}$ which corresponds to a point $x \in \mathcal{X}$. Then the initial distance is

$$\|v_0 - (v_0 + \epsilon z)\| = \epsilon\|z\| = \epsilon.$$

After n steps the distance becomes

$$\begin{aligned} \|A^{(n)}(x)v_0 - A^{(n)}(x)(v_0 + \epsilon z)\| &= \|A^{(n)}(x)\epsilon z\| \\ &\leq \epsilon\|A^{(n)}(x)\|\|z\| = \epsilon\|A^{(n)}(x)\| \end{aligned}$$

Thus, we have that

$$\frac{\|A^{(n)}(x)\epsilon z\|}{\epsilon\|z\|} \leq \|A^{(n)}(x)\|$$

Suprimum over all vectors with norm 1 leads to equality.

Example 3.8. Look at the matrix from example 3.6:

$$A_V = \begin{bmatrix} \cos(c) & \frac{1}{c}\sin(c) \\ -c\sin(c) & \cos(c) \end{bmatrix}.$$

There we saw that

$$A_V^{(n)} = \begin{bmatrix} \cos(nc) & \frac{1}{c}\sin(nc) \\ -c\sin(nc) & \cos(nc) \end{bmatrix}.$$

Thus, the norm is bounded below by 1 and above by $\left|c\right|$ or $\left|\frac{1}{c}\right|$. Therefore, the Lyapunov exponent is equal to 0.

Example 3.9. Consider the matrices from example 3.1

$$A_0 = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma^{-1} \end{pmatrix} \quad A_1 = \begin{pmatrix} \sigma^{-1} & 0 \\ 0 & \sigma \end{pmatrix}, \quad \text{where } \sigma > 1.$$

with probability p_0 and p_1 respectively. By symmetry we get that

$$\log\|A^{(n)}\| = \begin{cases} \left|m_0 - m_1\right| \log\sigma & \text{if } m_0 \neq m_1 \\ 0 & \text{if } m_0 = m_1 \end{cases}$$

Since

$$\lim_{n\to\infty} \frac{\left|m_0 - m_1\right|}{n} = \left|p_0 - p_1\right|$$

the Lyapunov exponent of the system becomes

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \log\|A^{(n)}\| = \left|p_1 - p_2\right| \log\left|\sigma\right|.$$

Example 3.10. Consider the matrices

$$A_0 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad A_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

with $p_0 = p_1 = 1/2$. Then

$$A^{(n)} = \begin{pmatrix} 2^{k_1} & 0 \\ 0 & 2^{k_2} \end{pmatrix} \quad \text{or} \quad A^{(n)} = \begin{pmatrix} 0 & 2^{k_1} \\ 2^{k_2} & 0 \end{pmatrix}$$

for some numbers $0 \leq k_1, k_2 \leq n$. We claim that

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \log\|A^{(n)}\| = \frac{\log(2)}{4}.$$

Proof. Denote

$$a_n = \mathbb{E}(\log(\|A_n A_{n-1} \cdots A_1\|).$$

We want to find

$$\lim_{n\to\infty} \frac{a_n}{n}.$$

Let us try to find $a_n - a_{n-1}$, this can be written as

$$\frac{1}{2}\log 2 \cdot \mathbb{P}\left(A^{(n-1)} = \begin{pmatrix} 2^k & 0 \\ 0 & 2^l \end{pmatrix} \text{ or } \begin{pmatrix} 0 & 2^k \\ 2^l & 0 \end{pmatrix} \text{ and } k \geq l\right)$$

since an increase in norm only happens if the n'th matrix is equal to A_0, $A^{(n-1)}$ is on the given form and $k \geq l$. We can split this equation in two parts, one where $k > l$ and one where $k = l$

$$= \frac{1}{4}\log 2 + \frac{1}{2}\log 2 \cdot \mathbb{P}\left(A^{(n-1)} = \begin{pmatrix} 2^k & 0 \\ 0 & 2^k \end{pmatrix} \text{ or } \begin{pmatrix} 0 & 2^k \\ 2^k & 0 \end{pmatrix}\right)$$

We have used that

$$\mathbb{P}\left(A^{(n)} = \begin{pmatrix} 2^k & 0 \\ 0 & 2^l \end{pmatrix} \text{ or } \begin{pmatrix} 0 & 2^k \\ 2^l & 0 \end{pmatrix} \text{ and } k > l\right) = \frac{1}{2}$$

which can be shown by induction. Now let

$$p_n = \mathbb{P}\left(A^{(n)} = \begin{pmatrix} 2^k & 0 \\ 0 & 2^k \end{pmatrix} \text{ or } \begin{pmatrix} 0 & 2^k \\ 2^k & 0 \end{pmatrix}\right).$$

Thus,

$$a_n = a_{n-1} + \frac{1}{4}\log 2 + \frac{1}{2}\log 2 \cdot p_{n-1}$$

which results in

$$a_n = \frac{n}{4}\log 2 + \frac{1}{2}\log 2 \cdot \frac{p_0 + \cdots p_{n-1}}{n}.$$

By the last equation, if we can show that

$$p_n \to 0 \text{ as } n \to \infty$$

we are done. We know that $A^{(n)}$ is on the form

$$\begin{pmatrix} 2^k & 0 \\ 0 & 2^l \end{pmatrix} \text{ or } \begin{pmatrix} 0 & 2^k \\ 2^l & 0 \end{pmatrix}.$$

Now consider all 2^n ways we can combine A_0 and A_0, and let $d(i, n)$ be the total number of resulting matrices after n steps with $i = k - l$. We see that $-n \leq i \leq n$. Note that $p_n = \frac{1}{2^n} d(0, n)$. At $n = 1$ we have either

$$\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Thus, $d(-1, 1) = 0, d(0, 1) = 1$ and $d(1, 1) = 1$. The next values for $d(i, n)$ are shown in the table below

i:	-3	-2	-1	0	1	2	3	4
$n = 1$:				1	1			
$n = 2$:			1	1	1	1		
$n = 3$:		1	1	2	2	1	1	
$n = 4$:	1	1	3	3	3	3	1	1

Note that this is similar to Pascal's triangle, just that every number is repeated twice. Actually, this is true, $d(i,n)$ follows the rule:

$$d(i,n) = d(i-1,n-1) + d(-1,n-1)$$

and one can show by induction that $d(i,n)$ is equal to the entries in Pascal's triangle. Thus,

$$p_n = \frac{1}{2^n} d(0,n) = \frac{1}{2^n} \cdot \begin{cases} \binom{n}{n/2} & \text{if } n \text{ is even} \\ \binom{n}{(n+1)/2} & \text{if } n \text{ is odd.} \end{cases}$$

Say n is even for simplicity, we get that this is

$$p_n = \frac{1}{2^n}\binom{n}{n/2} = \frac{1}{2^n}\frac{n!}{((n/2)!)^2}.$$

By Stirling's approximation we get

$$p_n \approx \frac{1}{2^n}\frac{\sqrt{2\pi n}(n/e)^n}{(\sqrt{\pi n}(n/2e)^{\frac{n}{2}})^2} = \sqrt{\frac{2}{n}} \to 0$$

as $n \to \infty$. We have shown that

$$\frac{a_n}{n} \to \frac{\log 2}{4} \text{ as } n \to \infty$$

which proves our claim. □

3.4. **Furstenberg and Kesten's theorem.** Furstenberg and Kesten first discovered their theorem in 1960 [4], eight years prior to Kingman's theorem. The theorem proves convergence for random product of matrices. Birkhoff's ergodic theorem is often viewed as a generalization of the law of large numbers. Similarly, can one view Kingman's theorem as a generalization of Furstenberg-Kesten's theorem.

