2013

Recurrence times in stochastic processes and dynamical systems

Mimoon Ibrahim Ismael
University of Wollongong

School of Mathematics and Applied Statistics

Recurrence Times in Stochastic Processes and Dynamical Systems

Mimoon Ibrahim Ismael
Master of Mathematics

Supervisors

Professor Rodney Nillsen and Professor Graham Williams

This thesis is presented as required for the

Award of the Degree of

Master of Science-Research in Mathematics

of the

University of Wollongong

August 2013

Certification

I, Mimoon Ibrahim Ismael, declare that this thesis, submitted in fulfillment of the requirements for the award of Master by Research in the School of Mathematics and Applied Statistics, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Mimoon Ibrahim Ismael
August, 2013

Acknowledgements

I wish to convey a massive and deep thanks for my supervisors who are Professor Rodney Nillsen and Professor Graham Williams for their many helpful, wonderful and detailed comments on the thesis.

Especially, I would like to thank my parents, my brothers, and my sisters for their tolerance, patience, and constant encouragement during all phases of my study.

I am also grateful to all those people who have encouraged and supported me during all phases of my study whether they are friends or teachers.

Last but not least, I wish to thank all staff of the library at the University of Wollongong for their kindly help in finding some bibliographic items.

Abstract

In this work we consider various aspects of recurrence times in stochastic processes and dynamical systems. The first part of the thesis is set in the context of zero-one stochastic processes. Here, by a zero-one stochastic process is meant a sequence of functions on a given set where each function takes values of either zero or one. The discussion is primarily concerned with stationary processes and is a rigourous and, in some aspects, a more general discussion of work of P. Kasteleyn. A connection between notions of recurrence in zero-one stochastic processes and dynamical systems admitting an invariant probability is established. The later part of the thesis presents new results in some special dynamical systems. These results are mainly to do with calculating the standard deviation of recurrence times and discussing the finiteness of the standard deviation, and are related to the existing literature.

Contents

1 Introduction **1**

2 Probabilities and probability functions **5**
2.1 Introduction: notation and definitions 5
2.2 Probabilities 6

3 Stochastic processes and dynamical systems **10**
3.1 Introduction: stochastic processes 10
3.2 Notations and definitions 10
3.3 Stationary processes 13
3.4 Relationship between zero-one stochastic processes and dynamical systems 13

4 Basic identities **18**
4.1 Introduction 18
4.2 Notations and preliminaries 18
4.3 The main identities 22

5 Recurrence times in stochastic processes **28**
5.1 Introduction: the notion of recurrence 28
5.2 Poincaré recurrence 29
5.3 Recurrence times in dynamical systems 34
5.4 The average of the recurrence time 36
5.5 The dynamical systems formulation 37

6 The standard deviation of the recurrence time **39**
6.1 Introduction 39
6.2 The standard deviation formula 39
6.3 The standard deviation in dynamical systems 46

7 Piecewise linear functions and their dynamical systems **48**
7.1 Introduction . . . 48
7.2 Piecewise linear functions . . . 48
7.3 The standard deviation for some piecewise linear transformations 56

8 Recurrence and standard deviation in a finite dynamical system **75**
8.1 Introduction . . . 75
8.2 Preliminaries . . . 76
8.3 Recurrence times and the average in a finite dynamical system 78
8.4 The standard deviation of the recurrence time . . . 81

9 Standard deviation of recurrence times in infinite discrete systems **94**
9.1 Introduction . . . 94
9.2 Finite sums of dynamical systems . . . 94
9.3 Countable sums of finite dynamical systems . . . 99
9.4 Recurrence and standard deviation in a sum of systems . . . 101

10 Conclusion **109**

Chapter 1

Introduction

There are phenomena which occur repeatedly in the natural world and in human affairs. Some of these phenomena are predictable, such as the rising of the sun every twenty four hours. However, other phenomena, such as earthquakes, are not predictable, despite the use of sophisticated technology in attempts to do so.

In mathematical terms, phenomena of repetition and recurrence may be studies in terms of stochastic processes (or times series) and, more particularly, in terms of dynamical systems. A zero-one stochastic processes is a sequence (X_n) of functions having a common domain, and having a common co-domain $\{0, 1\}$. The parameter n may be interpreted as time. Then, if $X_n(x) = 1$, we might say that if the process started in a state x, then "1" is observed at time n. Then, if $m > n$ and $X_m(x) = 1$, we might say that "1" has again been observed, or that "1" has *recurred* after a further elapse of time $m - n$.

In a dynamical system, we are given a set S and a transformation $f : S \longrightarrow S$. Given $x \in S$, the sequence $x, f(x), f^2(x), \ldots$ in S is called the *orbit* of x (note that we use the notation $f^n = f \circ f \circ \ldots \circ f$, where the composition is taken n times). Given a subset $U \subseteq S$, and given $x \in U$, we can consider whether there is another point $f^n(x)$ in the orbit of x that is also in U. The connection with stochastic processes comes about as follows. Let χ_U denote the characteristic function of U, that is

$$\chi_U(x) = \begin{cases} 1, & \text{if } x \in U, \\ 0, & \text{if } x \notin U. \end{cases}$$

Then, define $X_n : S \longrightarrow \{0, 1\}$ by putting $X_n(x) = \chi_U\Big(f^n(x)\Big)$. Then, (X_n) is a stochastic process, and recurrence phenomena in the process (X_n) are equivalent to recurrence phenomena in the dynamical system (S, f).

One aspect of this thesis is to examine parts of the work of Kasteleyn [18]. The aim here is to present some of Kasteleyn's work in a more formalised and mathematically rigorous way, to make explicit the underlying assumptions, and to provide complete proofs in so doing. Here, the aim also is to generalise Kasteleyn's approach in some respects, and produce some efficiencies.

Further, another aim of the thesis is to obtain new results concerning recurrence, and the standard deviation of recurrence times, for some particular dynamical systems. Three main types of dynamical systems are considered.

The first type is where the system (S, f) consists of a bounded interval of real numbers and $f : S \longrightarrow S$ is a piecewise linear transformation in the sense that on each subinterval of S that is in a given partition of S, f has a graph that is a straight line segment.

The second type of system (S, f) is where S is a finite set and f acts as a cyclic permutation on S.

Third type of system (S, f) is arising from an infinite "sum" of finite discrete systems, where $S = \bigcup_{j\in\mathbb{N}} S_j$, the S_j are disjoint sets and $f = f_j$ acts as a cyclic permutation on the whole of S_j, for all $j \in \mathbb{N}$. For all of these systems, new results are obtained for the standard deviation of recurrence times. This work in places sheds new light on some of the results in Kasteleyn, and it is also related to earlier work on the moments of recurrence times by Blum and Rosenblatt [4].

The following now gives an idea of the contents of each Chapter, and a more detailed view of the structure of the thesis. Detailed definitions and more technical issues are dealt with in the body of the thesis.

Chapter 2 introduces the notion of a probability and a probability function. A probability is a finitely additive set function, of total mass 1, on an algebra of sets. A probability function P is a probability on an algebra $\mathcal{B}$ of sets such that if (A_n) is a sequence of disjoint sets in $\mathcal{B}$ and $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}$, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

These are introduced to distinguish between different assumptions needed in the work of Kasteleyn [18] (see also [1], [5], and [11]).

Chapter 3 introduces zero-one stochastic processes and stationary processes, which play an important role. Technical preliminary results are derived and the relationship between stochastic processes and dynamical systems is discussed (see [3], [9], [10], and [17]).

Chapter 4 deals with some basic identities to be used in connection with the study of recurrence and recurrence times in stochastic processes. These results are due to Kasteleyn [18], but the approach here is more general and is more explicit in relation to the underlying definitions and assumptions.

Chapter 5 presents some of the main results in Kasteleyn [18] concerning recurrence. The Poincaré Theorem specifies when an initial state will recur within a finite time, or maybe approximately recur within a finite time. Also, Kac's formula says that the average time for a given event to recur is inversely proportional to the probability of the event. These results are derived for stochastic processes and specialised to dynamical systems. Again, there is an emphasis here on clarifying assumptions and giving complete proofs. Whereas the recurrence time is an observation of how long it has taken an event to recur, the standard deviation of recurrence times measures "how much" the various recurrence times deviate from the average value of the recurrence time as given by Kac's formula. Thus, the standard deviation is related to the "predictability" of the recurrence times (for more discussion and clarification see [12], [7], [19], [24], [8], and [28]).

Chapter 6 is concerned with work on the standard deviation of recurrence times, as originally looked at by Blum and Rosenblatt [4] and Kasteleyn [18]. There is also a discussion in [20, pages 270-284]. The discussion depends upon the technical tools previously developed.

Chapter 7 is concerned with a discussion of dynamical systems where the underlying set S is an interval and $f : S \longrightarrow S$ is a transformation that is piecewise linear, as previously described. New results are obtained for estimating the standard deviation in some of these systems. Related ideas and calculations are presented to obtain new results in a system $\Big([0,1), f\Big)$ where $f : [0,1) \longrightarrow [0,1)$ is the fractional part of $x+1/q$, where $q \in \{2,3,\ldots\}$ is specified in advance.

Chapter 8 deals with a finite dynamical system where the underlying set $S = \{u_0, u_1, \ldots, u_{|S|-1}\}$ is a finite set and $f : S \longrightarrow S$ is a cyclic permutation of S. (Here $|S|$ denotes the number of elements in a finite set S.) The system in Chapter 7 can be regarded more abstractly as a system (S, f), where S is a finite set and f is a cyclic permutation of S. New results are obtained for the standard deviation of the recurrence times which relate to Kac's formula for the average of recurrence times.

At the end, Chapter 9 deals with a countable "sum" of finite discrete dynamical systems where $S = \bigcup_{j\in\mathbb{N}} S_j$, put

$$f : S \longrightarrow S \text{ by } f(x) = f_j(x) \text{ for all } x \in S_j,$$

and f_j acts as a cyclic permutation on the whole of S_j. New results are obtained for the standard deviation of the recurrence times.

We now introduce some of the general notations used in the thesis. Others will be introduced later, as needed.

Given sets A and B, the union of the sets is denoted by $A \cup B$, and the intersection of sets is denoted by $A \cap B$. When A is a subset of B, we denoted this by $A \subseteq B$, and the complement of A is denoted by A^c. If A, B are sets, $A \cap B^c$ may be denoted by $A - B$. The union of sets $A_1, A_2, \ldots$ is denoted by $\bigcup_{n=1}^{\infty} A_n$, and the intersection of sets $A_1, A_2, \ldots$ is denoted by $\bigcap_{n=1}^{\infty} A_n$. The empty set is denoted by $\emptyset$. If A, B are sets, the Cartesian product of A, B is

$$A \times B = \Big\{(a, b) : a \in A \text{ and } b \in B\Big\}.$$

Similarly, if we have sets $A_1, A_2, \ldots, A_n$, the Cartesian product of the sets is

$$A_1 \times A_2 \times \ldots \times A_n = \Big\{(a_1, a_2, \ldots, a_n) : a_j \in A_j \text{ for all } j = 1, 2, \ldots, n\Big\}.$$

A Cartesian product of the form $A \times A \times \ldots \times A$, taken r times, is denoted by A^r.

The set of positive integers or natural numbers is denoted by $\mathbb{N} = \{1, 2, \ldots\}$, the set of integers is denoted by $\mathbb{Z}$, the set of non-negative integers is denoted by $\mathbb{Z}_+$, and the set of real numbers is denoted by $\mathbb{R}$. If (x_n) is a sequence of non-negative terms, the sum of the sequence is denoted by $\sum_{n=1}^{\infty} x_n$. A function mapping domain A into co-domain B is denoted by $f : A \longrightarrow B$, a transformation f on a set A is a function $f : A \longrightarrow A$, the image of the set A under function f is denoted by $f(A)$, and the inverse image of the set A under function f is denoted by $f^{-1}(A)$. Given functions f and g, the composition of function f with function g is denoted by $f \circ g$, and the n^{th} iterate of a transformation f is denoted by $f^n = f \circ f \circ \cdots \circ f$, where composition is taken n times. The end of a proof is denoted by $\square$.

Chapter 2

Probabilities and probability functions

2.1 Introduction: notation and definitions

We will give important notation and definitions considering probability in this section. Also, some properties of probabilities and probability functions are discussed, as they will be used in later chapters.

Definition 2.1.1.

A family $\mathcal{B}$ of subsets of a set S is said to be an *algebra of subsets* of S when:

(1) $\emptyset, S \in \mathcal{B}$.

(2) If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$.

(3) If $A, B \in \mathcal{B}$, then $A \cup B \in \mathcal{B}$.

Definition 2.1.2.

Property (3) in definition (2.1.1) immediately implies that $\mathcal{B}$ is closed under finite unions: $\bigcup_{j=1}^{n} A_j \in \mathcal{B}$ whenever $A_1, A_2, \ldots, A_n \in \mathcal{B}$. Also, if $A_1, A_2, \ldots, A_n \in \mathcal{B}$, then

$$\bigcap_{j=1}^{n} A_j = \bigcap_{j=1}^{n} \left(A_j^c\right)^c = \left(\bigcup_{j=1}^{n} A_j^c\right)^c \in \mathcal{B}.$$

So, a finite intersection of sets in $\mathcal{B}$ is also in $\mathcal{B}$.

If in addition $\mathcal{B}$ is closed under countable unions, that is, if

$$\bigcup_{j=1}^{\infty} A_j \in \mathcal{B} \text{ when } A_j \in \mathcal{B} \text{ for every } j \in \mathbb{N},$$

then $\mathcal{B}$ is called a $\sigma - algebra$ [22, page 102](see also [27, page 308]).

Definition 2.1.3.

Let S be a set and let $\mathcal{B}$ be an algebra of subsets of S. A *probability* P on $\mathcal{B}$ is a function $P : \mathcal{B} \to [0, 1]$ such that the following hold:

(1) $P(S) = 1$ and $P(\emptyset) = 0$.

(2) If we have $A, B \in \mathcal{B}$ and $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.

When $A \in \mathcal{B}$, we call A an *event* and $P(A)$ is called the *probability* of A or, in other words, the probability of the event A.

Definition 2.1.4.

A $probability\ function$ on $\mathcal{B}$ is a probability on $\mathcal{B}$ such that if $A_n \in \mathcal{B}$ for all $n \in \mathbb{N}, A_n \cap A_m = \emptyset$ for all $m \neq n$, and $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}$, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \tag{2.1}$$

A probability is often referred to in the literature as a $finitely\ additive\ measure$ with a total mass of 1. Also, a probability function is often referred to as a *countably additive measure* or $\sigma - additive\ measure$ with a total mass of 1.

2.2 Probabilities

In this section, let S be a set, let $\mathcal{B}$ be an algebra of subsets of S, and let P be a probability function on $\mathcal{B}$. Note that if $A, B \in \mathcal{B}$ and $A \subseteq B$, then $P(A) \leq P(B)$. This fact may be used without explicit reference.

Lemma 2.2.1. *Let P be a probability function on $\mathcal{B}$. Then if (A_n) is a sequence of sets in $\mathcal{B}$, and if $A \in \mathcal{B}$ are such that $A \subseteq \bigcup_{n=1}^{\infty} A_n$, then* $P(A) \leq \sum_{n=1}^{\infty} P(A_n)$.

Proof. Put $V_1 = A_1$, $V_2 = A_2 \cap A_1^c$ and so on, putting $V_n = A_n \cap (\bigcup_{k=1}^{n-1} A_k)^c$. Now we will show $V_1 \cap V_2 = \emptyset$. Let us assume $V_1 \cap V_2 \neq \emptyset$. Let $x \in V_1 \cap V_2$ then $x \in V_2 \Rightarrow x \in A_2 \cap V_1^c \Rightarrow x \notin V_1$. As $x \in V_1$ and $x \in V_1^c$, we have a contradiction. Thus, $V_1 \cap V_2 = \emptyset$.

In the general case we will show that the sets V_n are pairwise disjoint.

We have

$$\begin{aligned} x \in V_n &\Rightarrow x \in A_n \cap (A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_{n-1})^c, \\ &\Rightarrow x \notin A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_{n-1}, \\ &\Rightarrow x \notin A_1, x \notin A_2, x \notin A_3, \ldots, x \notin A_{n-1}, \\ &\Rightarrow x \notin A_j, \text{for all } 1 \leq j \leq n-1, \\ &\Rightarrow x \notin V_j, \text{ since } V_j \subseteq A_j. \end{aligned}$$

Thus,

$$V_n \cap V_j = \emptyset, \text{ if } 1 \leq j \leq n-1. \tag{2.2}$$

In general for all $m, n \in \mathbb{N}$ where $m \neq n$ then if $m < n$, by (2.2), $V_n \cap V_m = \emptyset$. Also if $m > n$, then $V_n \cap V_m = \emptyset$ by (2.2).

Also we will show $V_1 \cup V_2 = A_1 \cup A_2$. From the definitions we have

$$\begin{aligned} V_1 \cup V_2 &= A_1 \cup V_2 \\ &= A_1 \cup (A_2 \cap A_1^c) \\ &= (A_1 \cup A_2) \cap (A_1 \cup A_1^c) \\ &= A_1 \cup A_2. \end{aligned}$$

In general we will show $\bigcup_{j=1}^{r} V_j = \bigcup_{j=1}^{r} A_j$. We will use mathematical induction. The case $r = 1$ is true as $V_1 = A_1$, by definition.
Now we assume the result true for r. Then,

$$\bigcup_{j=1}^{r} V_j = \bigcup_{j=1}^{r} A_j.$$

So for $r + 1$ we have

$$\begin{aligned} & V_1 \cup V_2 \cup ... \cup V_r \cup V_{r+1} \\ = \;& (A_1 \cup A_2 \cup ... \cup A_r) \cup V_{r+1} \\ = \;& \Big(A_1 \cup A_2 \cup ... \cup A_r\Big) \cup \left(A_{r+1} \cap \Big(\bigcup_{j=1}^{r} A_j\Big)^c\right) \\ = \;& \Big(A_1 \cup A_2 \cup ... \cup A_{r+1}\Big) \cap \left((A_1 \cup A_2 \cup ... \cup A_r) \cup \Big(\bigcup_{j=1}^{r} A_j\Big)^c\right) \\ = \;& A_1 \cup A_2 \cup ... \cup A_{r+1}. \end{aligned}$$

So the statement holds for $r+1$. It is now true by induction that for all n,

$$\bigcup_{j=1}^{n} V_j = \bigcup_{j=1}^{n} A_j.$$

Therefore, $V_j \in \mathcal{B}$ and $A = A \cap (\bigcup_{n=1}^{\infty} A_n) = A \cap (\bigcup_{n=1}^{\infty} V_n) = \bigcup_{n=1}^{\infty} A \cap V_n$. We know that $V_n \in \mathcal{B}$ and also $A \cap V_n \in \mathcal{B}$, and as P is a probability function, and the sets $A \cap V_n$ are disjoint,

$$P(A) = \sum_{n=1}^{\infty} P(A \cap V_n) \leq \sum_{n=1}^{\infty} P(A_n).$$

□

Example 2.2.1.