Let $F : \mathcal{X} \times R^d \to \mathcal{X} \times R^d$ be given by $F(x,v) = (f(x), A(x)v)$ for a measurable function $A : \mathcal{X} \to GL_d$. Let $L_1(\mu)$ denote the space of μ-integrable functions on $\mathcal{X}$.

Theorem 3.11. *If* $\log\|A\| \in L_1(\mu)$ *then*

$$\lambda(x) = \lim_{n\to\infty} \frac{1}{n} \log\|A^{(n)}(x)\|$$

exists μ *a.e.*$x \in \mathcal{X}$. *Moreover,* $\lambda(x)$ *is an invariant function and* μ*-integrable.*

Proof. This follows directly from Theorem 2.5. Define

$$\phi_n(x) = \log\|A^{(n)}(x)\|.$$

We showed in example 2.2 that this is a subadditive function satisfying equation (2.1) and by the assumptions we know that $\phi_1(x) \in L_1(\mu)$. Thus the assumptions of Theorem 2.5 are satisfied and the conclusions of the theorem follows. □

This theorem and example 3.5 shows that

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \log\|\Pi_n\|$$

exists almost surely.

The study of Lyapunov exponents originated in the late 19th century from the fundamental work of the Russian mathematician Aleksandr Lyapunov [26] on the stability of solutions of differential equations. As discussed in the remarks throughout the chapter, one consider linear equations of the type

$$v'(t) = B(t)v(t)$$

where $B(t)$ is a bounded function from $\mathbb{R}$ to the set of $d \times d$ matrices. In the 1960-80, work of several mathematicians such as Furstenberg, Kesten and Kingman made the study of Lyapunov exponents in to a very active research field, with strong ties to other areas of mathematics and physics. As seen, even in basic systems computing Lyapunov exponents can be quite complicated. As a result, several numerical methods have emerged and has become a research field by its own [27].

4. Longest increasing subequence of a permutation

We will now study a different application of Kingman's theorem, the concept of longest increasing subsequence of a random permutation. Random permutation has been studied immensely in the last half century, and spans several branches of mathematics. It is also a classical computer science problem, where it is an example of a dynamical programming problem and has applications in other areas such as physics, random matrix theory and bioinformatics [28].

Let S_n be the group of permutations of order n. An increasing subsequence of a permutation $\pi : i \to \pi(i)$ is a subsequence such that

$$i_1 < i_2 < \cdots < i_k; \quad \pi(i_1) < \pi(i_2) < \cdots < \pi(i_k).$$

Similarly, a decreasing subsequence is when

$$i_1 < i_2 < \cdots < i_k; \quad \pi(i_1) > \pi(i_2) > \cdots > \pi(i_k).$$

Example 4.1. Look at the permutation π

$$2 \quad 6 \quad 8 \quad 4 \quad 1 \quad 5 \quad 7 \quad 3.$$

This permutation has several increasing subsequences, but only one of length 4:

$$2 \quad 4 \quad 5 \quad 7.$$

We call this the longest increasing subsequence of π. Similarly we have the longest decreasing subsequence of π, here we actually have three decreasing subsequences of length 3;

$$8 \quad 5 \quad 3, \quad 6 \quad 4 \quad 1 \quad \text{and} \quad 6 \quad 5 \quad 3.$$

We have a more precise definition.

Definition 4.1. Let π be a permutation of order n. Then we let $L(\pi)$ to be the maximum length of an increasing subsequence.

$L(\pi) = \max\{1 \leq k \leq n : \pi \text{ has an increasing subsequence of length } k\}$.

Similarly, we define $D(\pi)$ to be the maximum length of a decreasing subsequence.

$D(\pi) = \max\{1 \leq k \leq n : \pi \text{ has a decreasing subsequence of length } k\}$.

Example 4.2. Consider the permutation in example 4.1. We have that $L(\pi) = 4$ and $D(\pi) = 3$.

Now define the sequence l_n

$$l_n = \frac{1}{n!} \sum_{\pi \in S_n} L(\pi).$$

That is the average over $L(\pi)$ of all permutation of order n. Similarly define

$$d_n = \frac{1}{n!} \sum_{\pi \in S_n} D(\pi).$$

By symmetry note that $l_n = d_n$. Our goal is to use Kingman's theorem to find the asymptotic behaviour of l_n. It turns out that the expected value of the longest increasing sub-sequence in a random permutation is asymptotic to $\sqrt{n}$. The goal of this section will be to prove the following theorem.

Theorem 4.3.

$$\lim_{m \to \infty} \frac{l_n}{\sqrt{n}} \to c.$$

For a constant c, where $0 < c < \infty$.

4.1. **First bounds.** Before we get to the details, let us try to get some bounds on c.

Theorem 4.4 (Erdos-Szekeres theorem)**.** *If* $\pi \in S_n$ *and* $n > rs$ *for some integers* $r, s \in \mathbb{N}$*, then either* $L(\pi) > r$ *or* $D(\pi) > s$*. Also,* $L(\pi)D(\pi) \geq n$*.*

Proof. For each x in the permutation π defined the pair $(i, j)_x$ as follows: i is the length of the longest increasing sequence starting at x and j is the length of the longest decreasing sequence ending at x. We now have n pairs and we claim that they are all distinct. Indeed, assume that $(i, j)_x = (k, l)_y$ and without loss of generality that x comes before y in the permutation. If $x > y$ then there is a longer decreasing sequence ending at y than l, namely the decreasing sequence of length j ending at x with y added at the end. This sequence has length $l + 1$. Then if $x < y$, then there is a longer increasing sequence staring at x than i, namely the increasing sequence of length k starting at y with x added at the beginning. This sequence has length $i+1$. Hence, we have n distinct pairs of numbers. Now we must have that either $L(\pi) > r$ or $D(\pi) > s$ since if (i, j) are all in $\{1, 2, \cdots, r\} \times \{1, 2, \cdots, s\}$ then we have a contradiction of what we just proved. That is, since $n > rs$ there are not enough pairs in $\{1, 2, \cdots, r\} \times \{1, 2, \cdots, s\}$ to make n pairs distinct. In the same way we know that the pairs (i, j) are all in $\{1, 2, \cdots, L(\pi)\} \times \{1, 2, \cdots, D(\pi)\}$, thus since they are all distinct we get that $L(\pi)D(\pi) \geq n$. □

Example 4.5. Consider the special case when $n = N^2$ for some $N \in \mathbb{N}$. Then there exists a permutation π_n such that $L(\pi) = D(\pi) = N$.

Proof. The permutation we are seeking is a block permutation with N blocks of size N. The first block is the numbers from $N^2 - N + 1$ to N^2 in increasing order. Then the next block is the numbers from $N^2 - 2N + 1$ to $N^2 - N$ in increasing order. And so on until the last block, which is the numbers 1 to N in increasing order, see figure 5. Here are the permutations for $N = 2, 3$:

$$3 \quad 4 \quad 1 \quad 2$$
$$7 \quad 8 \quad 9 \quad 4 \quad 5 \quad 6 \quad 1 \quad 2 \quad 3$$

One can easy see that for both of these permutations both $L(\pi)$ and $D(\pi)$ are equal to 2 and 3 respectively. This is true for any N. Let us first see this for L. Any block would be an increasing subsequence of length N. But one can not combine different blocks to create a increasing subsequence since the blocks are decreasing. For D pick one element for each block and this will create a decreasing subsequence of length N. One can not have two or more elements from the same block as this would not create a decreasing subsequence. Thus, both the maximal increasing and decreasing subsequence are of length N. □

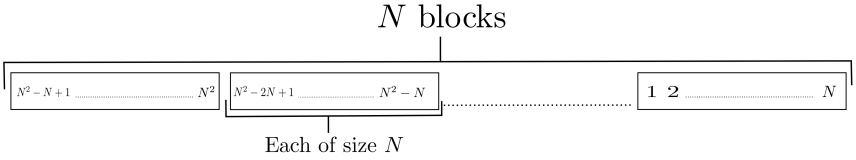

FIGURE 5. The block-permutation from example 4.5. The permutation is of order N^2 and has $L(\pi) = D(\pi) = N$.

We will use Erdos-Szekere's theorem to get a lower bound for l_n.

Proposition 4.1. *For all $n \geq 1$ we have*

$$l_n \geq \sqrt{n}.$$

Proof. Remember that l_n is defined as the average value of $L(\pi)$ and by symmetry it is also the average value of $D(\pi)$. Thus, we get that

$$l_n = \frac{1}{n} \sum_{\pi \in S_n} \frac{L(\pi) + D(\pi)}{2} = \mathbb{E}\left(\frac{L(\pi) + D(\pi)}{2} \right). \tag{4.1}$$

The inequality of arithmetic and geometric means states that

$$\frac{x + y}{2} \geq \sqrt{xy} \quad x, y \geq 0.$$

Using this on equation (4.1) and the fact that $L(\pi)D(\pi) \geq n$ we get that

$$\mathbb{E}\left(\frac{L(\pi)+D(\pi)}{2}\right) \geq \mathbb{E}(\sqrt{L(\pi)D(\pi)}) \geq \mathbb{E}(\sqrt{n}) \geq \sqrt{n}$$

Thus, we get that

$$l_n \geq \sqrt{n}.$$

□

Note that this is not an asymptotic bound, but it is true for all n. That is, the expected value of the longest increasing subsequence is larger than $\sqrt{n}$ for all n. From this can one asks what is a typical value for $L(\pi)$ and which values are rare to obtain. There is always a possibility to get 1 or n, but this happens only once for each n. Consider for example all permutations of order 5. Clearly we have one permutation with $L = 1$ and one with $L = 5$. In table 2 we have listed how many of them obtain each number from 1 to 5. This is also shown in figure 6, here have we plotted the histogram for $n = 7, 8, 9, 10$. We clearly see that the majority of permutations have similar L. This indicates that the variance can not be that large. We also calculated the exact value for l_n for different n. These numbers are given in table 1 where also the reference value $\sqrt{n}$ are listed. Those numbers suggest that $\sqrt{n}$ is a rather conservative lower bound, we will later see improved lower bounds.

TABLE 1. Exact value of l_n for $2 \leq n \leq 9$. As comparison $\sqrt{n}$ and $e\sqrt{n}$ are included to illustrate proposition 4.1 and proposition 4.2 respectively.

n	l_n	$\sqrt{n}$	$e\sqrt{n}$
2	1.5	1.41	3.64
3	2	1.73	4.71
4	2.42	2	5.44
5	2.79	2.23	6.08
6	3.14	2.45	6.66
7	3.47	2.65	7.29
8	3.77	2.83	7.69
9	4.06	3	8.15

TABLE 2. Number of permutations with $L(\pi) = i$, $1 \leq i \leq n$ for $n = 5$

i	$\#\pi$
1	1
2	41
3	61
4	16
5	1

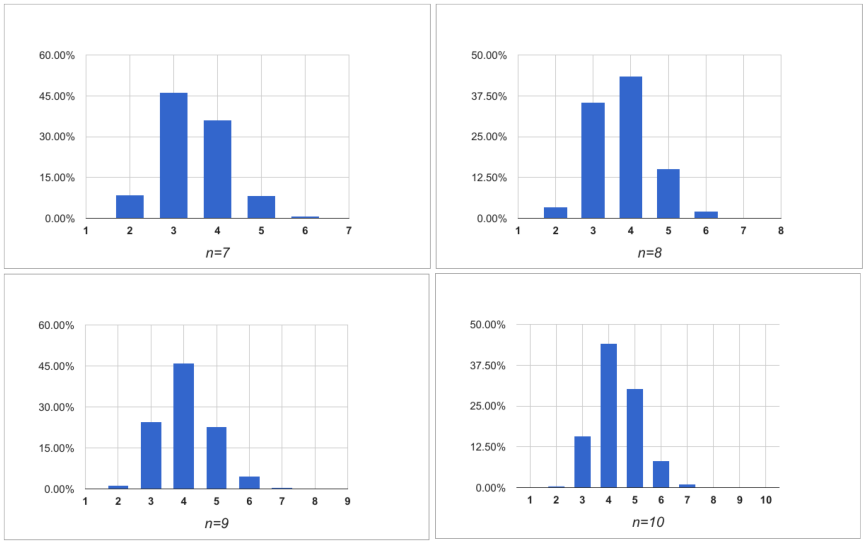

FIGURE 6. Histogram of $L(\pi)$ for $n = 7, 8, 9, 10$.

We have a lower bound for c. Next we will see an upper bound.