In this example we let $\mathcal{B}$ be the collection of subsets of $\mathbb{R}$ that are either countable or have countable complements. We show that $\mathcal{B}$ is an algebra of subsets of $\mathbb{R}$.
1) $\emptyset$ is countable and so belongs to $\mathcal{B}$. Also the complement of $\mathbb{R}$ is countable and so $\mathbb{R}$ belongs to $\mathcal{B}$.
2) There are two cases, the first when $A \in \mathcal{B}$ is countable. Observe that $(A^c)^c = A$, and so $A^c \in \mathcal{B}$. The second case is when A^c is countable, then $A^c \in \mathcal{B}$ by the definition of $\mathcal{B}$.
3) we consider A as the finite union of the sets A_j, where some of the A_j are countable, and the others have countable complements. Let the sets A_j that are countable be denoted by B_j, and those that have countable complements be denoted by C_j. Then, $\cup A_j = (\cup B_j) \cup (\cup C_k)$, and $B_j, C_k \in \mathcal{B}$ for all j, k.

Now we have to show if $A, B \in \mathcal{B}$ then $A \cup B \in \mathcal{B}$ in this case where A is countable and B^c is countable. Observe that

$$(A \cup B)^c = A^c \cap B^c \subseteq B^c \text{ so } A \cup B \in \mathcal{B}.$$

Thus, $\mathcal{B}$ is an algebra.

There are many probabilities on $\mathcal{B}$. For example, if $x \in \mathbb{R}$ then we put

$$\delta_x(A) = \begin{cases} 1, & \text{if } x \in A. \\ 0, & \text{if } x \notin A \ . \end{cases}$$

For all $A \in \mathcal{B}$. Then, $\delta_x(A)$ is probability on $\mathcal{B}$.

Example 2.2.2.

Let $\mathcal{B}$ be the collection of all subsets of $\mathbb{R}$. If $A \in \mathcal{B}$, we put

$$P(A) = \begin{cases} 1, & \text{if } A^c \text{ is countable.} \\ 0, & \text{if } A^c \text{ is uncountable.} \end{cases}$$

We will check whether P is a probability on $\mathcal{B}$ where $P : \mathcal{B} \rightarrow \{0, 1\}$. We have $\mathbb{R}$ is an uncountable set and $\mathbb{R}^c$ is countable because it is the empty set, that is $P(\mathbb{R}) = 1$. Let $A, A^c \subseteq \mathbb{R}$ be uncountable. Then $\mathbb{R} = A \cup A^c$, where this union is disjoint. Then, $1 = P(\mathbb{R}) \neq P(A) + P(A^c)$ as $P(A) = P(A^c) = 0$. Thus, P is not a probability on $\mathcal{B}$.

Lemma 2.2.2. *Let S be a set, let P be a probability function on an algebra $\mathcal{B}$ of subsets of S. Let $A_1, A_2, \ldots \in \mathcal{B}$ and assume that*

$$A_1 \supseteq A_2 \supseteq \ldots .$$

Put $A = \bigcap_{j=1}^{\infty} A_j$ and assume that $A \in \mathcal{B}$. Then $P(A) = \lim\limits_{j\to\infty} P(A_j)$.

Proof. Let (A_j) be a decreasing sequence in $\mathcal{B}$ and so also we have $(P(A_j))$ is non-negative and decreasing, so the $\lim_{j\to\infty} P(A_j)$ exists. Now we have

$$A_1 = A \cup (A_1 \cap A_2^c) \cup (A_2 \cap A_3^c) \cup \ldots \cup (A_j \cap A_{j+1}^c) \cup \ldots ,$$

and we have P is a probability function on algebra $\mathcal{B}$, so

$$\begin{aligned} P(A_1) &= \sum_{j=1}^{\infty} P(A_j \cap A_{j+1}^c) + P(A) \\ &= \sum_{j=1}^{\infty} \Big(P(A_j) - P(A_{j+1}) \Big) + P(A). \end{aligned}$$

Since

$$\sum_{j=1}^{\infty} \Big(P(A_j) - P(A_{j+1}) \Big) = P(A_1) - \lim_{j\to\infty} P(A_j),$$

we have

$$P(A_1) = P(A_1) - \lim_{j\to\infty} P(A_j) + P(A),$$

and

$$0 = \lim_{j\to\infty} -P(A_j) + P(A).$$

Therefore,

$$P(A) = \lim_{j\to\infty} P(A_j).$$

□

Chapter 3

Stochastic processes and dynamical systems

3.1 Introduction: stochastic processes

A zero-one stochastic process can be viewed as an infinite sequence of functions $X_0, X_1, \ldots, X_n, \ldots$, where each function takes its values in $\{0, 1\}$. That is, $X_n : S \longrightarrow \{0, 1\}$ is a given function for all $n \in \mathbb{Z}_+$. The parameter n may be taken to represent the time, so that if $x \in S$, at time n we make an observation $X_n(x)$, and we observe either a 0 or a 1. We can take successive observations $X_0(x), X_1(x), X_2(x), \ldots$, starting from time 0, since we keep x fixed. General references are [2], [3], [9], [10], and [17].

3.2 Notations and definitions

We will introduce some definitions and examples which relate to stochastic processes and dynamical systems.

Definition 3.2.1.

Let S be a set and for each $n \in \mathbb{Z}_+$ let $X_n : S \to \{0,1\}$. Let $\mathcal{B}$ be an algebra of subsets of S and P a probability on $\mathcal{B}$. Now we say $(S, \mathcal{B}, (X_n), P)$, or simply $(S, (X_n))$, is a $zero - one\ stochastic\ process$ if for every $n \in \mathbb{Z}_+$ and $d \in \{0, 1\}$, the event $\{x : x \in S \text{ and } X_n(x) = d\} \in \mathcal{B}$. Then for every $x \in S$, $(X_n(x))$ is a sequence of zeros and ones. Note that P need only be a probability, not a probability function. Now we assume that $d \in \{0, 1\}$ and $n \in \mathbb{Z}_+$. We define $A(n|d)$ by

$$A(n|d) = \{x : x \in S \,\text{and}\, X_n(x) = d\}.$$

As $(S,(X_n))$ is a stochastic process, $A(n|d) \in \mathcal{B}$. Now let $n_1, n_2, \ldots, n_r \in \mathbb{Z}_+$ with $0 \leq n_1 < n_2 < \ldots < n_r$, and let $d_j \in \{0,1\}$ for all $j = 1, 2, \ldots, r$. We define

$$A(n_1, \ldots, n_r|d_1, \ldots, d_r) = \{x : x \in S \,\text{and}\, X_{n_j}(x) = d_j, \text{for all } j = 1, 2, \ldots, r\}.$$

Observe that

$$A(n_1, \ldots, n_r|d_1, \ldots, d_r) = \bigcap_{j=1}^{r} A(n_j|d_j),$$

so that $A(n_1, \ldots, n_r|d_1, \ldots, d_r) \in \mathcal{B}$, as $\mathcal{B}$ is an algebra and as each $A(n_j|d_j) \in \mathcal{B}$. Then we define $P(n_1, \ldots, n_r|d_1, \ldots, d_r)$ by

$$P(n_1, \ldots, n_r|d_1, \ldots, d_r) = P\Big(\{x : X_{n_j}(x) = d_j, \text{for all } j = 1, 2, \ldots, r\}\Big).$$

Thus,

$$P(n_1, \ldots, n_r|d_1, \ldots, d_r) = P\Big(A(n_1, \ldots, n_r|d_1, \ldots, d_r)\Big).$$

We put

$$A(d_1, d_2, \ldots, d_r) = A(0, 1, 2, \ldots, r-1|d_1, d_2, \ldots, d_r),$$

and

$$P(d_1, d_2, \ldots, d_r) = P\Big(A(d_1, d_2, \ldots, d_r)\Big).$$

Now in case $d = (d_1, d_2, \ldots, d_r)$, r is called *the length* of d. In this case $P(d_1, d_2, \ldots, d_r)$ may be denoted by $P(d)$. If d is the empty sequence, then it has length 0. Also, if we have $d = (d_1, \ldots, d_r)$ and $c = (c_1, \ldots, c_s)$, then we define (d, c) to be the sequence $(d_1, \ldots, d_r, c_1, \ldots, c_s)$. So for $P(d, 0)$ we will have

$$(d, 0) = (d_1, d_2, \ldots, d_r, 0),$$

so that

$$P(d, 0) = P\Big(\{x : X_0(x) = d_1, X_1(x) = d_2, \ldots, X_{r-1}(x) = d_r, X_r(x) = 0\}\Big). \tag{3.1}$$

And also if we consider $P(d, 1)$,

$$P(d, 1) = P\Big(\{x : X_0(x) = d_1, X_1(x) = d_2, \ldots, X_{r-1}(x) = d_r, X_r(x) = 1\}\Big). \tag{3.2}$$

Therefore,

$$P(d, 0) + P(d, 1) = P\Big(\{x : X_0(x) = d_1, X_1(x) = d_2, \ldots, X_{r-1}(x) = d_r\}\Big) = P(d),$$

because both equations (3.1) and (3.2) involve disjoint unions of two sets in $\mathcal{B}$. Consequently we will have

$$P(d) = P(d, 0) + P(d, 1). \tag{3.3}$$

Similarly, if d is a sequence finite of symbols in $\{0, 1\}$,

$$P(d) = \sum_{c \in \{0,1\}^r} P(d, c), \text{ for all } r \in \mathbb{N}.$$

This is proved below in Theorem 3.2.1.

Example 3.2.1.

Consider $P(2, 3, 4|1, 0, 1)$. By the definition,

$$P(2, 3, 4|1, 0, 1) = P\Big(\{x : X_2(x) = 1, X_3(x) = 0, X_4(x) = 1\}\Big).$$

Example 3.2.2.

If we consider $P(0, 1, 2, 11, 14|1, 1, 0, 1, 1)$, we will have

$$\begin{aligned} & P(0, 1, 2, 11, 14|1, 1, 0, 1, 1) \\ = \; & P\Big(\{x : X_0(x) = 1, X_1(x) = 1, X_2(x) = 0, X_{11}(x) = 1, X_{14}(x) = 1\}\Big). \end{aligned}$$

Now, some general calculations are in the following Theorem.

Theorem 3.2.1. *Let $d = (d_1, d_2, \ldots, d_s)$ where $d_j \in \{0, 1\}$, for all $j \in \{1, 2, \ldots, s\}$. Then,*

$$P(d) = \sum_{c \in \{0,1\}^r} P(d, c), \text{ for all } r \in \mathbb{N}.$$

Proof. To prove this we will use mathematical induction on r. Let $d = (d_1, d_2, \ldots, d_s)$ where $d_j \in \{0, 1\}$ for each j.
The case when $r = 1$ is true by (3.3). We assume the result is true for r. Now we will check for $r + 1$. We have

$$\begin{aligned} P(d) &= \sum_{c \in \{0,1\}^r} P(d, c) \\ &= \sum_{c \in \{0,1\}^r} P(d, c, 0) + P(d, c, 1), \text{ by (3.3)}, \\ &= \sum_{c' \in \{0,1\}^{r+1}} P(d, c') \\ &= \sum_{c \in \{0,1\}^{r+1}} P(d, c), \end{aligned}$$

so the statement holds for $r + 1$ in place of r. □

3.3 Stationary processes

Definition 3.3.1.

Let $(S, \mathcal{B}, P, (X_n))$ be a zero-one stochastic process. Then the process is called *stationary* if for all $d_1, d_2, \ldots, d_r \in \{0,1\}$, and for all $s \in \mathbb{Z}_+$, we have

$$P(d_1, d_2, \ldots, d_r) = \sum_{c \in \{0,1\}^s} P(c, d_1, d_2, \ldots, d_r).$$

As in [5, page 104] and [10, page 94], this definition means that the probability of observing the consecutive occurrence of the symbols $d_1, d_2, \ldots, d_r$ remains the same over time. Note that

$$\sum_{c \in \{0,1\}^s} P(c, d_1, d_2, \ldots, d_r) = P(s, s+1, \ldots, s+r-1 | d_1, d_2, \ldots, d_r).$$

Consequently, the stationarity of $(S, \mathcal{B}, P, (X_n))$ is equivalently to requiring

$$P(0, 1, \ldots, r-1 | d_1, d_2, \ldots, d_r) = P(s, s+1, \ldots, s+r-1 | d_1, d_2, \ldots, d_r),$$

for all $r \in \mathbb{N}, d_1, d_2, \ldots, d_r \in \{0,1\}$, and $s \in \mathbb{Z}_+$.

That is,

$$P(d_1, d_2, \ldots, d_r) = P(s, s+1, \ldots, s+r-1 | d_1, d_2, \ldots, d_r).$$

From the definition above if we are given a finite sequence $(d_1, d_2, \ldots, d_r) \in \{0,1\}^r$ then,

$$P(0, d) + P(1, d) = \sum_{c \in \{0,1\}} P(c, d) = P(d), \text{ by stationarity.}$$

3.4 Relationship between zero-one stochastic processes and dynamical systems

Definition 3.4.1.

Let $f : S \to S$ be a transformation on a set S. Then, (S, f) is called a *dynamical system*. As well, given an algebra $\mathcal{B}$ of subsets of S and a probability P on $\mathcal{B}$, then for each $U \in \mathcal{B}$ we can define a zero-one stochastic process as follows.

Definition 3.4.2.

If $U \subseteq S$, we define

$$\chi_U(x) = \begin{cases} 1, & \text{if } x \in U, \\ 0, & \text{if } x \notin U. \end{cases}$$

We call the function χ_U the *characteristic* function of U.
So

$$\chi_U \circ f(x) = \chi_U(f(x)) = \begin{cases} 1, & \text{if } f(x) \in U, \\ 0, & \text{if } f(x) \notin U. \end{cases}$$

Now we use the definition that

$$f^{-1}(U) = \{x : f(x) \in U\}, \text{ where } x \in f^{-1}(U) \Leftrightarrow f(x) \in U.$$

Thus,

$$\chi_U \circ f(x) = \begin{cases} 1, & \text{if } x \in f^{-1}(U), \\ 0, & \text{if } x \notin f^{-1}(U). \end{cases}$$

Therefore, $\chi_U \circ f = \chi_{f^{-1}(U)}$.

Definition 3.4.3.

Let (S, f) be a dynamical system, let $\mathcal{B}$ be an algebra of subsets of S and let P be a probability on $\mathcal{B}$. Assume that for all $U \in \mathcal{B}$, $f^{-1}(U) \in \mathcal{B}$. Then, for a given $U \in \mathcal{B}$ with $(S, f, \mathcal{B}, P)$ we define an *associated zero − one stochastic process* $(S, \mathcal{B}, (X_n), P)$ by putting

$$X_n(x) = \chi_U(f^n(x)), \text{for all}\, n \in \mathbb{N}\, \text{and}\, x \in S.$$

That is,

$$X_n = \chi_U \circ f^n = \chi_{f^{-n}(U)}.$$

Note that $X_n(x) \in \{0, 1\}$. Also, note that if $d \in \{0, 1\}$,

$$\{x : x \in S \text{ and } X_n(x) = d\}.$$

We prove below that this set equals either $f^{-n}(U)$ or $f^{-n}(U)^c$, but in either case the set is in $\mathcal{B}$. Then, by definition $(S, \mathcal{B}, (X_n), P)$ will be a zero-one stochastic process.

To show that this is a zero-one stochastic process, by definition 3.2.1 we must show that for each $n = 1, 2, \ldots$ and $d \in \{1, 0\}$,

$$\{x : x \in S \text{ and } X_n(x) = d\} \in \mathcal{B}.$$

Observe that in particular if $d = 0$ then

$$\{x : x \in S \text{ and } X_n(x) = 0\} = \{x : f^n(x) \notin U\},$$

and if $d = 1$ then

$$\{x : x \in S \text{ and } X_n(x) = 1\} = \{x : f^n(x) \in U\}.$$

We have to show $f^{-n}(U^c), f^{-n}(U) \in \mathcal{B}$. However, $f^{-n}(U) = f^{-1}\Big(f^{-(n-1)}(U)\Big)$ and by mathematical induction on n it will be enough to show $f^{-1}(U) \in \mathcal{B}$ wherever $U \in \mathcal{B}$. Since, if $U \in \mathcal{B}$ then

$$f^{-1}(U) \in \mathcal{B}, f^{-1}\Big(f^{-1}(U)\Big) = f^{-2}(U) \in \mathcal{B}, f^{-1}\Big(f^{-2}(U)\Big) = f^{-3}(U) \in \mathcal{B},$$

and so on.

Now, for future use, we define

$$\begin{aligned} U_1 &= \{x : x \in U, f(x) \in U\}, \\ U_2 &= \{x : x \in U, f(x) \notin U, f^2(x) \in U\}, \end{aligned}$$

and so on. Then

$$U_n = \{x : x \in U, f(x) \notin U, \ldots, f^{n-1}(x) \notin U, f^n(x) \in U\}.$$

In particular, for $r \geq 2$,

$$\begin{aligned} &A\underbrace{(1, 0, 0, \ldots, 1)}_{r \text{ times}} \\ &= \{x : x \in S, \chi_U(x) = 1, \chi_U(f^j(x)) = 0 \text{ for } j = 1, \ldots, r-2, \chi_U(f^{r-1}(x)) = 1\} \\ &= \{x : x \in U, f(x) \notin U, \ldots, f^{r-2}(x) \notin U, f^{r-1}(x) \in U\} \\ &= U_r. \end{aligned}$$

Also, more generally

$$A(n_1, \ldots, n_r | d_1, \ldots, d_r) = \{x : x \in S \text{ and } \chi_U(f^{n_j}(x)) = d_j \text{ for } j = 1, 2, \ldots, r\}.$$

Example 3.4.1.

Consider (S, f) a dynamical system where $f : S \to S$ with $S = \{a, b, c, d\}$ and $U = \{b, c, d\} \subseteq S$. Let $\mathcal{B}$ be the algebra of all subsets of S, and let P be a probability on $\mathcal{B}$ where $P(U)$ is defined as the number of elements in U divided by 4. Let $f(a) = b, f(b) = a, f(c) = d$ and $f(d) = c$. Now,

$$X_0(b) = \chi_U(b) = 1 \text{ and } X_1(b) = \chi_{f^{-1}(U)}(b) = \chi_U(f(b)) = \chi_U(a) = 0,$$

and

$$X_2(b) = \chi_{f^{-2}(U)}(b) = \chi_U(f^2(b)) = \chi_U(b) = 1.$$

That is, $X_0(b) = 1, X_1(b) = 0, X_2(b) = 1$.

Definition 3.4.4.

Let (S, f) be a dynamical system, let $\mathcal{B}$ be an algebra of subsets of S, and let P be a probability on $\mathcal{B}$. Assume that $f^{-1}(C) \in \mathcal{B}$ for all $C \in \mathcal{B}$. We say that f is *P-invariant* if for all $C \in \mathcal{B}, P(f^{-1}(C)) = P(C)$.