Proposition 4.2.

$$\limsup_{n\to\infty} \frac{l_n}{\sqrt{n}} \leq e$$

Proof. For $1 \leq k \leq n$ let $X_{k,n}$ be the number of increasing subsequences of length k in a random permutation of length n. We want to compute the expected value of $X_{k,n}$. Note that this is the sum, over all $\binom{n}{k}$ subsequences of length k, of the probability that the subseqeuence is increasing. This probability is $\frac{1}{k!}$. Thus,

$$\mathbb{E}X_{k,n} = \frac{1}{k!}\binom{n}{k}.$$

We want to use this to bound the probability that $L(\pi)$ is a least k.

$$P(L(\pi) \geq k) = P(X_{k,n} \geq 1) \leq \mathbb{E}X_{k,n} = \frac{1}{k!}\binom{n}{k}.$$

Now let us consider this number. First we can write it out:

$$\frac{1}{k!}\binom{n}{k} = \frac{1}{(k!)^2}\frac{n!}{(n-k)!}.$$

The last term can be bounded by n^k:

$$\frac{n!}{(n-k)!} \leq n^k.$$

For the first term we need Sterling's formula

$$n! \geq \left(\frac{n}{e}\right)^n.$$

Thus,

$$\frac{1}{(k!)^2} \leq \left(\frac{e}{k}\right)^{2k}.$$

Combining this we get that

$$\frac{1}{k!}\binom{n}{k} \leq n^k \left(\frac{e}{k}\right)^{2k}.$$

Now for a $\delta > 0$ let $k = \lceil (1+\delta)e\sqrt{n} \rceil$. Here $\lceil x \rceil$ denotes the *ceil-function*, which is the least integer greater than or equal to x. Then we get that

$$n^k \left(\frac{e}{k}\right)^{2k} \leq \left(\frac{1}{1+\delta}\right)^{2k} \leq \left(\frac{1}{1+\delta}\right)^{2(1+\delta)e\sqrt{n}} = e^{-\log(1+\delta)2(1+\delta)e\sqrt{n}}$$

Note that this bound converges to zero at exponential rate in $\sqrt{n}$ as $n \to \infty$. Thus,

$$P(L(\pi) \geq k) \leq e^{-c_\delta \sqrt{n}}$$

for some positive constant c_δ depending on δ. Since $L(\pi) \leq n$, we will now use that

$$l_n = \mathbb{E}(L(\pi)) \leq P(L(\pi) < k)k + P(l(\pi) \geq k)n.$$

So we get that

$$l_n \leq (1+\delta)e\sqrt{n} + e^{-c_\delta\sqrt{n}}n.$$

Taking limits we get that

$$\limsup_{n\to\infty} \frac{l_n}{\sqrt{n}} \leq (1+\delta)e.$$

Since δ was arbitrary we are done. $\square$

Proposition 4.1 gave us a lower bound of l_n for all n. To the contrary proposition 4.2 is an asymptotic bound. Which means that for some $n \in \mathbb{N}$ it is possible that $l_n \geq e\sqrt{n}$. Table 1 shows that this is not the case up till $n = 9$. Although we can not say for certain what the upper bound of l_n is for a given n, the proof of the proposition gives us some clues. If we look at the proof of the proposition we see that $(1+\delta)e\sqrt{n}$ is not just a bound for the average of $L(\sigma_n)$ but also for the typical value. Namely the value that is obtained with a probability close to 1 for large n. We will need this later, so we have a corollary from the previous proposition.

Corollary 4.3. *For any $\alpha > e$ we have that for all n that*

$$P(L(\pi) > \alpha\sqrt{n}) \leq Ce^{-c\sqrt{n}}$$

for some constants $C, c > 0$ that are depended on α but not on n.

4.2. **Hammersley process.** As mentioned at the beginning of the section, we want to use Kingman's theorem to prove the asymptotic behaviour of l_n. We actually need Ligget's version since it is more general and covers this specific application. In order to do this, we need to reformulate our problem such that the theorem is indeed applicable. The idea is to view the problem more in a geometric way. We want to express a permutation as points on the plane and apply spatial probability theory. Analyzing these points in the plane as a Poisson point process is known as a Hammersley process, and was introduced by the mathematician while studying this problem. We therefore need a way to express a random permutation of order n by n points in the plane and vise versa. We define a permutation based on the set of points by: The point with the i'th smallest y-coordinate has the $\pi(i)$'th smallest x-coordinate. Reverse, a permutation π results in the set of points $(i, \pi(i))_{i=1}^n$.

Example 4.6. Let us go back to the permutation in example 4.1 and consider π:

$$2 \quad 6 \quad 8 \quad 4 \quad 1 \quad 5 \quad 7 \quad 3.$$

See figure 7 for a plot of this permutation.

The following arguments are based on the proof of Theorem 1.6 in [29]

We have a way to represent a permutation as a geometric object, but how does our problem translate with this representation? We are looking for an increasing subsequecne, that means we have an ordering on the numbers in order to compare them to each other. We need this in $\mathbb{R}^2$ as well. Let $\preceq$ be a partial order on $\mathbb{R}^2$ where $(x_1, y_1) \preceq (x_2, y_2)$

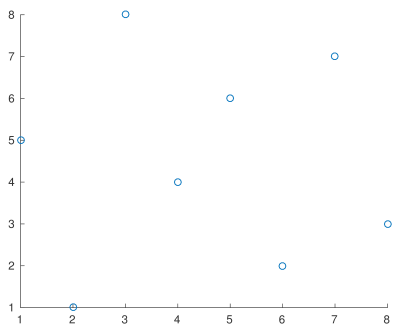

FIGURE 7. The permutation from example 4.6 represented as points in the plane.

if $x_1 \leq x_2$ and $y_1 \leq y_2$. Let A be a set of points in $\mathbb{R}^2$. An increasing subset of A is a subset where every two points in A are comparable with respect to $\preceq$. Let $L(A)$ be the length of the longest increasing subset of A.

Since l_n is the expected value of a random permutation, we need to make sure that the points chosen and the permutations they represent are uniformly distributed. Therefore an algorithm of generating this points is needed to ensure that they are indeed equally distributed. Let $X_1, \cdots, X_n$ be a sequence of independent random variables distributed on the uniform distribution $U[0,1]$. We want to define a permutation σ_n according to these random variables: define $\sigma_n(j)$ for each $1 \leq j \leq n$ to be the number k such that X_j is the kth smallest among $X_1, \cdots, X_n$. This is well defined since $X_1, \cdots, X_n$ takes distinct values with probability 1. Given a permutation we can easily go back to a set of points by letting $X_i = \pi(i)$ and $Y_i = i$.

Example 4.7. Let $n = 4$ and pick four random numbers,

$$X_1 = 0.5847, \quad X_2 = 0.9481, \quad X_3 = 0.0610, \quad X_4 = 0.5846.$$

Sorting them gives us

$$X_3 < X_4 < X_1 < X_2.$$

Then we have by construction that

$$\sigma(1) = 3$$

since X_1 is the 3rd smallest number. And so forth we get that

$$\sigma(2) = 4, \quad \sigma(3) = 1, \quad \sigma(4) = 2.$$

Thus, the resulting permutation is therefor

$$3 \quad 4 \quad 1 \quad 2.$$

Proposition 4.4. *The permutation σ_n defined in the preceding paragraph is uniformly distributed in S_n.*

Proof. Let σ_n be a permutation, we want to show that the probability to obtain this permutation by the described method is $1/n!$. Let $\sigma_n(1) = k_1$, then we must find the probability that X_1 is the k_1th smallest number among $X_1, \cdots, X_n$. That happens with a $1/n$ probability. Then $\sigma_n(2) = k_2 \neq k_1$, and the probability that X_2 is the k_2th smallest number among $X_1, \cdots, X_n$ given that X_1 is the k_1th smallest number is $1/(n-1)$. Thus, using induction we get that the probability to pick this exact permutation is $1/n!$. □

Now let $A_n(s,t)$ denote a set of n random points chosen independently and uniformly random from a rectangle $[0,s] \times [0,t]$ where $s, t > 0$. We want to show that the maximal increasing subset size $L(A_n(s,t))$ is a random variable with the same distribution as $L(\sigma_n)$. If we represent the set as $A_n(s,t) = \{(X_k, Y_k)\}_{k=1}^n$, where both $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ are independent random variables, with X_i distributed uniformly in $[0,s]$ and with Y_i distributed uniformly in $[0,t]$. Let π_n and ν_n be the resulting permutations from $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ respectively. Now let $\sigma_n = \pi_n \circ \nu_n^{-1}$. This is also a uniformly random permutation of order n.

Example 4.8. Continuing from example 4.7, let X_i be the random numbers given there and π the resulting permutation. Pick four new random numbers from $[0,1]$:

$$Y_1 = 0.2851, \quad Y_2 = 0.8277 \quad Y_3 = 0.1910 \quad Y_4 = 0.4425.$$

This results in a permutation ν

$$2 \quad 4 \quad 1 \quad 3.$$

Then we get that $\pi \circ \nu^{-1}$ is

$$1 \quad 3 \quad 2 \quad 4$$

which has an increasing subsequence of length 3. The set

$$A = \{(X_k, Y_k)\}_{k=1}^4$$

is plotted in figure 8. We see that the maximal increasing subset is indeed of size three.

With this example in mind we have motivation for our next proposition.

Proposition 4.5. $L(A_n(s,t)) = L(\sigma_n)$.

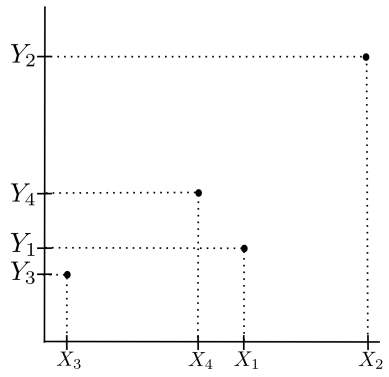

FIGURE 8. A plot of the random points from example 4.8.

Proof. First pick $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_3$ from the uniform distribution $U[0,1]$ and form the permutations π, ν and $\sigma = \pi \circ \nu^{-1}$ as described. For each $1 \leq j \leq n$ there exists a $1 \leq k \leq n$ such that

$$\nu_n^{-1}(k) = \pi_n^{-1}(j)$$

and

$$(X_{\pi_n^{-1}(j)}, Y_{\nu_n^{-1}(k)})$$

forms a point in the coordinate system. So an increasing subset of size l means that there are two increasing sequences of length l

$$j_1 < j_2 < \cdots < j_l, \quad 1 \leq j_i \leq n$$
$$k_1 < k_2 < \cdots < k_l, \quad 1 \leq k_i \leq n$$

such that

$$X_{\pi_n^{-1}(j_1)} < X_{\pi_n^{-1}(j_2)} < \cdots < X_{\pi_n^{-1}(j_l)}$$

and

$$Y_{\pi_n^{-1}(k_1)} < Y_{\pi_n^{-1}(k_2)} < \cdots < Y_{\pi_n^{-1}(k_l)}$$

with the property that

$$\nu_n^{-1}(k_i) = \pi_n^{-1}(j_i), \quad 1 \leq i \leq l.$$

But this is the same as

$$j_i = \sigma_n(k_i), \quad 1 \leq i \leq l.$$

But what do we have here, an increasing subsequence of length l. Indeed, for

$$k_1 < k_2 < \cdots < k_l : \sigma_n(k_1) < \sigma_n(k_2) < \cdots < \sigma_n(k_l).$$

We have showed that an increasing subset results in an increasing subsequence, so we have proved that $L(A_n(s,t) \leq L(\sigma_n)$. The reverse is more obvious, a permutation σ_n generates n points on the form

$(\sigma_n(i), i)$ for $1 \leq i \leq n$. Thus, if there is an increasing subsequence of length l:

$$i_1 < i_2 < \cdots < i_l : \quad \sigma_n(i_1) < \sigma_n(i_2) < \cdots < \sigma_n(i_l).$$

Thus, we have l points which forms an increasing subset:

$$(\sigma_n(i_1), i_1) \preceq (\sigma_n(i_l), i_l) \preceq \cdots \preceq (\sigma_n(i_l), i_l).$$

□

We have a way to represent our problem geometrically. The next proposition states that the distribution of these points are the same as a Poisson point process when conditioned on the number of points. This allows us to use the properties of a Poisson point process trying to find the asymptotic behaviour to l_n.

Let Π be a Poisson point process in the plane with unit intensity. Define $N(s,t) = \big|\Pi \cap ([0,s] \times [0,t])\big|$.

Proposition 4.6. *Conditioned on the event $N(s,t) = n$, the distribution of the set $\Pi \cap ([0,s] \times [0,t])$ is the same as the random set $A_n(s,t)$.*

For a proof see chapter 2 in [30]. Knowing this we now define a new random variable:

$$Y_{s,t} = L(\Pi \cap ([s,t) \times [s,t)))$$

We easily see that this is a superadditive sequence:

$$Y_{0,m} + Y_{m,n} \leq Y_{0,n}$$

since you can always find a increasing subsequence of length $Y_{0,m}+Y_{m,n}$ in the box $[0,n) \times [0,n)$ by combining the two increasing subsequences in the box $[0,m) \times [0,m)$ and $[m,n) \times [m,n)$ with length $Y_{0,m}$ and $Y_{m,n}$ respectively. This of course makes us think about Kingman's theorem.