Theorem 3.4.1. *Let S be a set, let $\mathcal{B}$ be an algebra of subsets of S, and let P be a probability on $\mathcal{B}$. Let $f : S \to S$ be a P-invariant transformation on S and let $A \in \mathcal{B}$ be given. Then, for $n \in \mathbb{Z}_+$ and $x \in S$ put*

$$X_n(x) = \chi_A(f^n(x)).$$

Then, the zero-one stochastic process $(S, \mathcal{B}, (X_n), P)$ is stationary.

Proof. Let $d_1, d_2, \ldots, d_r \in \{0, 1\}$ and let $s \in \mathbb{Z}_+$. Then,

$$\begin{aligned}
&P(s, s+1, \ldots, s+r-1 | d_1, d_2, \ldots, d_r) \\
&= P\Big(\bigcap_{j=0}^{r-1} A(s+j|d_{j+1})\Big), \\
&= P\Big(\bigcap_{j=0}^{r-1} \{x : X_{s+j}(x) = d_{j+1}\}\Big), \\
&= P\Big(\bigcap_{j=0}^{r-1} \{x : \chi_A(f^{s+j}(x)) = d_{j+1}\}\Big), \\
&= P\Big(\bigcap_{j=0}^{r-1} \{x : \chi_A(f^j(f^s(x))) = d_{j+1}\}\Big), \\
&= P\Big(\bigcap_{j=0}^{r-1} \{x : \chi_{f^{-j}A}(f^s(x)) = d_{j+1}\}\Big), \\
&= P\Big(\bigcap_{j=0}^{r-1} f^{-s}\{y : \chi_{f^{-j}A}(y) = d_{j+1}\}\Big), \\
&= P\Big(f^{-s}\Big(\bigcap_{j=0}^{r-1} \{y : \chi_{f^{-j}A}(y) = d_{j+1}\}\Big)\Big), \\
&= P\Big(\bigcap_{j=0}^{r-1} \{y : \chi_{f^{-j}A}(y) = d_{j+1}\}\Big), \text{as } f \text{ is P-invariant}, \\
&= P(0, 1, \ldots, r-1 | d_1, d_2, \ldots, d_r).
\end{aligned}$$

Thus, $(S, \mathcal{B}, (X_n), P)$ is stationary, by definition. □

Note that to apply this Theorem we must first check that $f^{-1}(C) \in \mathcal{B}$ for each $C \in \mathcal{B}$.

Example 3.4.2.

This example is discussed in more details in [20, page 150]. Take $S = \Big[0, 1\Big)$, take $\mathcal{B}$ to be the algebra of subsets of S that are finite unions of intervals , and let

$$f(x) = \begin{cases} 2x, & \text{if } 0 \leq x < 1/2, \\ 2x - 1, & \text{if } 1/2 \leq x < 1. \end{cases}$$

We define a probability P on $\mathcal{B}$ by putting, for any interval $I, P(I) = \textit{length of } I$ and then extend this in an obvious way to be defined on all of $\mathcal{B}$. It is possible to show that f is P-invariant (see [20]). If

$$U = [0, 1/2) \in \mathcal{B},$$

$$X_n(x) = \chi_U(f^n(x)) = \begin{cases} 1, & \text{if the } n^{th} \text{ digit in the binary expansion of } x \text{ is } 0, \\ 0, & \text{if the } n^{th} \text{ digit in the binary expansion of } x \text{ is } 1. \end{cases}$$

Now, by using Theorem 3.4.1 the zero-one stochastic process $([0, 1), \mathcal{B}, (X_n), P)$ is stationary.

Chapter 4

Basic identities

4.1 Introduction

In this Chapter we introduce some further definitions concerning probabilities and stochastic processes. Some properties and identities are discussed and proved for future use. The discussion is strongly influenced by Kasteleyn [18], but here there is more emphasis on precise definitions and clarification of underlying assumptions. Some of the results are more general than the corresponding results in [18]; in particular Lemma 4.2.1 and also parts of Theorem 4.3.1.

4.2 Notations and preliminaries

Consider a stochastic process as in Section 3.2 and we note that if $d \in \{0, 1\}$, then the set

$$X_t^{-1}(d) = \{x : x \in S \text{ and } X_t(x) = d\} \in \mathcal{B}, \tag{4.1}$$

for each $t = 0, 1, \ldots$. If we have a finite number of symbols, $d_1, \ldots, d_r \in \{0, 1\}$, and we form the set

$$\{x : x \in S \text{ and } X_0(x) = d_1, X_1(x) = d_2, \ldots, X_{r-1}(x) = d_r\}, \tag{4.2}$$

we observe that this set equals

$$\bigcap_{j=0}^{r-1} \{x : x \in S \text{ and } X_j(x) = d_{j+1}\} = \bigcap_{j=0}^{r-1} X_j^{-1}(d_{j+1}). \tag{4.3}$$

It follows from (4.1) and (4.3) that the set in (4.2) is a finite intersection of sets in $\mathcal{B}$ and so it must be in $\mathcal{B}$ because $\mathcal{B}$ is an algebra. That is, we have

$$\{x : x \in S \text{ and } X_0(x) = d_1, X_1(x) = d_2, \ldots, X_{r-1}(x) = d_r\} \in \mathcal{B}.$$

The set of all finite sequences of zeros and ones is denoted by Σ. Also, the set of all sequences of zeros and ones, finite or infinite, is denoted by Σ_∞. Thus, if $d = (d_1, \ldots, d_r) \in \Sigma$, has length r, as in Section 3.2, we put

$$A(d) = \Big\{x : X_j(x) = d_{j+1}, \text{ for } j = 0, 1, 2, \ldots, r-1\Big\} = \bigcap_{j=0}^{r-1} X_j^{-1}(d_{j+1}).$$

Note that $A(d) \in \mathcal{B}$ as it's a finite intersection of sets in $\mathcal{B}$. If $c = \{c_1, c_2, \ldots, c_s\} \in \Sigma$, and $d \in \{0, 1\}$, then $cd^\infty \in \Sigma_\infty$ is the sequence $\{c_1, c_2, \ldots, c_s, d_1, d_2, \ldots\}$.

Now if $d \in \Sigma_\infty$, we put

$$A(d) = \bigcap_{j=0}^{\infty} \{x : x \in S \text{ and } X_j(x) = d_{j+1}\},$$

for all $j \in \mathbb{N}$. However, we can not say that $A(d) \in \mathcal{B}$ or that $P(A(d))$ is defined. We will show how to extend the definition of P to include such sets. Observe that if $d \in \Sigma$, and d is a empty sequence, then d has length zero. Now, we proceed as above, allowing for the infinite sequence of symbols. The intersection of sets equals

$$\bigcap_{j=0}^{\infty} \{x : x \in S \text{ and } X_j(x) = d_{j+1}\} = \bigcap_{j=0}^{\infty} X_j^{-1}(d_{j+1}).$$

Now if $d \in \Sigma_\infty$, it may be that $A(d) \notin \mathcal{B}$. But in this case we observe that $(P(A(d_1, \ldots, d_r))_{r=1}^{\infty}$ is a decreasing sequence and we put

$$\widetilde{P}(A(d)) = \lim_{r\to\infty} P(d_1, d_2, \ldots, d_r).$$

Because $(P(d_1, d_2, \ldots, d_r))$ is a decreasing and non-negative sequence, the limit $\widetilde{P}(d)$ exists, for all $d \in \Sigma_\infty$. An important property of a probability function P is that it has a type of continuity property as expressed in equation (2.1). In the case when $A(d) \in \mathcal{B}$, $d \in \Sigma_\infty$, and P is a probability function, we have

$$A(d) = \bigcap_{n=1}^{\infty} (A(d_1, \ldots, d_n)),$$

and by Lemma 2.2.2,

$$\begin{aligned} P(A(d)) &= \lim_{n\to\infty} P(A(d_1, \ldots, d_n)), \\ &= \widetilde{P}(A(d)). \end{aligned}$$

Otherwise, if P is just a probability and $A(d) \in \mathcal{B}$, we have

$$\begin{aligned}\widetilde{P}(A(d)) &= \lim_{n\to\infty} P(A(d_1,\ldots,d_n)) \\ &\geq \lim_{n\to\infty} P(A(d)) \\ &= P(A(d)).\end{aligned}$$

The following Lemma will be used to derive identities concerning the probability of certain events in stationary zero-one stochastic processes. This Lemma is a generalization of a result in Kasteleyn [18].

Lemma 4.2.1. *Let* $a_m \geq 0$ *for all* $m \in \mathbb{N}$*, and let* $a_0 = 0$*. Assume that* $0 = a_0 \leq a_1 \leq a_2 \leq a_3 \ldots$*. Let there be* $c_0 \geq c_1 \geq c_2 \geq \ldots \geq 0$ *and define* b_m *by*

$$b_m = c_{m-1} - c_m, \text{ for all } m = 1, 2, \ldots.$$

Then,

$$\sum_{m=1}^{\infty} a_m b_m = \sum_{m=1}^{\infty} (a_m - a_{m-1})(c_{m-1} - \lim_{s\longrightarrow\infty} c_s). \tag{4.4}$$

Proof. If $s \in \mathbb{N}$ and $q \in \mathbb{Z}_+$ with $s > q$, then

$$\begin{aligned}\sum_{m=q+1}^{s} a_m b_m &= \sum_{m=q+1}^{s} a_m(c_{m-1} - c_m) \\ &= \sum_{m=q+1}^{s} a_m c_{m-1} - \sum_{m=q+1}^{s} a_m c_m \\ &= \sum_{m=q+1}^{s} (a_{m-1}c_{m-1} - a_m c_m) - \sum_{m=q+1}^{s} (a_{m-1} - a_m)c_{m-1} \\ &= (a_q c_q - a_s c_s) - \sum_{m=q+1}^{s} (a_{m-1} - a_m)c_{m-1}. \end{aligned} \tag{4.5}$$

Now, in the special case where $q = 0$, as $a_0 = 0$, equation (4.5) becomes

$$\begin{aligned}\sum_{m=1}^{s} a_m b_m &= -a_s c_s - \sum_{m=1}^{s} (a_{m-1} - a_m)c_{m-1} \\ &= \sum_{m=1}^{s} (a_{m-1} - a_m)c_s - \sum_{m=1}^{s} (a_{m-1} - a_m)c_{m-1} \\ &= \sum_{m=1}^{s} (a_m - a_{m-1})(c_{m-1} - c_s). \end{aligned} \tag{4.6}$$

Then, for $1 \leq q < s$, we have

$$\begin{aligned}
\sum_{m=1}^{\infty} a_m b_m &= \sum_{m=1}^{q} a_m b_m + \lim_{s \longrightarrow \infty} \sum_{m=q+1}^{s} a_m b_m \\
&\geq \sum_{m=1}^{q} a_m b_m + \lim_{s \longrightarrow \infty} \sum_{m=q+1}^{s} a_{q+1} b_m \\
&= \sum_{m=1}^{q} a_m (c_{m-1} - c_m) + \lim_{s \longrightarrow \infty} \sum_{m=q+1}^{s} a_{q+1} b_m \\
&= \sum_{m=1}^{q} (a_{m-1} c_{m-1} - a_m c_m) + \sum_{m=1}^{q} (a_m - a_{m-1}) c_{m-1} + \lim_{s \longrightarrow \infty} \sum_{m=q+1}^{s} a_{q+1} b_m \\
&= -a_q c_q - \sum_{m=1}^{q} (a_{m-1} - a_m) c_{m-1} + a_{q+1} \lim_{s \longrightarrow \infty} \sum_{m=q+1}^{s} (c_{m-1} - c_m) \\
&= -a_q c_q + a_{q+1} c_q - a_{q+1} \lim_{s \longrightarrow \infty} c_s - \sum_{m=1}^{q} (a_{m-1} - a_m) c_{m-1} \\
&= -a_{q+1} \lim_{s \longrightarrow \infty} c_s - \sum_{m=1}^{q+1} (a_{m-1} - a_m) c_{m-1} \\
&= \Big(\sum_{m=1}^{q+1} (a_{m-1} - a_m) \Big) \lim_{s \longrightarrow \infty} c_s - \sum_{m=1}^{q+1} (a_{m-1} - a_m) c_{m-1} \\
&= \sum_{m=1}^{q+1} (a_m - a_{m-1})(c_{m-1} - \lim_{s \longrightarrow \infty} c_s).
\end{aligned}$$

Letting $q \longrightarrow \infty$ gives

$$\sum_{m=1}^{\infty} a_m b_m \geq \sum_{m=1}^{\infty} (a_m - a_{m-1})(c_{m-1} - \lim_{s \longrightarrow \infty} c_s). \tag{4.7}$$

On the other hand, we can use equation (4.6). Then, we have

$$\begin{aligned}
\sum_{m=1}^{\infty} a_m b_m &= \lim_{s \longrightarrow \infty} \sum_{m=1}^{s} a_m b_m \\
&= \lim_{s \longrightarrow \infty} \sum_{m=1}^{s} (a_m - a_{m-1})(c_{m-1} - c_s) \\
&\leq \lim_{s \longrightarrow \infty} \sum_{m=1}^{s} (a_m - a_{m-1})(c_{m-1} - \lim_{k \longrightarrow \infty} c_k) \\
&= \sum_{m=1}^{\infty} (a_m - a_{m-1})(c_{m-1} - \lim_{s \longrightarrow \infty} c_s). \qquad (4.8)
\end{aligned}$$

Now, if we compare equation (4.7) and equation (4.8) we have

$$\sum_{m=1}^{\infty} a_m b_m = \sum_{m=1}^{\infty} (a_m - a_{m-1})(c_{m-1} - \lim_{s \longrightarrow \infty} c_s).$$

□

Definition 4.2.1.

Let (x_n) be a sequence of numbers. The sequence (Δx_n) is given by

$$\Delta x_1 = x_1, \text{ and } \Delta x_n = x_{n+1} - x_n, \text{ for } n \geq 1.$$

The sequence Δx_n may be written as (Δx_n).

4.3 The main identities

We now apply Lemma 4.2.1 to stochastic processes.

Lemma 4.3.1. *Let $(S, \mathcal{B}, (X_n), P)$ be a zero-one stochastic process as in Section 3.2. Assume that $\widetilde{P}(10^\infty) = 0$. Then,*

$$\sum_{m=1}^{\infty} mP(10^{m-1}1) = \sum_{m=1}^{\infty} P(0^{m-1}1). \qquad (4.9)$$

Also, if the process is stationary, then $\widetilde{P}(10^\infty) = 0$.

Proof. We take $a_m = m$, for $m = 0, 1, 2, \ldots$, $b_m = P(10^{m-1}1)$, for $m = 1, 2, \ldots$, and $c_m = P(10^m)$, for $m = 0, 1, 2, \ldots$

$$b_m + c_m = P(10^{m-1}1) + P(10^{m-1}0) = P(10^{m-1}) = c_{m-1}.$$

Hence, $c_m \leq c_{m-1}$ and $b_m = c_{m-1} - c_m$, for $m = 1, 2, \ldots$. Also, we have $a_0 = 0, c_0 = P(1)$ and

$$\lim_{m \longrightarrow \infty} c_m = \lim_{m \longrightarrow \infty} P(10^m) = \widetilde{P}(10^\infty) = 0.$$

Now, by using equation (4.4) of Lemma 4.2.1 with $\widetilde{P}(10^\infty) = 0$, we have

$$\sum_{m=1}^{\infty} mP(10^{m-1}1) = \sum_{m=1}^{\infty} P(10^{m-1}).$$

Finally, if the process is stationary,

$$P(10^{m-1}) + P(00^{m-1}) = P(0^{m-1}),$$

so if we let $m \longrightarrow \infty$ we have

$$\widetilde{P}(10^\infty) + \widetilde{P}(0^\infty) = \widetilde{P}(0^\infty),$$

which gives $\widetilde{P}(10^\infty) = 0$.

□

Theorem 4.3.1. *Let $(S, \mathcal{B}, (X_n), P)$ be a zero-one stationary stochastic process with $y \in \Sigma$. Then, the following hold:*

(i) $P(0y) + P(1y) = P(y0) + P(y1) = P(y)$.

(ii) $P(0^m1) = P(10^m)$.

(iii) $\widetilde{P}(y0^\infty) = 0$ *if $y \in \Sigma$ and contains at least a single one.*

(iv) $\lim_{m \to \infty} P(0^m y) = 0$ *if y contains at least a single one.*

(v) $\sum_{m=0}^{\infty} P(0^m1) + \widetilde{P}(0^\infty) = 1$.

(vi) $\sum_{m=1}^{\infty} m^r P(y0^{m-1}1) = \sum_{m=0}^{\infty} \Delta(m^r) \Big[P(y0^m) - \widetilde{P}(y0^\infty) \Big]$, *for $r \in \mathbb{N}$.*

(vii) $\sum_{m=1}^{\infty} m^r P(10^{m-1}y) = \sum_{m=0}^{\infty} \Delta(m^r) \Big[P(0^m y) - \lim_{s \longrightarrow \infty} P(0^s y) \Big]$, *for $r \in \mathbb{N}$.*

Proof. Proof of (i). The equation $P(0y)+P(1y) = P(y)$ follows immediately from stationarity. Also,

$$\begin{aligned} P(y0) + P(y1) &= P\big(A(y0)\big) + P\big(A(y1)\big) \\ &= P\big(A(y)\big) \\ &= P(y), \end{aligned}$$

as $A(y) = A(y0) \cup A(y1)$, and this is a disjoint union in $\mathcal{B}$.

Proof of property (ii). Note that

$$P(0^m) = P(0^m1) + P(0^{m+1}).$$

That is

$$P(0^m1) = P(0^m) - P(0^{m+1}).$$

Also, by using stationarity (see Definition 3.3.1),

$$P(10^m) = P(0^m) - P(0^{m+1}).$$

Thus, the both sides of statement (ii) are equal to $P(0^m) - P(0^{m+1})$.

Proof of property (iii). We may assume that y is of the form $(y_1, y_2, \ldots, y_{\ell-1}, 1)$, for some $\ell \in \mathbb{N}$. Note that

$$\begin{aligned} 0 &\leq \widetilde{P}(y0^\infty) \\ &= \lim_{k\longrightarrow\infty} P(y0^k) \\ &= \lim_{k\longrightarrow\infty} P(A(y0^k)) \\ &= \lim_{k\longrightarrow\infty} P\Big(\{x : X_0(x) = y_1, \ldots, X_{\ell-1} = y_\ell, X_\ell = 1, X_{\ell+1} = 0, \ldots, X_{\ell+k} = 0\}\Big) \\ &\leq \lim_{k\longrightarrow\infty} P\Big(\{x : X_\ell = 1, X_{\ell+1} = 0, \ldots, X_{\ell+k} = 0\}\Big), \\ &\quad \text{by using stationarity,} \\ &= \lim_{k\longrightarrow\infty} P(10^k) \\ &= \widetilde{P}(10^\infty). \end{aligned} \tag{4.10}$$

Now, by using Lemma 4.3.1, we obtain $\widetilde{P}(10^\infty) = 0$. Thus, $\widetilde{P}(y0^\infty) = 0$ by using (4.10).