Proposition 4.7. *The sequence $(-Y_{m,n})_{m.n}$ satisfies the conditions in theorem 2.11, including the ergodic property.*

Proof. We already saw that the sequence satisfies the subadditive property. We must show the three other properties. First, for each $k \geq 0$ the sequence

$$\{Y_{nk,(n+1)k}, n \geq 1\}$$

is i.i.d, thus by example 2.6 and 2.8 the sequence is both stationary and ergodic. Condition number (3) is straight forward. Since the number of points is equally distributed throughout the space,

$$\{Y_{m,m+k}, k \geq 1\}$$

clearly does not depend on m. We also have that

$$Y_{0,1} \leq N(s,t).$$

Thus,

$$\mathbb{E}\big|Y_{0,1}\big| \leq \mathbb{E}\big|N(s,t)\big| = 1.$$

We are left to show that for each n

$$\mathbb{E} - Y_{0,n} > c_0 n$$

for some constant $c_0 > -\infty$. This is the same as showing

$$\mathbb{E}Y_{0,n} < c_0 n.$$

We have

$$\begin{aligned}
\mathbb{E}Y_{0,n} &= \sum_k \mathbb{E}(Y_{0,n}|N(0,n)=k)P(N(0,n)=k) \\
&= \sum_k \mathbb{E}(L(A_k(0,n)))P(N(0,n)=k) \\
&= \sum_k \mathbb{E}(\sigma_k)P(N(0,n)=k)
\end{aligned}$$

We know from proposition 4.2 that $\limsup_{k\to\infty} \mathbb{E}(\sigma_k) \leq e\sqrt{k}$. We therefore know that for an $\epsilon > 0$ there exist a $M \in \mathbb{N}$ such that $\mathbb{E}(\sigma_k) \leq (e+\epsilon)\sqrt{k}$ for all $k > M$. We can rewrite the last expression for $\mathbb{E}Y_{0,n}$:

$$\begin{aligned}
\mathbb{E}Y_{0,n} &= \sum_k \mathbb{E}(\sigma_k)P(N(0,n)=k) \\
&= \sum_{k\leq M} \mathbb{E}(\sigma_k)P(N(0,n)=k) + \sum_{k>M} \mathbb{E}(\sigma_k)P(N(0,n)=k) \\
&\leq \sum_{k>M} (e+\epsilon)\sqrt{k}P(N(0,n)=k) \\
&\leq \sum_{k=0}^{\infty} (e+\epsilon)\sqrt{k}P(N(0,n)=k) = (e+\epsilon)\mathbb{E}\sqrt{Z_n}
\end{aligned}$$

Here Z_n denotes the random variable over how many points there are in the box $[0,n)\times[0,n)$. Since ϵ was arbitrary we get that $\mathbb{E}Y_{0,n} \leq e\mathbb{E}\sqrt{Z_n}$. Cauchy-Schwarz inequality says that $\mathbb{E}\sqrt{Z_n} \leq \sqrt{\mathbb{E}Z_n}$, and since clearly $\mathbb{E}Z_n = n^2$. we get that

$$\mathbb{E}Y_{0,n} < c'n.$$

□

Now that we know that $(-Y_{m,n})_{m.n}$ satisfies Ligget's version of Kingman's theorem, we get that

$$\frac{Y_{0,n}}{n} \to \gamma \text{ as } n \to \infty. \tag{4.2}$$

Remember that we are interested in $L(\sigma_n)$, we will now try to relate this limit to get what we desire. Now for each $n \geq 1$ define a random variable T_n by

$$T_n = \inf\{t > 0 : \left|\Pi \cap ([0,t] \times [0,t])\right| = n\},$$

and consider the set

$$\frac{1}{T_{n+1}}(\Pi \cap ([0, T_{n+1}) \times [0, T_{n+1})).$$

This is a random set of points the lives inside the unit square $[0,1] \times [0,1]$. By proposition 4.6 its joint distribution is that of n independent uniformly random points in $[0,1]^2$. In particular the random variable

$$Y_{0,T_{n+1}} = L(\Pi \cap ([0, T_{n+1}) \times [0, T_{n+1}))$$

is equal in distribution to $L(\sigma_n)$. We want to examine the asymptotic behaviour of T_n. To do that define a new sequence of random variables defined by

$$S_0 = 0 \text{ and } S_n = T_n^2 \text{ for } n \geq 1.$$

Then we get that

$$\begin{aligned} S_n &= \inf\{s > 0 : \left|\Pi \cap ([0, \sqrt{s}] \times [0, \sqrt{s}])\right| = n\} \\ &= \{s > 0 : M(s) = n\}, \end{aligned}$$

where $M(s) = \left|\Pi \cap ([0, \sqrt{s}] \times [0, \sqrt{s}])\right|$. We get from the definition of the Poisson point process that $(M(s))_{s \geq 0}$ is a one-dimensional Poisson process of unit intensity on $[0, \infty)$. From theorem 1.2 we have that $W_n = S_n - S_{n-1}$ are i.i.d exponential random variables with mean 1. Thus,

$$\frac{1}{n} S_n = \frac{1}{n} \sum_{k=1}^{n} W_k \to 1 \text{ as } n \to \infty$$

by the strong law of large numbers. This is equivalent to

$$\frac{1}{\sqrt{n}} T_n \to 1 \text{ almost surely as } n \to \infty.$$

Combining this with equation (4.2) we get that

$$\frac{Y_{0,T_{n+1}}}{\sqrt{n}} = \frac{T_{n+1}}{\sqrt{n}} \cdot \frac{Y_{0,T_{n+1}}}{T_{n+1}} \to \gamma \text{ almost surely as } n \to \infty.$$

We know that $Y_{0,T_{n+1}}$ has the same distribution as $L(\sigma_n)$, and since almost surly convergence implies convergence in probability we get that $L(\sigma_n)/\sqrt{n}$ converges to γ in probability. We discussed earlier what we can expect of $L(\pi)$, this limit gives us an indication of that. Not only is the expected value of $L(\pi)$, as we will see shortly, in the order of $\sqrt{n}$, but as n grows larger we can almost be certain that $L(\pi)$ is in the order of $\sqrt{n}$. In fact, freak occurrences as $L(\pi)$ being 1 or n will almost certainly never happen, which is logical.

We have now showed asymptotic behaviour for $L(\sigma_n)$,we can use this to prove theorem 4.3.

Proof of theorem 4.3. Let $\delta > 0$. We have that

$$\begin{aligned}
\big|\frac{l_n}{\sqrt{n}} - \gamma\big| &= \big|n^{-1/2}\mathbb{E}L(\sigma_n) - \gamma\big| \leq \mathbb{E}\big|n^{-1/2}L(\sigma_n) - \gamma\big| \\
&= \mathbb{E}\left(\big|n^{-1/2}L(\sigma_n) - \gamma\big|1_{\left|n^{-1/2}L(\sigma_n)-\gamma\right|\leq\delta}\right) \\
&+ \mathbb{E}\left(\big|n^{-1/2}L(\sigma_n) - \gamma\big|1_{\left|n^{-1/2}L(\sigma_n)-\gamma\right|>\delta, L(\sigma_n)\leq 3\sqrt{n}}\right) \\
&+ \mathbb{E}\left(\big|n^{-1/2}L(\sigma_n) - \gamma\big|1_{\left|n^{-1/2}L(\sigma_n)-\gamma\right|>\delta, L(\sigma_n)> 3\sqrt{n}}\right)
\end{aligned}$$

The first term is at most δ. The second one can be bounded by

$$(3+\gamma)P\left(\big|n^{-1/2}L(\sigma_n) - \gamma\big| > \delta\right).$$

For the last term we have that

$$\mathbb{E}\left(\big|n^{-1/2}L(\sigma_n) - \gamma\big|1_{\left|n^{-1/2}L(\sigma_n)-\gamma\right|>\delta, L(\sigma_n)> 3\sqrt{n}}\right) \leq C(\sqrt{n}+\gamma)e^{-c\sqrt{n}}$$

by corollary 4.3. Combining this we get that

$$\limsup_{n\to\infty}\big|\frac{l_n}{\sqrt{n}} - \gamma\big| \leq \delta.$$

This ends the proof since δ was arbitrary. □

4.3. **Improved bounds and further research.** We have successfully showed that l_n converges to $c\sqrt{n}$. We also showed that $1 \leq c \leq e$. Logan and Shepp (1977) [31] and Vershik and Kerov(1977) [32] showed independently that $c = 2$. We will not do this, but give another lower bound using Poisson point process.

Proposition 4.8. $c \geq (8/\pi)^{\frac{1}{2}} > 1.59$

Proof. Let Π be the Poisson point process in $\mathbb{R}^2$ with unit intensity. Choose a sequence (x_i, y_i) of points in Π as follows: (x_1, y_1) is the point of Π which minimize the value $x + y$ with $x \geq 0$ and $y \geq 0$. Then inductively pick points $(x_i, y_i) \in \Pi \cap (x_{i-1}, \infty) \times (y_{i-1}, \infty)$ which minimize the value $x+y$ in $(x_{i-1}, \infty) \times (y_{i-1}, \infty)$. The sequence (x_i, y_i) forms an increasing subset of Π. Let

$$t(n) = \max(x_n, y_n)$$

then $Y_{0,t(n)} \geq n$. Consider the difference $a_n = x_n - x_{n-1}$, they are independent and identically distributed. We want to calculate the mean of a_1. The probability that the first point of Π in the first quadrant is in A' is the probability that there are some points in A_1 minus the probability that there are some points in A_0, see figure 9.

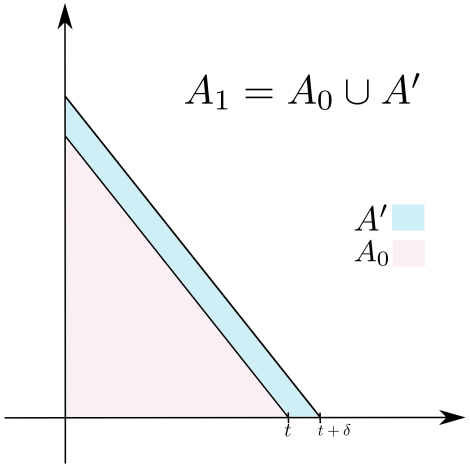

FIGURE 9. Calculation of the mean of a_1 from the proof of proposition 4.8. The probability that the first point of Π in the first quadrant is in A' is the probability that there are some points in A_1 minus the probability that there are some points in A_0.

Since we have a Poisson point process we can calculate the exact probabilities:

$$P(N(A_0) > 0) = 1 - e^{\frac{t^2}{2}}$$

$$P(N(A_1) > 0) = 1 - e^{\frac{-(t+\delta)^2}{2}}.$$

Thus,

$$P(\text{ first point of } \Pi \cap R^+ \text{ is in } A') = e^{-\frac{t^2}{2}} - e^{-\frac{(t+\delta)^2}{2}}.$$

We are interested in the probability density function:

$$\begin{aligned} f_d &= \lim_{\delta \to 0} \frac{P(\text{ first point of } \Pi \cap R^+ \text{ is in } A')}{\delta} \\ &= \lim_{\delta \to 0} \frac{e^{\frac{t^2}{2}} - e^{\frac{(t+\delta)^2}{2}}}{\delta} = \lim_{\delta \to 0} \frac{e^{\frac{t^2}{2}}(1 - e^{-t\delta - \delta^2/2})}{\delta} \\ &= \lim_{\delta \to 0} \frac{e^{\frac{t^2}{2}}(t\delta + (t\delta)^2/2 + \cdots)}{\delta} \to te^{-\frac{t^2}{2}}. \end{aligned}$$

Where we used the Taylor expansion of the exponential function. Thus, the expected value of a_1 is

$$\mathbb{E}a_1 = \mathbb{E}x_1 = \frac{1}{2}\mathbb{E}(x_1 + y_1) = \frac{1}{2}\int_0^\infty t(te^{-\frac{t^2}{2}})dt = \frac{\sqrt{\pi}}{2\sqrt{2}}.$$

We have that

$$x_n = \sum_{i=1}^{n} a_i,$$

thus we get by the strong law of large numbers [22]

$$\lim_{n\to\infty} \frac{x_n}{n} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} a_i = \left(\frac{\pi}{8}\right)^{\frac{1}{2}}.$$

Similarly for y_n we get

$$\lim_{n\to\infty} \frac{y_n}{n} = \left(\frac{\pi}{8}\right)^{\frac{1}{2}}.$$

Thus,

$$\lim_{n\to\infty} \frac{t(n)}{n} = \lim_{n\to\infty} \max\left(\frac{x_n}{n}, \frac{y_n}{n}\right) = \max\left(\lim_{n\to\infty}\left(\frac{x_n}{n}, \frac{y_n}{n}\right)\right) = \left(\frac{\pi}{8}\right)^{\frac{1}{2}}.$$

Therefore

$$c = \lim_{n\to\infty} \frac{Y_{0,t(n)}}{t(n)} \geq \lim_{n\to\infty} \frac{n}{t(n)} = \left(\frac{8}{\pi}\right)^{\frac{1}{2}} > 1.59.$$

□

They study of longest increasing subsequence involves surprisingly many branches of mathematics. David Aldous and Persi Diaconis wrote a great paper on patience sorting and its connections with the topic [28]. There they discuss several interesting results one get from studying permutations. Patience sorting is a sorting algorithm that takes name after the the card game patience, also known as solitaire. There they show that the number of steps the algorithm uses to sort a deck of cards is equal to the longest increasing subsequence of a permutation.

Another topic they discuss is that it is conjectured that

$$l_n = 2\sqrt{n} - \mu_\infty n^{1/6} + o(n^{1/6}), \quad \sigma(L_n) = \sigma_\infty n^{1/6} + o(n^{1/6})$$

where μ_∞ and σ_∞ are constants and $\sigma(L_n)$ is the standard derivation. This was proved by Baik, Deift and Johansson in 1998 [33], with numerical values $\mu_\infty = 1.7111\cdots$ and $\sigma_\infty = 0.902\cdots$. In the same paper they proved that $L(\sigma_n)$ follows the Tracy-Widom distribution. More precisely they proved that

$$\mathbb{P}\left(\frac{L(\sigma_n) - 2\sqrt{n}}{n^{1/6}} \leq x\right) \to F_2(x) \quad \text{as } n \to \infty,$$

where $F_2(x)$ follows the Tracy-Widom distribution with parameter 2, see figure 10. This is known as the Baik-Deift-Johansson theorem.