Proof of property (iv). Note that

$$\lim_{m\to\infty} P(0^m y) \geq 0.$$

Let y have at least one 1. Then, put $y = (0^r, d)$, where $d = (1, d_2, d_3, \ldots, d_\ell)$ for some $r, \ell \in \mathbb{Z}_+$ and $d \in \Sigma$. Then, for $m \in \mathbb{Z}_+$

$$P(0^m y) \leq P(0^{r+m}1) = P(0^{r+m}) - P(0^{r+m+1}).$$

However, $P(0^{r+m}) - P(0^{r+m+1}) \longrightarrow 0$ as $m \longrightarrow \infty$. Hence, $\lim\limits_{m\to\infty} P(0^m y) = 0$.

Proof of property (v). We will use mathematical induction on k to prove

$$\sum_{m=0}^{k} P(0^m 1) + P(0^{k+1}) = 1. \tag{4.11}$$

For $k = 0$ equation (4.11) is

$$P(1) + P(0) = 1,$$

which is true. Also when $k = 1$ equation (4.11) is

$$P(1) + P(01) + P(00) = 1.$$

Since $P(01) + P(00) = P(0)$, (4.11) is true for $k = 1$.

We suppose the result is true for k so

$$\sum_{m=0}^{k} P(0^m 1) + P(0^{k+1}) = 1.$$

Now we will show it is true for $k + 1$. We want to show

$$\sum_{m=0}^{k+1} P(0^m 1) + P(0^{k+2}) = 1. \tag{4.12}$$

We have

$$\sum_{m=0}^{k+1} P(0^m 1) = \sum_{m=0}^{k} P(0^m 1) + P(0^{k+1}1) = 1 - P(0^{k+1}) + P(0^{k+1}1), \tag{4.13}$$

Also we have

$$P(0^{k+1}) = P(0^{k+1}0) + P(0^{k+1}1) = P(0^{k+1}1) + P(0^{k+2}). \tag{4.14}$$

If we compare the result of equation (4.13) and (4.14) with equation (4.12), then the statement holds for $k+1$.

Now letting $k \longrightarrow \infty$ in equation (4.11) we get

$$\lim_{k\longrightarrow\infty} \sum_{m=0}^{k} P(0^m 1) + P(0^{k+1}) = 1.$$

Now, $\lim_{k\longrightarrow\infty} P(0^{k+1}) = \widetilde{P}(0^\infty)$ by definition, so we deduce that

$$\sum_{m=0}^{\infty} P(0^m 1) + \widetilde{P}(0^\infty) = 1.$$

Proof of property (vi). We want now to use equation (4.4) of Lemma 4.2.1 to prove it. We take $a_m = m^r, b_m = P(y0^{m-1}1)$, then

$$P(y0^{m-1}1) + P(y0^m) = P(y0^{m-1}).$$

So if we take $c_m = P(y0^m)$, we have $b_m = c_{m-1} - c_m$. Also, $c_0 = P(y)$, and $a_0 = 0$. Then equation (4.4) of Lemma 4.2.1 gives

$$\begin{aligned}
\sum_{m=1}^{\infty} m^r P(y0^{m-1}1) &= \sum_{m=1}^{\infty} \Big(m^r - (m-1)^r\Big)\Big[P(y0^{m-1}) - \lim_{s\longrightarrow\infty} P(y0^s)\Big] \\
&= \sum_{m=0}^{\infty} \Big((m+1)^r - m^r\Big)\Big[P(y0^m) - \lim_{s\longrightarrow\infty} P(y0^s)\Big] \\
&= \sum_{m=0}^{\infty} \Delta(m^r)\Big[P(y0^m) - \widetilde{P}(y0^\infty)\Big].
\end{aligned}$$

Proof of property (vii). We will use a similar technique of the derivation for property (vi). Taking $a_m = m^r, b_m = P(10^{m-1}y)$, then by using stationarity

$$P(10^{m-1}y) + P(0^m y) = P(0^{m-1}y).$$

So, if $c_m = P(0^m y)$, we have $b_m = c_{m-1} - c_m$. Now, using equation (4.4) of Lemma 4.2.1 with $c_0 = P(y)$ and $a_0 = 0$, we will get

$$\begin{aligned}\sum_{m=1}^{\infty} m^r P(10^{m-1}y) &= \sum_{m=1}^{\infty} \Big(m^r - (m-1)^r\Big)\Big[P(0^{m-1}y) - \lim_{s\longrightarrow\infty} P(0^s y)\Big] \\ &= \sum_{m=0}^{\infty} \Big((m+1)^r - (m)^r\Big)\Big[P(0^m y) - \lim_{s\longrightarrow\infty} P(0^s y)\Big] \\ &= \sum_{m=0}^{\infty} \Delta(m^r)\Big[P(0^m y) - \lim_{s\longrightarrow\infty} P(0^s y)\Big].\end{aligned}$$

□

Chapter 5

Recurrence times in stochastic processes

5.1 Introduction: the notion of recurrence

This chapter introduces the idea of recurrence for stochastic processes. Some standard results are proved including Poincaré's Recurrence Theorem. We also relate the ideas to dynamical system. Some general references which discuss the topic are [8], [7], [12], [19], [24], and [28].

Let S be a set and for each $t = 0, 1, \ldots$ let $X_t : S \longrightarrow \{0, 1\}$. We interpret t as time and call $X_t(x)$ the *state* of the process at time t, given $x \in S$. Points of S can be thought of as "initial states" of the "system"$(S, (X_t))$. If $x \in S$, the (first) arrival time is "how long" it takes to have $X_t(x) = 1$. The (first) recurrence time is "how long" it takes to have $X_t(x) = 1$ for a second time. We make the following definitions. For $x \in S$,

$$a_1(x) = \min\{t : t \geq 0, X_t(x) = 1\},$$

$$a_2(x) = \min\{t : t > a_1(x) \text{ and } X_t(x) = 1\},$$

and so on. In general

$$a_r(x) = \min\{t : t > a_{r-1}(x) \text{ and } X_t(x) = 1\}.$$

Note that if the minimum does not exist, we put $a_r(x) = \infty$. We will define $a_1(x)$ to be the *first arrival time* and $a_2(x)$ to be the *second arrival time* and so on. That is, for $r = 1, 2, \ldots$ $a_r(x)$ is the r^{th} arrival time for x. Also, the differences between successive arrival times, given a state x, are called recurrence times. That is, if $x \in S$, for $k = 1, 2, \ldots$ we define r_k

$$r_k(x) = a_{k+1}(x) - a_k(x).$$

Then $r_k(x)$ is called the k^{th} *recurrence time* for x if it is finite. Otherwise, we put $r_k(x) = \infty$. Also, we put $a_0(x) = 0$ and $r_0(x) = a_1(x)$.

Example 5.1.1.

Let $x \in S$ and suppose that $X_0(x) = \cdots = X_{11}(x) = 0$, $X_{12}(x) = 1$. Let $X_{13}(x) = \cdots = X_{22}(x) = 0$, $X_{23}(x) = 1$. Then, the first arrival time is $a_1(x) = 12$, and $a_2(x) = 23$ is the second arrival time. The first recurrence time is $r_1(x) = a_2(x) - a_1(x) = 11$.

5.2 Poincaré recurrence

Let S be a set and let $\mathcal{B}$ be an algebra of subsets of S. Let P be a probability on $\mathcal{B}$ and let $f : S \longrightarrow S$, so that (S, f) is a dynamical system. We assume that $f^{-1}(A) \in \mathcal{B}$ for all $A \in \mathcal{B}$. Then, if $U \subseteq S$, it was pointed out in Chapter 3 there is an associated stochastic process $(S, \mathcal{B}, (X_n), P)$, where $X_n(x) = \chi_{f^{-n}(U)}(x)$. The notions of arrival and recurrence in $(S, (X_n))$ can therefore be interpreted in terms of the dynamical systems (S, f). According to Poincaré's Recurrence Theorem [13, page 8], [6, page 13], [23, page 34], [20, pages 248-255], in a dynamical system under certain conditions, "almost all" points x of a set will return to the set, or recur, in the sense that $f^n(x)$ will be in the set for some $n \in \mathbb{N}$. Some of these points may take a short time to return but other points may take a long time to return.

Now, let $(S, \mathcal{B}, (X_n), P)$ be a zero-one stochastic process. If we have $c = (c_1, \dots, c_r)$, $d = (d_1, \dots, d_t) \in \Sigma$, then $(c_1, \dots, c_r, d_1, \dots, d_t) \in \Sigma$. For any $r \in \mathbb{N}$,

$$P(0^r1) = P(0^r) - P(0^r0) = P(0^r) - P(0^{r+1}), \text{by using (3.3).}$$

Also, if the process is stationary,

$$P(10^r) = P(0^r) - P(00^r) = P(0^r) - P(0^{r+1}).$$

The following Lemma is a consequence of (iii) of Theorem 4.3.1, but is stated here as the context is different.

Lemma 5.2.1. *Let $(S, \mathcal{B}, (X_n), P)$ be a stationary zero-one stochastic process. If we have $d = (d', 1, 0^k)$ for some $d' = (d'_1, \dots, d'_\ell) \in \Sigma$ and $k \in \mathbb{Z}_+$, then $\widetilde{P}(d, 0^\infty) = 0$.*

The following Lemma generalizes of (v) of Theorem 4.3.1.

Lemma 5.2.2. *Let* $(S,\, \mathcal{B},\, (X_n), P)$ *be a zero-one stochastic process and let* $d \in \Sigma$ *be a finite sequence of zeros and ones. Then, for all* $n = 1, 2, \ldots,$ *we have*

$$P(d) = \sum_{k=0}^{n-1} P(d, 0^k, 1) + P(d, 0^n).$$

Also,

$$P(d) = \sum_{k=0}^{\infty} P(d, 0^k, 1) + \widetilde{P}(d, 0^\infty),$$

and if d has at least one non-zero entry and the process is stationary, then

$$P(d) = \sum_{k=0}^{\infty} P(d, 0^k, 1), \text{ and in particular } P(1) = \sum_{k=0}^{\infty} P(1, 0^k, 1). \tag{5.1}$$

Proof. We will use mathematical induction on n to prove the first statement of the Lemma but before we start to prove it, note that we interpret the expression $P(d, 0^0, 1)$ to be $P(d, 1)$. For $n = 1$, we have by using equation (3.3).

$$P(d) = P(d, 1) + P(d, 0),$$

so the statement is true for $n = 1$.

Now we assume the statement is true for some $n \in \mathbb{N}$. Thus,

$$P(d) = \sum_{k=0}^{n-1} P(d, 0^k, 1) + P(d, 0^n). \tag{5.2}$$

Now we will show the statement is true for $n + 1$. Now,

$$\begin{aligned}
& \sum_{k=0}^{n} P(d, 0^k, 1) + P(d, 0^{n+1}) \\
= & \sum_{k=0}^{n-1} P(d, 0^k, 1) + P(d, 0^n, 1) + P(d, 0^{n+1}), \\
= & \sum_{k=0}^{n-1} P(d, 0^k, 1) + P(d, 0^n), \text{ since } P(d, 0^n) = P(d, 0^n, 1) + P(d, 0^{n+1}), \\
= & \; P(d), \text{ by using equation (5.2).}
\end{aligned}$$

Thus, the statement holds for $n + 1$. By induction, the statement is true for all n.

Now by letting $n \to \infty$ and by using the definition of $\widetilde{P}$, we get

$$P(d) = \sum_{k=0}^{\infty} P(d, 0^k, 1) + \widetilde{P}(d, 0^\infty). \tag{5.3}$$

Also, if d has at least one non-zero entry, and as the process is stationary, we have $\widetilde{P}(d, 0^\infty) = 0$, by Lemma 5.2.1. Thus, in this case,

$$P(d) = \sum_{k=0}^{\infty} P(d, 0^k, 1).$$

In particular if $d = 1$ in this equation, then we have

$$P(1) = \sum_{k=0}^{\infty} P(1, 0^k, 1).$$

□

Definition 5.2.1.

Let P be a probability on an algebra $\mathcal{B}$ of subsets of S. A subset Z of S is called a *set of measure zero* or a *set of probability zero* if, for all $\varepsilon > 0$, there is a sequence (A_n) of sets in $\mathcal{B}$ such that

$$Z \subseteq \bigcup_{n=1}^{\infty} A_n \text{ and } \sum_{n=1}^{\infty} P(A_n) < \varepsilon.$$

Note that if Z is a set of measure zero, then if does not necessarily mean that $Z \in \mathcal{B}$. A special case is when, for all $\varepsilon > 0$, there is $A_\varepsilon \in \mathcal{B}$ such that

$$Z \subseteq A_\varepsilon \text{ and } P(A_\varepsilon) < \varepsilon.$$

Lemma 5.2.3. *Let $\mathcal{B}$ be an algebra of subsets of a set S and let P be a probability on $\mathcal{B}$. Let (A_n) be a sequence of subsets of S of measure zero. Then $\bigcup_{n=1}^{\infty} A_n$ is also a set of measure zero.*

Proof. Suppose that a sequence $A_1, A_2, A_3, \ldots$ of sets of measure zero is given, and let $\varepsilon > 0$. Then as each A_n is set of measure zero, by the definition above there is a sequence $(B_{nj})_{j=1}^{\infty}$ of sets in $\mathcal{B}$ such that

$$A_n \subseteq \bigcup_{j=1}^{\infty} B_{nj}, \text{ and } \sum_{j=1}^{\infty} P(B_{nj}) < \frac{\varepsilon}{2^n}.$$

Now consider the sequence

$$B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, \ldots, B_{1k}, B_{2k}, B_{kk}, \ldots.$$

Rename the above sequence to be $C_1, C_2, C_3, \ldots$. That is,

$$C_1 = B_{11}, C_2 = B_{12}, C_3 = B_{22}, \ldots,$$

and so on. Then,

$$A_1 \cup A_2 \cup A_3 \cup \ldots \subseteq \bigcup_{n=1}^{\infty} \bigcup_{j=1}^{\infty} B_{nj} = \bigcup_{n=1}^{\infty} C_n.$$

Thus,

$$\sum_{n=1}^{k} P\Big(C_n\Big) = \sum_{n=1}^{\infty} \Big(\sum_{j=1}^{\infty} P(B_{nj}) \Big) < \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon.$$

Therefore, $\bigcup_{n=1}^{\infty} A_n$ is a set of measure zero.

□

Lemma 5.2.4. *Let S be a set, let $\mathcal{B}$ be an algebra of subsets of S, and let P be a probability on $\mathcal{B}$. Let $A \in \mathcal{B}$, let (A_n) be a sequence of disjoint sets in $\mathcal{B}$ and let Z be a subset of S that is disjoint from all the sets A_n such that*

$$A = Z \cup \Big(\bigcup_{n=1}^{\infty} A_n \Big) \text{ and } P(A) = \sum_{n=1}^{\infty} P(A_n).$$

Then, Z has measure zero.

Proof. Note that $P(A) < \infty$ and let $\varepsilon > 0$. Then, as we have $P(A) = \sum_{n=1}^{\infty} P(A_n)$, there is $k \in \mathbb{N}$ such that

$$P(A) - \sum_{n=1}^{k} P(A_n) < \varepsilon.$$

Note now that

$$Z \subseteq A - \bigcup_{n=1}^{k} A_n.$$

Clearly,

$$A - \bigcup_{n=1}^{k} A_n \in \mathcal{B}.$$

So,

$$Z \subseteq A - \bigcup_{n=1}^{k} A_n \in \mathcal{B}.$$

Now by using additivity for finite disjoint unions, we have

$$P\Big(A - \bigcup_{n=1}^{k} A_n\Big) = P(A) - \sum_{n=1}^{k} P(A_n) < \varepsilon.$$

Thus, Z has measure zero, by the definition 5.2.1. □

Theorem 5.2.1. *Let $\mathcal{B}$ be an algebra of subsets of a set S, let P be a probability on $\mathcal{B}$, and for $n = 0, 1, \ldots,$ let $X_n : S \to \{0, 1\}$. Assume that $(S, \mathcal{B}\ , (X_n), P)$ is a stationary process, and assume $d = (d_1, d_2, \ldots, d_r) \in \Sigma$. Then, there is subset Z of S that has measure zero such that for all $x \notin Z$ we have: if*

$$X_0(x) = 1, X_1(x) = d_1, \ldots, X_r(x) = d_r,$$

then there are $n \in \mathbb{N}$ and $r \geq 0$ such that $n > r$ and $X_n(x) = 1$.

Proof. By using (5.1) in Lemma 5.2.2 and stationarity

$$P(1, d) = \sum_{k=0}^{\infty} P(1, d, 0^k, 1).$$

That is,

$$P\Big(A(1, d)\Big) = \sum_{k=0}^{\infty} P\Big(A(1, d, 0^k, 1)\Big). \tag{5.4}$$

If we put $A_j = A(1, d, 0^j, 1)$ then, we define Z by

$$Z = A(1, d) \cap \Big(\bigcup_{j=1}^{\infty} A_j\Big)^c.$$

So, noting that $A_j \subseteq A(1, d)$ for all j, we have

$$A(1, d) = Z \cup \Big(\bigcup_{j=1}^{\infty} A_j\Big),$$

and all the sets Z, A_j are mutually disjoint. Then, by applying Lemma 5.2.4, and using (5.4) we see that Z has measure zero. If $x \in A(1, d)$ but $x \notin Z$, $x \in \bigcup_{j=1}^{\infty} A_j$ so there is $j \in \mathbb{N}$ such that

$$x \in A(1, d, 0^j, 1).$$

Therefore,

$X_0(x) = 1, X_1(x) = d_1, \ldots, X_r(x) = d_r, X_{r+1}(x) = 0, \ldots, X_{r+j}(x) = 0$ and $X_{r+j+1}(x) = 1$.

Now if we put $n = j + r + 1$, then $n > r$ and $X_n(x) = 1$. □

The following theorem follows immediately from Theorem 5.2.1. This is the Poincaré Recurrence Theorem discussed at the start of Section 5.2 (see [13, page 8], [6, page 13], [23, page 34], [20, pages 248-255], [15, pages10-12] and [16]).