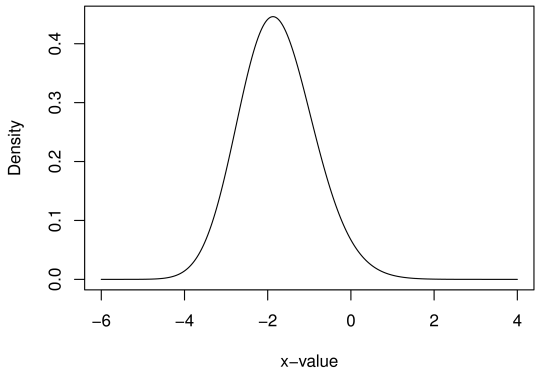

FIGURE 10. The density function $f_2(x) = F_2(x)$ associated with the Tracy-Widom distribution

Aldous and Diaconis also gives a proof in another paper that $c = 2$ [34], which is much easier proof than the original proofs of Logan and Shepp and Vershik and Kerov. Where both the original proofs are using random Young tableaux, Aldous and Diaconis used an interacting particle process.

Dan Romik wrote an excellent book on the subject, *The surprising mathematics of longest increasing subsequences* [29]. There he discusses all of this and much more. Sadly, the time constrains on this projects has prevented me to dig deeper into all of this.

References

[1] J. Michael Steele. Kingman's subadditive ergodic theorem. *Ann. Inst. H. Poincaré Probab. Statist.*, 25(1):93–98, 1989.

[2] J. F. C. Kingman. The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc. Ser. B*, 30:499–510, 1968.

[3] J. M. Hammersley and D. J. A. Welsh. First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. In *Proc. Internat. Res. Semin., Statist. Lab., Univ. California, Berkeley, Calif*, pages 61–110. Springer-Verlag, New York, 1965.

[4] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457–469, 1960.

[5] J. M. Hammersley. A few seedlings of research. pages 345–394, 1972.

[6] Karl Petersen. *Ergodic theory*, volume 2 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1983.

[7] Peter Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1982.

[8] Jonathan Marchini. Lecture 5 : The Poisson Distribution. `http://www.stats.ox.ac.uk/~marchini/teaching/L5/L5.notes.pdf`, November 10, 2008. [Online; accessed 1-May-2017].

[9] Sheldon M. Ross. *Introduction to probability models*. Elsevier/Academic Press, Amsterdam, eleventh edition, 2014.

[10] Hugo Duminil-Copin and Stanislav Smirnov. The connective constant of the honeycomb lattice equals $\sqrt{2+\sqrt{2}}$. *Ann. of Math. (2)*, 175(3):1653–1665, 2012.

[11] I. Jensen and A. J. Guttmann. Self-avoiding walks, neighbour-avoiding walks and trails on semiregular lattices. *J. Phys. A*, 31(40):8137–8145, 1998.

[12] George S. Lueker. Improved bounds on the average length of longest common subsequences. *J. ACM*, 56(3):Art. 17, 38, 2009.

[13] Marcos Kiwi, Martin Loebl, and Jiri Matousek. *Expected length of the longest common subsequence for large alphabets*, volume 2976 of *Lecture Notes in Comput. Sci.*, pages 302–311. Springer, Berlin, 2004.

[14] Marcelo Viana. *Lectures on Lyapunov exponents*, volume 145 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2014.

[15] George Casella and Roger L. Berger. *Statistical inference*. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990.

[16] Alexandr A. Borovkov. *Probability theory*, pages xxviii+733. Universitext. Springer, London, 2013. Translated from the 2009 Russian fifth edition by O. B. Borovkova and P. S. Ruzankin, Edited by K. A. Borovkov.

[17] Gennady Samorodnitsky. *Stochastic processes and long range dependence*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2016.

[18] James Polsinell. A Study of Probability and Ergodic theory with applications to Dynamical Systems. `http://www.d.umn.edu/math/Technical%20Reports/Technical%20Reports%202007-/TR%202010/TR_2010_1.pdf`, May 20, 2010. [Online; accessed 11-May-2017].

[19] Thomas M. Liggett. An improved subadditive ergodic theorem. *Ann. Probab.*, 13(4):1279–1285, 1985.

[20] Antonio Auffinger, Michael Damron, and Jack Hanson. 50 years of first passage percolation, 2015.

[21] S. R. Broadbent and J. M. Hammersley. Percolation processes. I. Crystals and mazes. *Proc. Cambridge Philos. Soc.*, 53:629–641, 1957.

[22] Rick Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fourth edition, 2010.

[23] J. Theodore Cox. The time constant of first-passage percolation on the square lattice. *Adv. in Appl. Probab.*, 12(4):864–879, 1980.

[24] O. Knill. Lyapunov Exponent. `http://www.math.harvard.edu/archive/118r_spring_05/handouts/lyapunov.pdf`, September 2, 2005. [Online; accessed 25 May-2017].

[25] Cosma Shalizi. Solution to homework 2. `http://www.stat.cmu.edu/~cshalizi/754/2006/notes/solutions-2.pdf`, 7 February 2006. [Online; accessed 04 June-2017].

[26] A. M. Lyapunov. *The general problem of the stability of motion.* Taylor & Francis, Ltd., London, 1992. Translated from Edouard Davaux's French translation (1907) of the 1892 Russian original and edited by A. T. Fuller, With an introduction and preface by Fuller, a biography of Lyapunov by V. I. Smirnov, and a bibliography of Lyapunov's works compiled by J. F. Barrett, Lyapunov centenary issue, Reprint of Internat. J. Control **5**5 (1992), no. 3 [MR1154209 (93e:01035)], With a foreword by Ian Stewart.

[27] Mark Pollicott. Maximal Lyapunov exponents for random matrix products. *Invent. Math.*, 181(1):209–226, 2010.

[28] David Aldous and Persi Diaconis. Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem. *Bull. Amer. Math. Soc. (N.S.)*, 36(4):413–432, 1999.

[29] Dan Romik. *The surprising mathematics of longest increasing subsequences.* Institute of Mathematical Statistics Textbooks. Cambridge University Press, New York, 2015.

[30] Roy L. Streit. *Poisson point processes.* Springer, New York, 2010. Imaging, tracking, and sensing.

[31] B. F. Logan and L. A. Shepp. A variational problem for random Young tableaux. *Advances in Math.*, 26(2):206–222, 1977.

[32] A. M. Vervsik and S. V. Kerov. Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of Young tableaux. *Dokl. Akad. Nauk SSSR*, 233(6):1024–1027, 1977.

[33] Jinho Baik, Percy Deift, and Kurt Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.

[34] D. Aldous and P. Diaconis. Hammersley's interacting particle process and longest increasing subsequences. *Probab. Theory Related Fields*, 103(2):199–213, 1995.