Theorem 5.2.2. *Let S be a set, let $\mathcal{B}$ be an algebra of subsets of S, let P be a probability on $\mathcal{B}$, and for each $n = 0, 1, \ldots,$ let $X_n : S \to \{0, 1\}$. Assume that $(S, \mathcal{B}, (X_n), P)$ is a stationary zero-one process. Then, there is a subset Z of S that has measure zero such that: if $x \in S$, $x \notin Z$ and $X_0(x) = 1$, there is $r > 0$ such that $X_r(x) = 1$.*

5.3 Recurrence times in dynamical systems

We return to the discussion of dynamical systems, recalling that if S is a set and f is a transformation on S then (S, f) is called a dynamical system. We saw in Section 3.4 that in some circumstances there are zero-one stochastic processes that can be associated with the dynamical system. We now consider how the notion of the arrival and recurrence times as they occur in a stochastic process can be interpreted in a dynamical system. Let S be a set, let $\mathcal{B}$ be an algebra of subsets of S and let f be a transformation on S. We suppose $U \subseteq S$ and we put

$$X_n(x) = \chi_U(f^n(x)), \text{ where } x \in S \text{ and } n = 0, 1, \ldots.$$

Thus,

$$X_n(x) = 1 \iff f^n(x) \in U.$$

Assume that $f^{-n}(A) \in \mathcal{B}$ for all $A \in \mathcal{B}$ and assume further that the set U above is in $\mathcal{B}$. We assume there is a probability P on $\mathcal{B}$. Then, $(S, \mathcal{B}, (X_n), P)$ is the stochastic process associated with the dynamical system (S, f), given $U \in \mathcal{B}$. We put

$$U_0 = U,$$

and

$$U_n = \{x : x \in U, f(x) \notin U, \ldots, f^{n-1}(x) \notin U, f^n(x) \in U \text{ for } n = 1, 2, \ldots\}.$$

That is,

$$U_0 = U \text{ and } U_n = U \cap f^{-1}(U^c) \cap f^{-2}(U^c) \ldots \cap f^{-n+1}(U^c) \cap f^{-n}(U) \in \mathcal{B}.$$

Observe that if $d = (d_1, d_2, \ldots, d_n)$,

$$\begin{aligned} P(d) =& P(A(d)) \\ =& P\Big(\{x : X_0(x) = d_1, X_1(x) = d_2, \ldots, X_{n-1}(x) = d_n\}\Big) \\ =& P\Big(\{x : \chi_U(f^j(x)) = d_{j+1}, j = 0, 1, 2, \ldots, n-1\}\Big). \end{aligned}$$

Thus,

$$\begin{aligned} P(U_n) =& P(10^{n-1}1) \\ =& P\Big(\{x : x \in U, f^j(x) \notin U, \text{ for } j = 1, 2, \ldots, n-1, f^n(x) \in U\}\Big) \\ =& P\Big(U \cap f^{-1}(U^c) \cap \ldots \cap f^{-n+1}(U^c) \cap f^{-n}(U)\Big). \end{aligned}$$

Also,

$$\begin{aligned} P(0^n) =& P\Big(\{x : x \notin U, f^j(x) \notin U, \text{ for } j = 1, 2, \ldots, n-1\}\Big) \\ =& P\Big(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{n-1}(U^c)\Big). \end{aligned}$$

Now the arrival times for the process $(S, (X_n))$ are denoted as in Section 5.1 by $a_n(x)$ for $n = 0, 1, 2, \ldots$. We see that for $x \in S$,

$$a_1(x) = \min\{n : n = 0, 1, 2, \ldots \text{ and } f^n(x) \in U\}.$$

Also, the recurrence times for the process $(S, (X_n))$ are denoted by $r_n(x)$ for $n = 1, 2, \ldots$, as in Section 5.2. We see that given $x \in U$,

$$r_1(x) = \min\{n : n \in \mathbb{N} \text{ and } f^n(x) \in U\}.$$

Thus, in the dynamical system (S, f), if $x \notin U$ the first arrival time is the first or minimum value $m \geq 1$ such that $f^m(x) \in U$. If $x \in U$, then the first arrival time is 0. The first recurrence time is the first or minimum value of m such that $f^m(x) \in U$, given that $x \in U$. Similar comments apply to the higher order arrival and recurrence times. The k^{th} arrival time $a_k(x)$ for x is the k^{th} value m such that $f^m(x) \in U$; and the k^{th} recurrence time for x is $a_k(x) - a_{k-1}(x)$.

5.4 The average of the recurrence time

We saw in Section 5.2 that "almost all" points x of a set will return to the set, or recur, in a dynamical system under certain conditions. Theorem 5.2.2 says that similar things happen for stationary stochastic process. However, we did not discuss how long the point takes to return to some state. In this section, we will examine *the average recurrence time* where $(S, \mathcal{B}\ , (X_n), P)$ is a stochastic process and note that we assume $P(A(1)) > 0$.
Observe that,

$$A(1) = \{x : x \in S \text{ and } X_0(x) = 1\} \in \mathcal{B}.$$

As described in Section 5.1 the first recurrence time is given by, for $x \in A(1)$,

$$r_1(x) = \min\{k : k \in \mathbb{N} \text{ and } X_k(x) = 1\}.$$

We now define the the average of the first recurrence time $\langle r_1 | X_0 = 1\rangle$ (sometimes called the *expectation*) of r_1 over $A(1)$ as

$$\langle r_1 | X_0 = 1\rangle = \frac{1}{P(A(1))} \sum_{m=1}^{\infty} mP(\{x : x \in A(1) \text{ and } r_1(x) = m\}). \tag{5.5}$$

Note that the set $\{x : x \in A(1) \text{ and } r_1(x) = m\}$ equals the set $A(10^{m-1}1)$ which is in $\mathcal{B}$.

Theorem 5.4.1. *Let S be a set, let $\mathcal{B}$ be an algebra of subsets of S and let P be a probability on $\mathcal{B}$. Let $(S, \mathcal{B}, (X_n), P)$ be a stationary zero-one stochastic process, suppose that $\widetilde{P}(0^\infty) = 0$ and $P(A(1)) = q > 0$ where*

$$A(1) = \{x : x \in S \text{ and } X_0(x) = 1\} \in \mathcal{B}.$$

Then, the average of the first recurrence time over $A(1)$ is

$$\langle r_1 | X_0 = 1\rangle = \frac{1}{q}\ .$$

Proof. We will use Theorem 4.3.1, together with $\widetilde{P}(0^\infty) = 0$, to prove the result.

We have by (5.5),

$$\langle r_1 | X_0 = 1\rangle = \frac{1}{q} \sum_{m=1}^{\infty} mP(10^{m-1}1),$$

and by using (vii) of Theorem 4.3.1 with $r = 1$ and $y = 1$ we get

$$\langle r_1 | X_0 = 1\rangle = \frac{1}{q} \sum_{m=1}^{\infty} \Big[P(0^m 1) - \lim_{N \longrightarrow \infty} P(0^N 1)\Big]. \tag{5.6}$$

Now, if we use (iv) of Theorem 4.3.1 we get $\lim_{N\longrightarrow\infty} P(0^N1) = 0$, so (5.6) gives

$$\langle r_1 | X_0 = 1 \rangle = \frac{1}{q} \sum_{m=1}^{\infty} P(0^m 1).$$

However, using (v) of Theorem 4.3.1 now gives $\sum_{m=1}^{\infty} P(0^m 1) = 1$, so we deduce that

$$\langle r_1 | X_0 = 1 \rangle = \frac{1}{q},$$

as required. □

5.5 The dynamical systems formulation

In this Section we take the notion and result of the previous Section to derive the average value of r_1 over U for a dynamical system. Suppose we have a dynamical system (S, f) where S is a set and U is a subset of S. Then, given $x \in U$, the first recurrence time is given by

$$r_1(x) = \min\{n \in \mathbb{N} \text{ and } f^n(x) \in U\}.$$

Note that for the corresponding stochastic process, $A(1) = U$ and using equation (5.5), the average value of r_1 over U equals

$$\frac{1}{P(U)} \sum_{m=0}^{\infty} mP\Big(\{x : x \in U \text{ and } r_1(x) = m\}\Big).$$

To apply Theorem 5.4.1, we need to have a dynamical system (S, f) together with an algebra $\mathcal{B}$ of subsets of S and a probability P on $\mathcal{B}$ for which f is P-invariant. The condition $\widetilde{P}(0^\infty) = 0$ in the Theorem 5.4.1 can be interpreted in this context as

$$\lim_{n\longrightarrow\infty} P(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-n}(U^c)) = 0.$$

Also, the P-invariance of f ensures that the stochastic process associated with (S, f) is stationary, by Theorem 3.4.1. So, we obtain the following result.

Theorem 5.5.1. *Let (S, f) be a dynamical system, let $\mathcal{B}$ be an algebra of subsets of S, and let P be a probability on $\mathcal{B}$ such that f is P-invariant. Let $U \in \mathcal{B}$ with $P(U) > 0$, and assume that*

$$\lim_{n \longrightarrow \infty} P(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-n}(U^c)) = 0.$$

Let r_1 be the first recurrence time of f over U. Then, the average value of r_1 over U is

$$\frac{1}{P(U)} \sum_{m=0}^{\infty} mP\Big(\{x : x \in U \text{ and } r_1(x) = m\}\Big) = \frac{1}{P(U)}.$$

Chapter 6

The standard deviation of the recurrence time

6.1 Introduction

As we have seen in Chapter Five Section 5.4, given a stationary stochastic process $(S, \mathcal{B}, (X_n), P)$ with $\widetilde{P}(0^\infty) = 0$ then the average of the first recurrence time r_1 over $A(1) = \{x : x \in S \text{ and } X_0(x) = 1\}$, with $P(A(1)) > 0$ is

$$\langle r_1 | X_0 = 1 \rangle = \frac{1}{q}, \text{ where } q = P(A(1)). \tag{6.1}$$

The "average scattering" of a function is gauged by the standard deviation of the function. We now define the standard deviation and derive a formula for it in the case of zero-one stationary stochastic processes. This formula is originally due J. R. Blum and J. I. Rosenblatt [4], but was proved also in P. W. Kasteleyn [18] (see also [20, pages 270-282]). The approach taken here differs from that of [4] and [18], in some respects, in that the formula is derived under somewhat weaker and more explicit assumptions.

6.2 The standard deviation formula

We now make some important definitions.

Definition 6.2.1.

Let S be a set, let $\mathcal{B}$ be an algebra of subsets of S, let P be a probability on $\mathcal{B}$, and let $\phi : S \longrightarrow \mathbb{Z}_+ \cup \{\infty\}$ be a function such that for each $m \in \mathbb{Z}_+, \phi^{-1}(m) \in \mathcal{B}$, and $\phi^{-1}(\{\infty\})$ has measure zero. Then, the *average* (or *expectation*) of ϕ over a set $A \in \mathcal{B}$ with $P(A) > 0$ is

$$E_A(\phi) = \frac{1}{P(A)} \sum_{m=0}^{\infty} mP(\phi^{-1}(m)).$$

Then, the *variance* $v_A(\phi)$ of ϕ over the set $A \in \mathcal{B}$ is

$$v_A(\phi) = \frac{1}{P(A)} \sum_{m=0}^{\infty} \Big(m - E_A(\phi)\Big)^2 P(\phi^{-1}(m)),$$

provided that $E_A(\phi) < \infty$. The standard deviation is defined to be

$$\sigma_A(\phi) = \sqrt{v_A(\phi)}.$$

Since we are only going to do the case of r_1 over $A = A(1)$, then perhaps we do not need the general definition.

Note that in our applications to stochastic processes

$$A = \{x : x \in S \text{ and } X_0(x) = 1\} = A(1),$$

and in our applications to dynamical system $A = U$ a given subset of S. For now, we deal with the stochastic processes case.

We assume that $q = P(A) > 0$. Then $r_1 : A \longrightarrow \mathbb{N} \cup \{\infty\}$, and under the condition of Theorem 5.2.1, $P(r_1^{-1}(\{\infty\})) = 0$. Observe that

$$r_1^{-1}(m) = A(10^{m-1}1), \text{ for } m \in \mathbb{N}.$$

Hence, according to the definition, the average of r_1 over $A(1)$ is

$$\begin{aligned} E_{A(1)}(r_1) &= \frac{1}{q} \sum_{m=0}^{\infty} mP(r_1^{-1}(m)) \\ &= \frac{1}{q} \sum_{m=0}^{\infty} mP(10^{m-1}1), \end{aligned}$$

and by Theorem 5.4.1,

$$E_{A(1)}(r_1) = \frac{1}{P(A(1))} = \frac{1}{q}.$$

Now, the *standard deviation* of r_1 over $A(1)$ is

$$\begin{aligned}\sigma_{A(1)}(r_1) &= \sqrt{\frac{1}{q}\sum_{m=1}^{\infty}\Big(m - E_{A(1)}(r_1)\Big)^2 P(10^{m-1}1)} \\ &= \sqrt{\frac{1}{q}\sum_{m=1}^{\infty}\Big(m - q^{-1}\Big)^2 P(10^{m-1}1)} \\ &= \sqrt{\frac{1}{q}\sum_{m=1}^{\infty}\Big(m^2 - 2q^{-1}m + q^{-2}\Big) P(10^{m-1}1)}. \qquad (6.2)\end{aligned}$$

Using (6.2), the aim is to derive the the standard deviation of the first recurrence time. However, before giving a formal discussion for the standard deviation of the first recurrence time, we need to prove Lemma below

Lemma 6.2.1. *Let S be a set, let $\mathcal{B}$ be an algebra of subsets of S and let P be a probability function on $\mathcal{B}$. Let $X \in \mathcal{B}$, let Z be a subset of S of measure zero, and let (Y_n) be a sequence of disjoint sets in $\mathcal{B}$ such that*

$$X = \Big(\bigcup_{j=1}^{\infty} Y_j\Big) \cup Z.$$

Then,

$$P(X) = \sum_{n=1}^{\infty} P(Y_n).$$

Proof. Let $W = X \cap \Big(\bigcup_{n=1}^{\infty} Y_n\Big)^c$. Then, $W \subseteq Z$ and so W has measure zero. Also,

$$X = W \cup \Big(\bigcup_{n=1}^{\infty} Y_n\Big),$$

and this union is disjoint. Then for $\varepsilon > 0$ there is a sequence (J_n) in $\mathcal{B}$ such that $W \subseteq \bigcup_{n=1}^{\infty} J_n$, and $\displaystyle\sum_{n=1}^{\infty} P(J_n) < \varepsilon$. We now have

$$X \subseteq \Big(\bigcup_{n=1}^{\infty} Y_n\Big) \cup \Big(\bigcup_{n=1}^{\infty} J_n\Big) \text{ and } \sum_{n=1}^{\infty} P(J_n) < \varepsilon.$$

Now, by using Lemma 2.2.1,

$$P(X) \leq \sum_{n=1}^{\infty} P(Y_n) + \sum_{n=1}^{\infty} P(J_n) < \sum_{n=1}^{\infty} P(Y_n) + \varepsilon.$$

As this holds for all $\varepsilon > 0$,

$$P(X) \leq \sum_{n=1}^{\infty} P(Y_n). \tag{6.3}$$

Now note that as P is a probability,

$$\begin{aligned} \sum_{n=1}^{r} P(Y_n) &= P\Big(\bigcup_{n=1}^{r} Y_n\Big), \text{ since the sets } Y_n \text{ are disjoint,} \\ &\leq P(X). \end{aligned}$$

Letting $r \longrightarrow \infty$ we obtain

$$\sum_{n=1}^{\infty} P(Y_n) \leq P(X). \tag{6.4}$$

Clearly, from (6.3) and (6.4) we have

$$P(X) = \sum_{n=1}^{\infty} P(Y_n).$$

□

Theorem 6.2.1. *Let $(S, \mathcal{B}, (X_n), P)$ be a stationary stochastic process, suppose that $\widetilde{P}(0^\infty) = 0$ with P a probability function on $\mathcal{B}$. Let $r_1 : A(1) \longrightarrow \mathbb{N}$ be the function which is the first recurrence time. Put $q = P(A(1))$ and assume $q > 0$. Then the standard deviation of r_1 over $A(1)$ is*

$$\sigma_{A(1)}(r_1) = \sqrt{q^{-1} - q^{-2} + 2q^{-1} \sum_{m=1}^{\infty} P(0^m)}.$$

Equivalently,

$$\sigma_{A(1)}(r_1) = \sqrt{3q^{-1} - q^{-2} - 2 + 2q^{-1} \sum_{m=2}^{\infty} P(0^m)}.$$

Consequently, if the serie $\sum_{m=1}^{\infty} P(0^m)$ is divergent, $\sigma_{A(1)}(r_1)$ is infinite; otherwise it is finite.

Proof. By using (6.2) we have

$$\sigma_{A(1)}(r_1) = \sqrt{\frac{1}{q}\sum_{m=1}^{\infty}\Big(m^2 - 2q^{-1}m + q^{-2}\Big)P(10^{m-1}1)}.$$

The goal now is to calculate the terms under the square root. Observe that

$$\begin{aligned} &\frac{1}{q}\sum_{m=1}^{\infty}\Big(m^2 - 2q^{-1}m + q^{-2}\Big)P(10^{m-1}1) \\ = \; &\frac{1}{q}\Big[\underbrace{\sum_{m=1}^{\infty} m^2 P(10^{m-1}1)}_{\text{first term}} - 2q^{-1}\underbrace{\sum_{m=1}^{\infty} mP(10^{m-1}1)}_{\text{second term}} + q^{-2}\underbrace{\sum_{m=1}^{\infty} P(10^{m-1}1)}_{\text{third term}}\Big]. \end{aligned} \tag{6.5}$$

We will show that the second and third terms are convergent and obtain a simple expression for the first term. For the first term, using Theorem 4.3.1 property (vii) with $y = 1$ and $r = 2$, the first term of (6.5) is

$$\sum_{m=1}^{\infty} m^2 P(10^{m-1}1) = \sum_{m=0}^{\infty}\Big((m+1)^2 - m^2\Big)\Big[P(0^m 1) - \lim_{s\longrightarrow\infty} P(0^s 1)\Big].$$

Now observe that $P(0^m 1) = P(0^m) - P(0^{m+1})$, so taking limits gives

$$\lim_{m\longrightarrow\infty} P(0^m 1) = 0.$$

It follows that

$$\begin{aligned} \sum_{m=1}^{\infty} m^2 P(10^{m-1}1) &= \sum_{m=0}^{\infty}(1+2m)P(0^m 1) \\ &= \sum_{m=0}^{\infty} P(0^m 1) + 2\sum_{m=0}^{\infty} mP(0^m 1). \end{aligned} \tag{6.6}$$

Now by using Theorem 4.3.1 property (vi) with $y = 0$ and $r = 1$ gives

$$\begin{aligned} \sum_{m=0}^{\infty} mP(0^m 1) &= \sum_{m=0}^{\infty}\Big[P(0^{m+1}) - \widetilde{P}(0^{\infty})\Big] \\ &= \sum_{m=1}^{\infty} P(0^m). \end{aligned} \tag{6.7}$$

Also,

$$\begin{aligned}\sum_{m=0}^{\infty} P(0^m 1) &= P(1) + \sum_{m=1}^{\infty}\Big[P(0^m) - P(0^{m+1})\Big] \\ &= P(1) + P(0) - \widetilde{P}(0^\infty) \\ &= P(S) \\ &= 1. \end{aligned} \tag{6.8}$$

We now see from (6.6),(6.7), and (6.8) above that the first term in (6.5) is

$$\sum_{m=1}^{\infty} m^2 P(10^{m-1}1) = 1 + 2\sum_{m=1}^{\infty} P(0^m). \tag{6.9}$$

We now consider the second term. It is immediately from Theorem 5.4.1 that the second term of (6.5) becomes

$$\sum_{m=1}^{\infty} mP(10^{m-1}1) = 1. \tag{6.10}$$

Finally, we show that the third term gives $\sum_{m=1}^{\infty} P(10^{m-1}1) = q$. We consider

$$\sum_{m=1}^{\infty} P(10^{m-1}1) = \sum_{m=0}^{\infty} P(10^m 1).$$

Observe that

$$A(1) = \Big(\bigcup_{m=0}^{\infty} A(10^m 1)\Big) \cup A(10^\infty).$$

We first prove that $A(10^\infty)$ has measure zero. We have

$$P(10^m) = P(0^m) - P(0^{m+1}),$$

and

$$A(10^\infty) \subseteq A(10^m).$$

We know by Theorem 4.3.1 property (iii), $\lim_{m\longrightarrow\infty} P(10^m) = 0$ and as $P(10^m) = P\Big(A(10^m)\Big)$, we see that $A(10^\infty)$ is a set of measure zero.

Now,

$$A(1) = \Big(\bigcup_{m=0}^{\infty} A(10^m 1)\Big) \cup A(10^\infty),$$

so as $A(10^\infty)$ has measure zero, Lemma 6.2.1 gives

$$\begin{aligned} P(A(1)) &= \sum_{m=0}^{\infty} P\Big(A(10^m 1)\Big) \\ &= \sum_{m=0}^{\infty} P(10^m 1). \end{aligned}$$

Therefore, the third term of (6.5) will be

$$\sum_{m=1}^{\infty} P(10^{m-1}1) = P(A(1)) = q. \tag{6.11}$$

Now putting (6.9), (6.10), and (6.11) in (6.5) gives

$$\begin{aligned} \Big[\sigma_A(r_1)\Big]^2 &= \frac{1}{q}\Big[1 + 2\sum_{m=1}^{\infty} P(0^m) - 2q^{-1} + q^{-2}(q)\Big] \\ &= q^{-1} - q^{-2} + 2q^{-1}\sum_{m=1}^{\infty} P(0^m). \end{aligned}$$

Hence, the standard deviation is

$$\sigma_A(r_1) = \sqrt{q^{-1} - q^{-2} + 2q^{-1}\sum_{m=1}^{\infty} P(0^m)}.$$

Note that $P(0) = 1 - P(1) = 1 - q$ so the standard deviation is also equal to

$$\begin{aligned} \sigma_A(r_1) &= \sqrt{q^{-1} - q^{-2} + 2q^{-1}\Big[1 - q + \sum_{m=2}^{\infty} P(0^m)\Big]} \\ &= \sqrt{3q^{-1} - q^{-2} - 2 + 2q^{-1}\sum_{m=2}^{\infty} P(0^m)}. \end{aligned}$$

□

6.3 The standard deviation in dynamical systems

In this section, we will discuss the special form of Theorem 6.2.1 in the case of certain dynamical systems. Let (S, f) be a dynamical system with P a probability function on an algebra $\mathcal{B}$ of subsets of S. Let $U \in \mathcal{B}$ with $P(U) > 0$. We assume that f is P-invariant. Then, if we consider the associated stationary stochastic process, $(S, \mathcal{B}, (X_n), P)$, as in Theorem 3.4.1, the condition $\widetilde{P}(0^\infty) = 0$ is equivalent to having

$$\lim_{n \longrightarrow \infty} P(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-n}(U^c)) = 0.$$

The following Theorem is the interpretation of Theorem 6.2.1 in the dynamical systems context.

Theorem 6.3.1. *Let (S, f) be a dynamical system, let $\mathcal{B}$ be an algebra of subsets of S, and let P be a probability function on $\mathcal{B}$ and assume f is P-invariant. Let $U \in \mathcal{B}$ with $P(U) > 0$. Let r_1 be the first recurrence time of f over U and suppose $\widetilde{P}(0^\infty) = 0$. The standard deviation can be infinite but it is finite if and only if*

$$\sum_{m=1}^{\infty} P\Big(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-m}(U^c)\Big) < \infty.$$

In this case, the standard deviation of r_1 over U is finite and equals
$\sigma_U(r_1) =$

$$\sqrt{-P(U)^{-2} + 3P(U)^{-1} - 2 + 2P(U)^{-1} \textstyle\sum_{m=1}^{\infty} P\Big(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-m}(U^c)\Big)}.$$

Proof. It follows from Theorem 6.2.1 with $\widetilde{P}(0^\infty) = 0$ that if we take $P(U) > 0$ in place of q and under the conditions of Theorem 3.4.1 then,

$$\sigma_U(r_1) = \sqrt{P(U)^{-1} - P(U)^{-2} + 2P(U)^{-1} \sum_{m=1}^{\infty} P(0^m)}.$$

Now observe that

$$\begin{aligned} P(0^m) &= P\Big(\{x : X_1(x) = 0, X_2(x) = 0, \ldots, X_{m-1}(x) = 0\}\Big), \\ &= P\Big(\{x : x \notin U, f(x) \notin U, \ldots, f^{(m-1)}(x) \notin U\}\Big), \\ &= P\Big(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-(m-1)}(U^c)\Big). \end{aligned}$$

That is, $\sigma_U(r_1)$

$$= \sqrt{P(U)^{-1} - P(U)^{-2} + 2P(U)^{-1} \sum_{m=0}^{\infty} P\Big(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-m}(U^c)\Big)}$$

$$= \sqrt{-P(U)^{-2} + 3P(U)^{-1} - 2 + 2P(U)^{-1} \sum_{m=1}^{\infty} P\Big(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-m}(U^c)\Big)}.$$

□

Note that this formula above tells us that the standard deviation depends on the subset U of S.

Chapter 7

Piecewise linear functions and their dynamical systems

7.1 Introduction

A main aim of this chapter is to introduce piecewise linear functions and study their behaviour in dynamical systems. If X is an interval and $f : X \longrightarrow \mathbb{R}$ is a function, then f is a *piecewise linear* function if there is a partition of X into a finite number of subintervals such that, on each interval J in the partition, f is given by $x \longmapsto ax + b$, for suitable constants a and b that generally depend on J.

In this Chapter, for a piecewise linear function of the type in Figure 7.1, some properties are derived which are used to obtain new estimates for the standard deviation of recurrence times in the associated dynamical systems. There is also a related discussion of some examples that are of use in subsequent parts of the thesis (for more discussion and clarification see [20]).

7.2 Piecewise linear functions

Definition 7.2.1.

Let X be a set. A *partition* of X is a family $X_1, X_2, \ldots, X_k$ of non-empty disjoint subsets of X such that

$$X = \bigcup_{j=1}^{k} X_j.$$

Definition 7.2.2.

Let X be an interval. A function $g : J \to \mathbb{R}$, where J is a subinterval of X, is called *linear* if there are $a, b \in \mathbb{R}$ such that $g(x) = ax + b$, for all $x \in J$. A function $f : X \to \mathbb{R}$ is called *piecewise linear* if there is a partition of X into subintervals of positive length such that f is a linear on each of these subintervals.

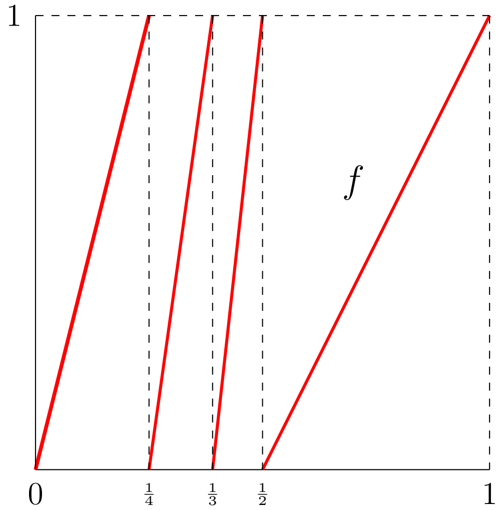

Figure 7.1. This illustrates the graph of a piecewise linear function on $[0, 1)$. The interval $[0, 1)$ is partitioned into the four subintervals $[0, 1/4), [1/4, 1/3), [1/3, 1/2)$, and $[1/2, 1)$. On each of these subintervals, f is linear in the sense that it is given by $x \longmapsto ax + b$, for a suitable choice of a and b. Note that the range of f on each interval in the partition is $[0, 1)$. The transformation f is an example of a completely stretching piecewise linear transformation, see the main text for the definition.

Definition 7.2.3.

Let $f : X \to X$ be a transformation on X where X is an interval of length 1. Now we define a *completely stretching piecewise linear transformation.* If there is $r \in \mathbb{N}$ with $r \geq 2$ and a partition $X_1, X_2, ..., X_r$ of X into r subintervals such that, for each $j \in \{1, 2, ..., r\}$, f is linear on X_j and its range on X_j is either X or X less one or two endpoints.

See Figure 7.2 for different examples of piecewise linear functions. On the right of Figure 7.2, the transformation is completely stretching but on the left of Figure 7.2, it is not.

Note that if f is completely stretching and is not linear on the whole of its domain, then f is not one-to-one.

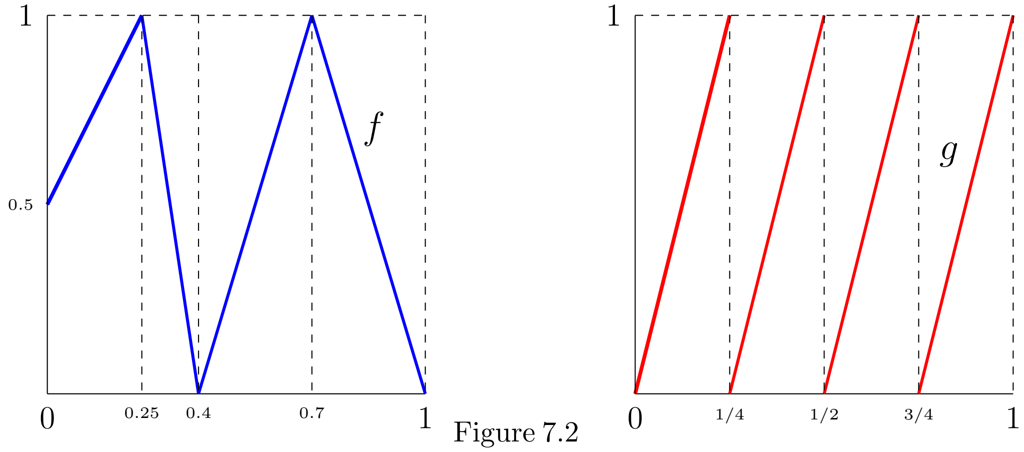

Figure 7.2

Figure 7.2. On the left, the Figure illustrates a continuous piecewise linear function that is linear on each of intervals $[0, 0.25), [0.25, 0.4), [0.4, 0.7)$ and $[0.7, 1]$. Note that $([0, 1], f)$ is a dynamical system. On the right, the Figure illustrates a completely stretching piecewise linear function g on $[0, 1)$. The function g is linear on $[0, 1/4), [1/4, 1/2), [1/2, 3/4)$ and $[3/4, 1)$ and on each of these the restriction of g has range $[0, 1)$. Also, $([0, 1), g)$ is a dynamical system.

Let f be a completely stretching piecewise linear function as in Definition 7.2.3. Let $k \in \mathbb{N}$ be given and for each choice $n_1, n_2, \ldots, n_k$ of k numbers in $\{1, 2, \ldots, r\}$, put

$$X_{n_1 n_2 \ldots n_k} = \{x : x \in X_{n_1}, f(x) \in X_{n_2}, \ldots, f^{k-1}(x) \in X_{n_k}\} = \bigcap_{j=1}^{k} f^{-j+1}(X_{n_j}). \tag{7.1}$$

In the case $k = 1$, (7.1) gives X_{n_1}, and there is no ambiguity. The set $X_{n_1 n_2 \ldots n_k}$ can be thought of as the set of initial states which lead to a prescribed evolution of the system over the first k time units. The requirement that $f^{j-1}(x) \in X_{n_j}$ for all $j = 1, 2, \ldots, k$ means that after the elapse of $j-1$ time units, the system is in a state lying in the prescribed set X_{n_j} for all $j = 1, 2, \ldots, k$.

Lemma 7.2.1. *Let X be an interval of length 1 and let $X_1, X_2, \ldots, X_r$ be a partition of X into r subintervals of positive length. Let $k \in \mathbb{N}$, and let f be a completely stretching piecewise linear transformation on X. Then, the set $X_{n_1 n_2 \ldots n_k}$ is a subinterval of X for all $n_1, n_2, ..., n_k \in \{1, 2, ..., r\}$. Also, the restriction of f^k to $X_{n_1 n_2 \ldots n_k}$ is a linear function whose range is either X, or X less one or two endpoints.*

Proof. The approach of the proof will be by mathematical induction. Let $int(J)$ denote the interior of any interval J, and let $k \in \mathbb{N}$. Consider the following statements:

(i) $X_{n_1 n_2 \ldots n_k}$ is a subinterval of X for all $n_1, n_2, ..., n_k \in \{1, 2, ..., r\}$.
(ii) f^k is linear on $X_{n_1 n_2 \ldots n_k}$ for all $n_1, n_2, ..., n_k \in \{1, 2, ..., r\}$.
(iii) f^k maps $int(X_{n_1 n_2 \ldots n_k})$ onto $int(X)$ for all $n_1, n_2, ..., n_k \in \{1, 2, ..., r\}$.

If f^k is linear on $X_{n_1 n_2 \ldots n_k}$ and maps $int(X_{n_1 n_2 \ldots n_k})$ onto $int(X)$, then f^k maps $X_{n_1 n_2 \ldots n_k}$ onto X or X less one or two endpoints. Hence, if $(i), (ii)$ and (iii) are true for some k, then the Lemma is also true for that value of k. However, $(i), (ii)$ and (iii) are true when $k = 1$, because of the assumed properties of the transformation f. So we proceed by mathematical induction.
Assume that $(i), (ii)$ and (iii) hold for k. Then,

$$\begin{aligned} X_{n_1 n_2 \ldots n_k n_{k+1}} &= \Big\{ x : x \in X_{n_1}, f(x) \in X_{n_2}, \ldots, f^{k-1}(x) \in X_{n_k}, f^k(x) \in X_{n_{k+1}} \Big\} \\ &= \Big\{ x : x \in X_{n_1 n_2 \ldots n_k} \text{ and } f^k(x) \in X_{n_{k+1}} \Big\}. \end{aligned}$$

Hence, $X_{n_1 n_2 \ldots n_k n_{k+1}}$ is an interval because it is assumed that f^k is linear on $X_{n_1 n_2 \ldots n_k}$. That is, (i) holds with $k+1$ in place of k.

Now observe that

$$\begin{aligned} & x \in X_{n_1 n_2 \ldots n_k n_{k+1}} \\ & \Longrightarrow x \in X_{n_1}, f(x) \in X_{n_2}, \ldots, f^{k-1}(x) \in X_{n_k}, f^k(x) \in X_{n_{k+1}} \\ & \Longrightarrow f(x) \in X_{n_2}, f\Big(f(x)\Big) \in X_{n_3}, \ldots, f^{k-2}\Big(f(x)\Big) \in X_{n_k}, f^{k-1}\Big(f(x)\Big) \in X_{n_{k+1}} \\ & \Longrightarrow f(x) \in X_{n_2 n_3 \ldots n_k n_{k+1}}. \end{aligned}$$

Hence, $f : X_{n_1 n_2 \ldots n_k n_{k+1}} \longrightarrow X_{n_2 n_3 \ldots n_k n_{k+1}}$ and f^k is linear on $X_{n_2 n_3 \ldots n_k n_{k+1}}$ by assumption (ii). However, as f is linear on X_{n_1}, f is linear on the smaller set $X_{n_1 n_2 \ldots n_k n_{k+1}}$. So we have

$$f : X_{n_1 n_2 \ldots n_k n_{k+1}} \longrightarrow X_{n_2 n_3 \ldots n_k n_{k+1}} \text{ and } f^k : X_{n_2 n_3 \ldots n_k n_{k+1}} \longrightarrow X,$$

where both f and f^k are linear. Since the composition of linear functions is linear, it follows that $f^{k+1} = f^k \circ f$ is linear on $X_{n_1 n_2 \ldots n_k n_{k+1}}$, thus showing that (ii) holds with $k+1$ in place of k.

Now, let $z \in int(X)$. By the induction assumption (iii), there is $y \in int(X_{n_2 n_3 \ldots n_k n_{k+1}})$ such that $f^k(y) = z$. Then, as f maps X_{n_1} onto either X or X less one or two endpoints, there is $x \in int(X_{n_1})$ such that $f(x) = y$. then,

$$f^{k+1}(x) = \Big(f^k \circ f\Big)(x) = f^k\Big(f(x)\Big) = f^k(y) = z.$$

Therefore, $x \in X_{n_1 n_2 \ldots n_k n_{k+1}}$, and it follows that f^{k+1} maps $int(X_{n_1 n_2 \ldots n_k n_{k+1}})$ onto $int(X)$, so that (iii) holds with $k+1$ in place of k. Thus, the Lemma follows by mathematical induction. □

Now, let X be an interval of length 1 and let $\mathcal{B}$ be the algebra of subsets of X consisting of finite unions of subintervals of X. For each subinterval A of X we define $\mu(A) =$ length of A. If $A \in \mathcal{B}$ and $A = \bigcup_{j=1}^{n} A_j$, where this is a disjoint union of intervals, then if we put $\mu(A) = \sum_{j=1}^{n} \mu(A_j)$, μ is well defined on $\mathcal{B}$ and $\mu(X) = 1$. For future use, note that μ is a length-preserving (see Definition 7.3.1 and [20]). Note that if f is a piecewise linear transformation on X, then $f^{-1}(C) \in \mathcal{B}$ for all $C \in \mathcal{B}$. Note that a completely stretching piecewise linear function is a length-preserving transformation(see [20, pages 245-246]).

Lemma 7.2.2. *Let X be an interval of length 1 and let $X_1, X_2, \ldots, X_r$ be a partition of X into $r \geq 2$ subintervals of positive length. Let f be a completely stretching piecewise linear transformation and $X_{n_1 n_2 \ldots n_k}$ defined as in (7.1). If $k \in \mathbb{N}$, $X_{n_1 n_2 \ldots n_k}$ is a subinterval of X for all $n_1, n_2, ..., n_k \in \{1, 2, ..., r\}$. Then,*

$$\mu(X_{n_1 n_2 \ldots n_k}) = \prod_{j=1}^{k} \mu(X_{n_j}).$$

Furthermore, if $\rho = \max\{\mu(X_j) : 1 \leq j \leq r\}$, then $0 < \rho < 1$ and

$$\mu(X_{n_1 n_2 \ldots n_k}) \leq \rho^k,$$

for all $k \in \mathbb{N}$ and all $n_1, n_2, ..., n_k \in \{1, 2, ..., r\}$.

Proof. We will use mathematical induction to prove the Lemma. Consider the first statement. If $k = 1$, the result is true. Assume the result is true for some k. We use the result of Lemma 7.2.1 that f^k is linear on $X_{n_1 n_2 \ldots n_k}$ and the range of f on $X_{n_1 n_2 \ldots n_k}$ is X, or X less one or two endpoints. Then,

$$\begin{aligned}
\mu(X_{n_1 n_2 \ldots n_k n_{k+1}}) &= \mu\Big(\{x : x \in X_{n_1}, f(x) \in X_{n_2}, \ldots, f^{k-1}(x) \in X_{n_k}, f^k(x) \in X_{n_{k+1}}\}\Big) \\
&= \mu\Big(\{x : x \in X_{n_1 n_2 \ldots n_k} \text{ and } f^k(x) \in X_{n_{k+1}}\}\Big) \\
&= \mu(X_{n_1 n_2 \ldots n_k}).\mu(X_{n_{k+1}}) \\
&= \mu(X_{n_{k+1}}).\prod_{j=1}^{k} \mu(X_{n_j}), \text{ by the inductive assumption,} \\
&= \prod_{j=1}^{k+1} \mu(X_{n_j}).
\end{aligned}$$

Thus, if the result holds for k, it holds also for $k+1$. So the first statement is true for all $k \in \mathbb{N}$ by induction.

Now, as $r \geq 2$, and as $X_1, X_2, \ldots, X_r$ partition X and have positive length, we have $0 < \mu(X_j) < 1$ for all $j \in \{1, 2, \ldots, r\}$. So, $0 < \rho < 1$ and it follows that

$$\mu(X_{n_1 n_2 \ldots n_k}) = \prod_{j=1}^{k} \mu(X_{n_j}) \leq \rho^k, \text{ for all } k \in \mathbb{N} \text{ and } n_1, n_2, ..., n_k \in \{1, 2, ..., r\}.$$

□

Lemma 7.2.3. *Let X be an interval of length 1 and let f be a completely stretching piecewise linear transformation on X. Let $X_1, X_2, \ldots, X_r$ be a partition of X into r subintervals of positive length. Let $k, s \in \mathbb{N}$ and for each $j \in \{1, 2, \ldots, s\}$, suppose there are given $n_{1j}, n_{2j}, ..., n_{kj} \in \{1, 2, \ldots, r\}$ and put $Y_j = X_{n_{1j} n_{2j} \ldots n_{kj}}$. Then,*

$$\mu(Y_1 \cap f^{-k}(Y_2) \cap \ldots \cap f^{-(s-1)k}(Y_s)) = \prod_{j=1}^{s} \mu(Y_j).$$

Proof. We will prove from the definitions of $X_{n_{1j} n_{2j} \ldots n_{kj}}$ and Y_j that

$$Y_1 \cap f^{-k}(Y_2) \cap \ldots \cap f^{-(s-1)k}(Y_s) = X_{n_{11} \ldots n_{k1} n_{12} \ldots n_{k2} n_{13} \ldots n_{ks-1} n_{1s} \ldots n_{ks}}. \quad (7.2)$$

We have

$$\begin{aligned}
&x \in X_{n_{11}\ldots n_{k1}n_{12}\ldots n_{k2}n_{13}\ldots n_{ks-1}n_{1s}\ldots n_{ks}}\\
&\iff x \in X_{n_{11}}, \ldots, f^{k-1}(x) \in X_{n_{k1}}, \ldots, f^{k(s-1)}(x) \in X_{n_{1s}}, f^{sk-1}(x) \in X_{n_{ks}}\\
&\iff x \in X_{n_{11}\ldots n_{k1}}, \ldots, f^{k}(x) \in X_{n_{12}\ldots n_{k2}}, \ldots, f^{k(s-1)}(x) \in X_{n_{1s}\ldots n_{ks}}\\
&\iff x \in Y_1, x \in f^{-k}(Y_2), \ldots, x \in f^{(s-1)k}(Y_s)\\
&\iff x \in Y_1 \cap f^{-k}(Y_2) \cap \ldots \cap f^{(s-1)k}(Y_s).
\end{aligned}$$

Hence, clearly now the statement (7.2) has been proven.

Now by Lemma 7.2.2,

$$\mu(Y_1 \cap f^{-k}(Y_2) \cap \ldots \cap f^{(s-1)k}(Y_s)) = \mu(X_{n_{11}}) \ldots \mu(X_{n_{k1}}) \ldots \mu(X_{n_{ks-1}})\mu(X_{n_{1s}}) \ldots \mu(X_{n_{ks}}).$$

Also, by using Lemma 7.2.2 again,

$$\mu(Y_1 \cap f^{-k}(Y_2) \cap \ldots \cap f^{(s-1)k}(Y_s)) = \prod_{j=1}^{s} \mu(Y_j).$$

□

Lemma 7.2.4. *Let X be an interval of length 1. Let $f : X \longrightarrow X$ be a piecewise linear transformation of X, where f is linear on each of the intervals $X_1, X_2, \ldots, X_r$ which form a partition of X. Let each X_j have positive length. Put*

$$\mathcal{P}_k = \Big\{X_{n_1 n_2 \ldots n_k} : n_1, n_2, \ldots, n_k \in \{1, 2, \ldots, r\}\Big\},$$

for $k = 1, 2, 3, \ldots$. Then, $\mathcal{P}_k$ is a partition of X. If f is completely stretching, $\mathcal{P}_k$ contains of r^k disjoint intervals of positive length.

Proof. We have $X_1, X_2, \ldots, X_r$ partitioning X and $f : X \longrightarrow X$ is piecewise linear transformation of X. Let $x \in X$. Then, given $k \in \mathbb{N}$, there are $n_1, n_2, \ldots, n_k \in \{1, 2, \ldots, r\}$ such that

$$x \in X_{n_1}, f(x) \in X_{n_2}, \ldots, f^{k-1}(x) \in X_{n_k}.$$

Then, $x \in X_{n_1 n_2 \ldots n_k}$. Then, the union of sets in $\mathcal{P}_k$, for a given k, is X. If $x \in X_{n_1 n_2 \ldots n_k} \cap X_{m_1 m_2 \ldots m_k}$, then

$$\begin{aligned}
&x \in X_{n_1} \cap X_{m_1} \Longrightarrow n_1 = m_1\\
&f(x) \in X_{n_2} \cap X_{m_2} \Longrightarrow n_2 = m_2\\
&\ldots\\
&f^{k-1} \in X_{n_k} \cap X_{m_k} \Longrightarrow n_k = m_k.
\end{aligned}$$

Hence, $n_1 = m_1, n_2 = m_2, \ldots, n_k = m_k$, and it follows that

$$X_{n_1 n_2 \ldots n_k} = X_{m_1 m_2 \ldots m_k}.$$

We deduce that the sets in $\mathcal{P}_k$ partition X. Note that $\mathcal{P}_k$ contains of r^k intervals when f is completely stretching (see Lemma 7.2.2).

□

Lemma 7.2.5. *Let f be a completely stretching piecewise linear transformation on interval X of length 1. Let W be a set that is a union of sets in the partition*

$$\mathcal{P}_k = \Big\{ X_{n_1 n_2 \ldots n_k} : n_1, n_2, \ldots, n_k \in \{1, 2, \ldots, r\} \Big\}.$$

Then, for $m = 1, 2, \ldots,$

$$\mu\Big(W \cap f^{-k}(W) \cap f^{-2k}(W) \cap \ldots \cap f^{-(m-1)k}(W) \Big) = \mu(W)^m.$$

Proof. There are disjoint sets $Y_1, Y_2, \ldots, Y_s$ in $\mathcal{P}_k$ such that $W = \bigcup_{j=1}^{s} Y_j$. Then,

$$\begin{aligned}
&\mu\Big(W \cap f^{-k}(W) \cap f^{-2k}(W) \cap \ldots \cap f^{-(m-1)k}(W) \Big) \\
&= \mu\Big[\Big(\bigcup_{j=1}^{s} Y_j \Big) \cap f^{-k}\Big(\bigcup_{j=1}^{s} Y_j \Big) \cap f^{-2k}\Big(\bigcup_{j=1}^{s} Y_j \Big) \cap \ldots \cap f^{-(m-1)k}\Big(\bigcup_{j=1}^{s} Y_j \Big) \Big] \\
&= \mu\Big[\Big(\bigcup_{j=1}^{s} Y_j \Big) \cap \Big(\bigcup_{j=1}^{s} f^{-k}(Y_j) \Big) \cap \Big(\bigcup_{j=1}^{s} f^{-2k}(Y_j) \Big) \cap \ldots \cap \Big(\bigcup_{j=1}^{s} f^{-(m-1)k}(Y_j) \Big) \Big] \\
&= \mu\Big[\bigcup_{j_0, j_1, j_2, \ldots, j_{m-1} \in \{1, 2, \ldots, s\}} Y_{j_0} \cap f^{-k}(Y_{j_1}) \cap f^{-2k}(Y_{j_2}) \cap \ldots \cap f^{-(m-1)k}(Y_{j_{m-1}}) \Big] \\
&= \sum_{j_0, j_1, j_2, \ldots, j_{m-1} \in \{1, 2, \ldots, s\}} \mu\Big[Y_{j_0} \cap f^{-k}(Y_{j_1}) \cap f^{-2k}(Y_{j_2}) \cap \ldots \cap f^{-(m-1)k}(Y_{j_{m-1}}) \Big] \\
&= \sum_{j_0, j_1, j_2, \ldots, j_{m-1} \in \{1, 2, \ldots, s\}} \prod_{t=0}^{m-1} \mu(Y_{j_t}), \text{ by Lemma 7.2.3,} \\
&= \Big(\sum_{j=1}^{s} \mu(Y_j) \Big)^m \\
&= \mu(W)^m.
\end{aligned}$$

□

7.3 The standard deviation for some piecewise linear transformations

Here, we are going to use the previous results to derive an estimate for standard deviation of the first recurrence time in a dynamical system (X, f) where X is a unit interval and f is a completely stretching piecewise linear transformation in X.

Theorem 7.3.1. *Let X be an interval of length 1 and let $X_1, X_2, \ldots, X_r$ be a partition of X into subintervals of X of positive length with $r > 1$. Let f be a completely stretching piecewise transformation on X that is linear on each interval X_j. Put $\rho = \max\{\mu(X_j) : 1 \leq j \leq r\}$, and let U be a subinterval of X of positive length. Then, in the dynamical system (X, f), the mean of the recurrence time r_U over U is $\mu(U)^{-1}$, and the standard deviation $\sigma(r_U)$ of r_U is finite. In fact, if we put*

$$k = 1 + \text{ the integer part of } \Big(\frac{\log(\mu(U)/2)}{\log(\rho)}\Big),$$

then

$$\sigma(r_U) \leq \sqrt{-\frac{1}{\mu(U)^2} + \frac{3}{\mu(U)} - 2 + \frac{k(1 + 2\rho^k - \mu(U))}{\mu(U) - 2\rho^k}}.$$

Proof. It is clear that the mean recurrence time is $\mu(U)^{-1}$ by Theorem 5.5.1.

If $U = X$ except possibly for end points, the standard deviation is zero and is finite, so we may assume that $0 < \mu(U) < 1$. Using Lemma 7.2.2, as $\rho < 1$ we see that the maximum length of an interval in the partition $\mathcal{P}_k$ tends to 0, as $k \longrightarrow \infty$. So, there is $k \in \mathbb{N}$ and a set Y that is finite union of sets in $\mathcal{P}_k$ such that $Y \subseteq U$. Later we will chose k as small as we can so that this still happens. The complement of Y in X is a set W that is also a finite union of sets in $\mathcal{P}_k$ such that $U^c \subseteq W$ (see Figure 7.3). Note that $0 < \mu(W) < 1$.

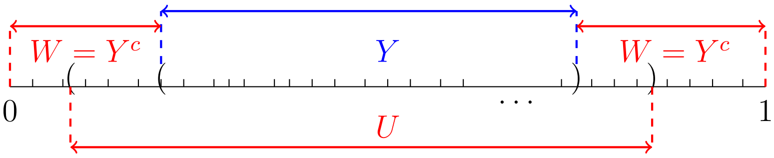

Figure 7.3. This illustrates ideas in part of the argument used in proving Theorem 7.3.1. The set Y is a finite union of intervals which are subsets of U. The set W is the complement of Y and $U^c \subseteq W$. The partition of $[0,1)$ into subintervals come from intervals in $\mathcal{P}_k$, where k is as in the main text, and the maximum possible length of an interval in $\mathcal{P}_k$ is $\rho^k < 1$.

Having chosen k as described and having chosen W as defined,

$$\sum_{n=1}^{\infty} \mu\Big(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-n}(U^c)\Big)$$

$$= \sum_{\ell=1}^{\infty} \left(\sum_{n=(\ell-1)k+1}^{\ell k} \mu\Big(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-n}(U^c)\Big) \right)$$

$$\le \sum_{\ell=1}^{\infty} \left(\sum_{n=(\ell-1)k+1}^{\ell k} \mu\Big(W \cap f^{-1}(W) \cap \ldots \cap f^{-n}(W)\Big) \right), \text{ as } U^c \subseteq W,$$

$$\le \sum_{\ell=1}^{\infty} \left(\sum_{n=(\ell-1)k+1}^{\ell k} \mu\Big(W \cap f^{-1}(W) \cap \ldots \cap f^{-[(\ell-1)k+1]}(W)\Big) \right),$$

as some terms in the intersection have been omitted,

$$\le \sum_{\ell=1}^{\infty} \left(\sum_{n=(\ell-1)k+1}^{\ell k} \mu\Big(W \cap f^{-k}(W) \cap \ldots \cap f^{-(\ell-1)k}(W)\Big) \right),$$

again by omitting some terms in the intersection,

$$= \sum_{\ell=1}^{\infty} k\mu(W)^{\ell}, \text{ by Lemma 7.2.5,}$$

$$= \frac{k\mu(W)}{1-\mu(W)} \tag{7.3}$$

$$< \infty.$$

As

$$\sum_{n=1}^{\infty} \mu\Big(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-n}(U^c)\Big) < \infty,$$

$$\lim_{n\longrightarrow\infty} \mu\Big(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-n}(U^c)\Big) = 0.$$

Now, if we interpret the dynamical system (X, f) as a stochastic process $(X,\ \mathcal{B},\ (X_n), P)$ as explained in Section 3.4, this means that the condition that $\widetilde{P}(0^\infty) = 0$ in Theorem 6.2.1 is satisfied. Hence, by (7.3), the standard deviation $\sigma(r_U)$ of r_U is finite. Note that for the estimate(7.3) for a given k we would achieve the best estimate when we pick $\mu(Y)$ as large as possible to get a minimum value of $\mu(W)$. Also k should be as small as possible.

The estimate in (7.3) may be made more explicit as follows. The sets in $\mathcal{P}_k$ partition X into intervals whose maximum possible length is ρ^k, by Lemma 7.2.2. Thus, if $k > \frac{\log(\mu(U)/2)}{\log(\rho)}$, that is if $\mu(U) > 2\rho^k$, there is a non-void set Y that is finite union of sets in $\mathcal{P}_k$ and is contained in U, and Y is maximal with respect to these properties. The estimate (7.3) applies to the complement W of Y. However, as at most two elements of $\mathcal{P}_k$ have points in both U and U^c, we have

$$\begin{aligned} \mu(Y) + 2\rho^k \geq \mu(U) \qquad &\Longrightarrow 1 - \mu(W) + 2\rho^k \geq \mu(U) \\ &\Longrightarrow 1 + 2\rho^k - \mu(U) \geq \mu(W) \\ &\Longrightarrow 1 > 1 + 2\rho^k - \mu(U) \geq \mu(W), \end{aligned} \tag{7.4}$$

where we have used the fact that $\mu(U) > 2\rho^k$.

Putting (7.4) into (7.3) and after using the fact that $\frac{x}{1-x}$ is increasing for $0 < x < 1$, we have

$$\sum_{n=1}^{\infty} \mu\Big(U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-n}(U^c)\Big) \leq \frac{k(1 + 2\rho^k - \mu(U))}{\mu(U) - 2\rho^k}.$$

Hence, the final parts of the result will follow from applying the formula of Theorem 6.3.1 for the standard deviation in dynamical systems. Therefore,

$$\sigma(r_U) \leq \sqrt{-\frac{1}{\mu(U)^2} + \frac{3}{\mu(U)} - 2 + \frac{k(1 + 2\rho^k - \mu(U))}{\mu(U) - 2\rho^k}}.$$

□

Example 7.3.1.

Consider a completely stretching piecewise transformation in the dynamical system (see Figure 7.4) $([0,1), f)$ where

$$f(x) = \begin{cases} 2x, & \text{if } 0 \leq x < \frac{1}{2} \\ 2x-1, & \text{if } \frac{1}{2} \leq x < 1. \end{cases}$$

Now the second iterate of f (see Figure 7.5) is given by

$$f^2(x) = \begin{cases} 4x, & \text{if } 0 \leq x < \frac{1}{4} \\ 4x-1, & \text{if } \frac{1}{4} \leq x < \frac{1}{2} \\ 4x-2, & \text{if } \frac{1}{2} \leq x < \frac{3}{4} \\ 4x-3, & \text{if } \frac{3}{4} \leq x < 1. \end{cases}$$

In general, the n^{th} iterate of f is

$$f^n(x) = \begin{cases} 2^n x, & \text{if } 0 \leq x < \frac{1}{2^n} \\ 2^n x - 1, & \text{if } \frac{1}{2^n} \leq x < \frac{2}{2^n} \\ \dots & \\ 2^n x - 2^n + 1, & \text{if } \frac{2^n-1}{2^n} \leq x < 1. \end{cases}$$

Now let $U = [\frac{1}{2}, 1) \subseteq [0,1)$. We apply Theorem 7.3.1, we find $\mu(U) = \frac{1}{2}$ and $\rho = \frac{1}{2}$. To find k,

$$\begin{aligned} k &= 1 + \text{ the integer part of } \Big(\frac{\log(\mu(U)/2)}{\log(\rho)}\Big) \\ &= 1 + \text{ the integer part of } \Big(\frac{\log((\frac{1}{2})/2)}{\log(\frac{1}{2})}\Big) \\ &= 1 + \text{ the integer part of } \Big(\frac{\log(\frac{1}{2})^2}{\log(\frac{1}{2})}\Big) \\ &= 1 + \text{ the integer part of } (2). \end{aligned}$$

Thus, $k = 3$. Now, by substituting $\mu(U) = \rho = \frac{1}{2}$ and $k = 3$ in Theorem 7.3.1 we will have

$$\begin{aligned} \sigma(r_{[\frac{1}{2},1)}) &\leq \sqrt{-\frac{1}{(\frac{1}{2})^2} + \frac{3}{(\frac{1}{2})} - 2 + \frac{3\Big(1 + 2(\frac{1}{2})^3 - (\frac{1}{2})\Big)}{(\frac{1}{2}) - 2(\frac{1}{2})^3}}, \text{ so} \\ \sigma(r_{[\frac{1}{2},1)}) &\leq \sqrt{-4 + 6 - 2 + 3(3)}, \text{ and} \\ \sigma(r_{[\frac{1}{2},1)}) &\leq 3. \end{aligned}$$

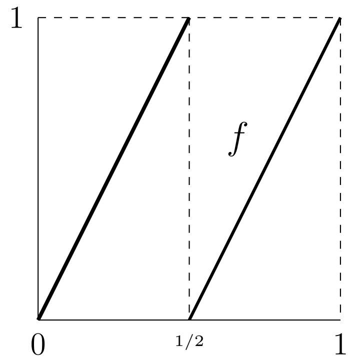

Figure 7.4. The Figure shows the graph of a special completely stretching piecewise transformation in a dynamical system $([0,1), f)$ where f is given by

$$f(x) = \begin{cases} 2x, & \text{if } 0 \leq x < \frac{1}{2} \\ 2x-1, & \text{if } \frac{1}{2} \leq x < 1. \end{cases}$$

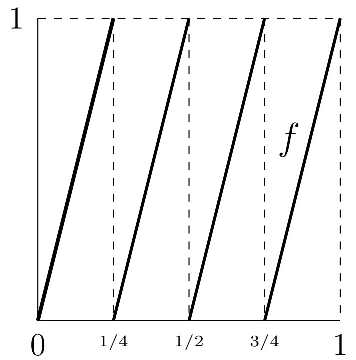

Figure 7.5. The Figure shows the graph of the second iterate of the function f in Figure 7.4, which is given by

$$f^2(x) = \begin{cases} 4x, & \text{if } 0 \leq x < \frac{1}{4} \\ 4x-1, & \text{if } \frac{1}{4} \leq x < \frac{1}{2} \\ 4x-2, & \text{if } \frac{1}{2} \leq x < \frac{3}{4} \\ 4x-3, & \text{if } \frac{3}{4} \leq x < 1. \end{cases}$$

Definition 7.3.1.

Let $f : X \longrightarrow X$ be a transformation on X where X is an interval of length 1. Let $\mathcal{B}$ be the algebra of subsets of X. Then, f is called *length preserving transformation* if for all $J \in \mathcal{B}$,

$$f^{-1}(J) \in \mathcal{B}, \text{ and } \mu(f^{-1}(J)) = \mu(J).$$

Example 7.3.2.

In connection with the analysis above, we take another example (see [20, pages 281-282]) to compare the results. In this example, the formula for the standard deviation (see also [18, page 822]) was derived by assuming f is a length preserving transformation on a bounded interval S and is given by

$$\sqrt{-\frac{\mu(S)^2}{\mu(U)^2}+3\frac{\mu(S)}{\mu(U)}-2+\frac{2}{\mu(U)}\sum_{n=1}^{\infty}\mu\Big(U^c\cap f^{-1}(U^c)\cap\ldots\cap f^{-(n-1)}(U^c)\cap f^{-n}(U^c)\Big)}. \tag{7.5}$$

Note that the formula of (7.5) is similar to the formula in Theorem 6.3.1. Consider $0 \le a \le \frac{1}{2}$ with $V = [0, a)$ and assume f is given by

$$f(x) = \begin{cases} 2x, & \text{if } 0 \le x < \frac{1}{2} \\ 2x-1, & \text{if } \frac{1}{2} \le x < 1. \end{cases}$$

Now observe that

$$V \cap f^{-1}(V) = [0, a) \cap \left(\left[0, \frac{a}{2}\right) \cup \left[\frac{1}{2}, \frac{a+1}{2}\right) \right) = \left[0, \frac{a}{2}\right),$$

and

$$\begin{aligned} V \cap f^{-1}(V) \cap f^{-2}(V) =& V \cap f^{-1}\left(V \cap f^{-1}(V)\right) \\ =& [0, a) \cap f^{-1}\left(\left[0, \frac{a}{2}\right)\right) \\ =& [0, a) \cap \left(\left[0, \frac{a}{4}\right) \cup \left[\frac{1}{2}, \frac{a+2}{4}\right)\right) \\ =& \left[0, \frac{a}{4}\right), \end{aligned}$$

and so on. We will get

$$V \cap f^{-1}(V) \cap f^{-2}(V) \cap \ldots \cap f^{-n}(V) = \left[0, \frac{a}{2^n}\right).$$

Hence,

$$\sum_{n=1}^{\infty} \mu\Big(V \cap f^{-1}(V) \cap f^{-2}(V) \cap \ldots \cap f^{-n}(V)\Big) = \sum_{n=1}^{\infty} \mu\Big(\Big[0, \frac{a}{2^n}\Big) \Big)$$
$$= \sum_{n=1}^{\infty} \frac{a}{2^n} = a.$$

Now putting $U = V^c = [a, 1)$, so $U^c = V$, $\mu(U) = 1 - a$ and the standard deviation of the recurrence time of r_U over U is

$$\sqrt{-\frac{1}{(1-a)^2} + \frac{3}{1-a} - 2 + \frac{2a}{1-a}} = \frac{\sqrt{a(3-4a)}}{1-a}.$$

Thus, if we put $U = \left[\frac{1}{2}, 1\right) \subseteq S = [0, 1)$, then the standard deviation equals $\sqrt{2}$ exactly. Hence, we can compare the result of this example with estimated value in the previous example where $\sqrt{2} \leq 3$. This is because the estimate of the standard deviation of the recurrence time of r_U over U in Theorem 7.3.1 drops some of the terms and also only applies to completely stretching piecewise linear transformations rather than just length preserving transformations.

Example 7.3.3.

Let $f : [0, 1) \longrightarrow [0, 1)$ be a dynamical system (see exercise 10 [20, page 284]) defined by

$$f(x) = frac(x + 1/q),$$

where $q \in \mathbb{N}$ is given with $q > 1$ and $0 \leq x < 1$.
We want first to show this function is length preserving. Observe that

$$f(x) = \begin{cases} \frac{1}{q} + x, & \text{if } 0 \leq x < (q-1)/q \\ x - \frac{(q-1)}{q}, & \text{if } (q-1)/q \leq x < 1. \end{cases}$$

Now

$$f(x) = \begin{cases} f_1(x), & \text{if } 0 \leq x < (q-1)/q \\ f_2(x), & \text{if } (q-1)/q \leq x < 1, \end{cases}$$

where $f_1(x) = 1/q + x, f_2(x) = x - \frac{(q-1)}{q}$, so that $f_1(x) : [0, \frac{q-1}{q}) \longrightarrow [\frac{1}{q}, 1)$,and $f_2(x) : [\frac{q-1}{q}, 1) \longrightarrow [0, \frac{1}{q})$.
That is, $f_1^{-1}(x) = x - 1/q$ and $f_2^{-1}(x) = x + (q-1)/q$. There are three cases

to show this function is length preserving.
Case one (see Figure 7.6): if we take $U = [a, b] \subseteq [\frac{1}{q}, 1)$ then,

$$f^{-1}(U) = f_1^{-1}(U) = [a - 1/q, b - 1/q].$$

Hence, $\mu(U) = b - a = \mu\Big(f^{-1}(U)\Big)$.

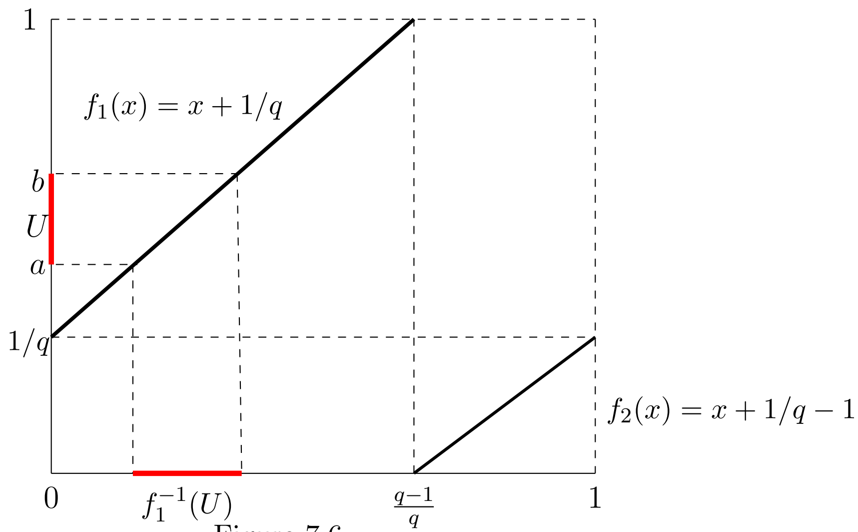

Figure 7.6

Figure 7.6. The Figure shows the graph of the transformation f on $[0, 1)$ which is given by

$$f(x) = \begin{cases} f_1(x) = x + 1/q, & \text{if } 0 \le x < (q-1)/q \\ f_2(x) = x + 1/q - 1, & \text{if } (q-1)/q \le x < 1. \end{cases}$$

Also, if an interval $U = [a, b] \subseteq [\frac{1}{q}, 1)$, then $f^{-1}(U) = f_1^{-1}(U)$ which has the same length as the interval U.

Case two (see Figure 7.7): Observe that $U = [a, b] \subseteq [0, \frac{1}{q})$ so

$$f^{-1}(U) = f_2^{-1}(U) = [a + (q-1)/q, b + (q-1)/q].$$

Thus, $\mu(U) = b - a = \mu\Big(f^{-1}(U)\Big)$.

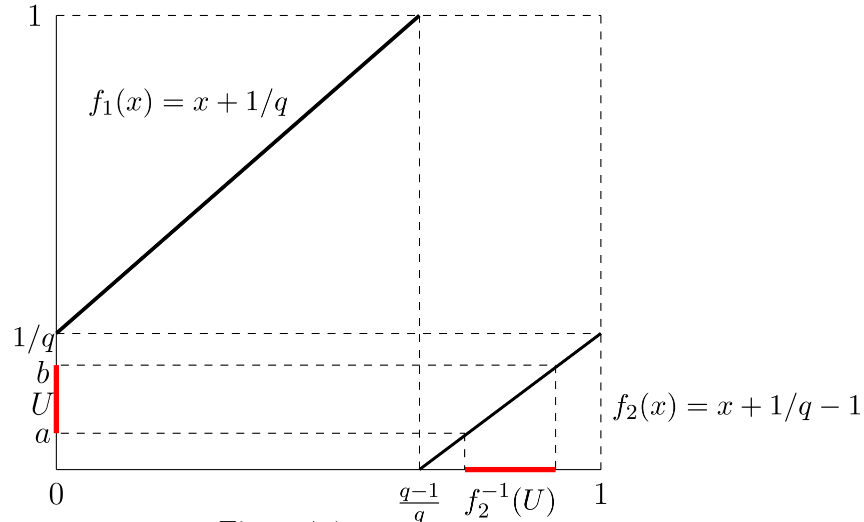

Figure 7.7. The Figure illustrates the inverse image of the interval $U = [a, b] \subseteq [0, \frac{1}{q})$ under f is $f^{-1}(U) = f_2^{-1}(U)$.

Final case (see Figure 7.8): we have $a \leq 1/q \leq b$ and $U = U_1 \cup U_2$ where $U_1 \subseteq [\frac{1}{q}, 1)$ and $U_2 \subseteq [0, \frac{1}{q})$. Now we have

$$U_1 = U \cap \Big[\frac{1}{q}, 1\Big) \text{ and } U_2 = U \cap \Big[0, \frac{1}{q}\Big).$$

That is, $U_1 \cap U_2 = \emptyset$ then,

$$\begin{aligned} f^{-1}(U) =& f^{-1}(U_1 \cup U_2) \\ =& f_1^{-1}(U_1) \cup f_2^{-1}(U_2), \end{aligned}$$

and this union is disjoint. Thus,

$$\begin{aligned} \mu\Big(f^{-1}(U)\Big) =& \mu\Big(f_1^{-1}(U_1)\Big) + \mu\Big(f_2^{-1}(U_2)\Big) \\ =& \mu(U_1) + \mu(U_2), \text{ by cases one and two,} \\ =& \mu(U). \end{aligned}$$

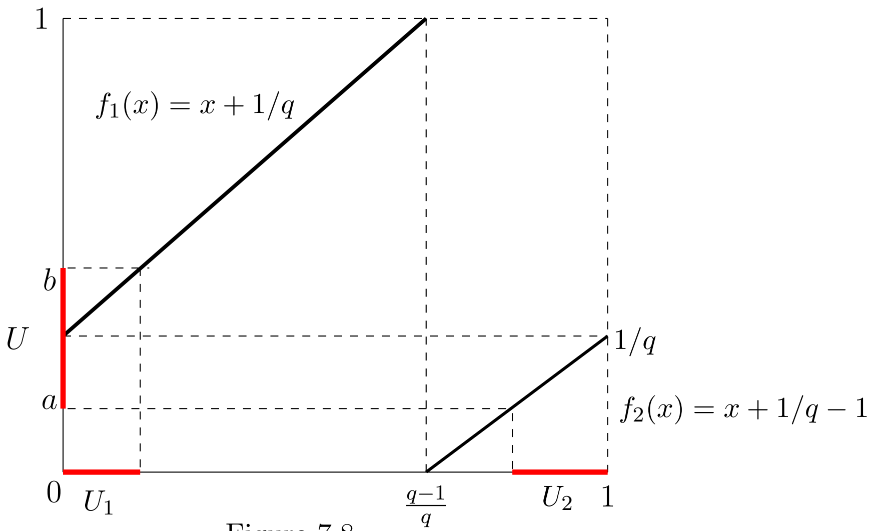

Figure 7.8. The Figure shows the inverse image of the interval U which is the union of the intervals U_1 and U_2.

So, the length of any interval is preserved and this means that f is a length preserving transformation (see [14, Chapter 8]).

Now, we return to find the standard deviation in this example. We will take the interval U with three different cases to show how the standard deviation depends upon the distribution of U.
Case I: We take $U = \left[0, \frac{1}{q}\right)$ and $\mu(U) = \dfrac{1}{q}$. Note that $r_1(x) = q(x)$ for all $x \in \left[0, \frac{1}{q}\right)$ so the standard deviation is 0 (see also the later discussion). We have $U^c = \left[\frac{1}{q}, 1\right) = \left[\frac{1}{q}, \frac{q}{q}\right)$.
Now,

$$\begin{aligned} f^{-1}(U^c) =&\{x : f(x) \in U^c\}, \\ =&\left[0, \frac{1}{q}\right) \cup \left[\frac{1}{q}, \frac{2}{q}\right) \cup \ldots \cup \left[\frac{q-2}{q}, \frac{q-1}{q}\right), \\ =&\left[0, \frac{q-1}{q}\right). \end{aligned}$$

So,

$$U^c \cap f^{-1}(U^c) = \left[\frac{1}{q}, \frac{q-1}{q}\right).$$

Also,

$$\begin{aligned} U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) =& U^c \cap f^{-1}\Big(U^c \cap f^{-1}(U^c)\Big), \\ =& \left[\frac{1}{q}, 1\right) \cap f^{-1}\Big(\left[\frac{1}{q}, \frac{q-1}{q}\right) \Big), \\ =& \left[\frac{1}{q}, 1\right) \cap \left[0, \frac{q-2}{q}\right), \\ =& \left[\frac{1}{q}, \frac{q-2}{q}\right), \end{aligned}$$

and so on. In general, for $j \le q-2$

$$U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) \cap \ldots \cap f^{-j}(U^c) = \left[\frac{1}{q}, \frac{q-j}{q}\right).$$

Now if $j \ge q-1$ then

$$U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-(j)}(U^c) \subseteq U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-(q-1)}(U^c).$$

That is,

$$U^c \cap f^{-1}(U^c) \cap \ldots \cap f^{-(q-1)}(U^c) = \left[\frac{1}{q}, \frac{1}{q}\right) = \emptyset.$$

Therefore,

$$\begin{aligned} \sum_{j=1}^{\infty} \mu\Big(U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) \cap \ldots \cap f^{-j}(U^c)\Big) =& \sum_{j=1}^{q-2} \mu\Big(\left[\frac{1}{q}, \frac{q-j}{q}\right) \Big), \\ =& \sum_{j=1}^{q-2} \frac{q-j-1}{q}, \\ =& \sum_{k=1}^{q-2} \frac{k}{q}, \\ =& \frac{(q-2)(q-1)}{2q}. \end{aligned}$$

By the argument above $P(0^n) = 0$ if $j > q-2$ therefore $\widetilde{P}(0^\infty) = 0$ and formula (7.5) will apply. Now, by putting this in the formula (7.5) the standard deviation will be

$$\sqrt{-q^2 + 3q - 2 + 2q\frac{(q-2)(q-1)}{2q}} = 0.$$

That verifies by means of the earlier formula that the recurrence time is constant and equals q. Of course, we do have

$$f(U) = \left[\frac{1}{q}, \frac{2}{q}\right), f^2(U) = \left[\frac{2}{q}, \frac{3}{q}\right), \ldots, f^{q-1}(U) = \left[\frac{q-1}{q}, 1\right), f^q(U) = \left[0, \frac{1}{q}\right) = U.$$

Case II: We take $U = \left[0, \frac{2}{q}\right)$ and $\mu(U) = 2/q$. So, $U^c = \left[\frac{2}{q}, 1\right)$ and

$$\begin{aligned} f^{-1}(U^c) &= \left[\frac{1}{q}, \frac{2}{q}\right) \cup \ldots \cup \left[\frac{q-2}{q}, \frac{q-1}{q}\right), \\ &= \left[\frac{1}{q}, \frac{q-1}{q}\right). \end{aligned}$$

So,

$$U^c \cap f^{-1}(U^c) = \left[\frac{2}{q}, \frac{q-1}{q}\right).$$

Also,

$$\begin{aligned} U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) =& U^c \cap f^{-1}\Big(U^c \cap f^{-1}(U^c)\Big), \\ &= \left[\frac{2}{q}, 1\right) \cap f^{-1}\left(\left[\frac{2}{q}, \frac{q-1}{q}\right)\right), \\ &= \left[\frac{2}{q}, 1\right) \cap \left[\frac{1}{q}, \frac{q-2}{q}\right), \\ &= \left[\frac{2}{q}, \frac{q-2}{q}\right). \end{aligned}$$

and so on. In general,

$$U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) \cap \ldots \cap f^{-j}(U^c) = \left[\frac{2}{q}, \frac{q-j}{q}\right), \text{ for } 1 \le j \le q-2.$$

Observe that if $j = q-2$, then

$$U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) \cap \ldots \cap f^{-j}(U^c) = \emptyset.$$

In particular, we have

$$U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) \cap \ldots \cap f^{-(q-3)}(U^c) = \left[\frac{2}{q}, \frac{3}{q}\right).$$

It follows that

$$\begin{aligned}\sum_{j=1}^{\infty} \mu\Big(U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) \cap \ldots \cap f^{-j}(U^c)\Big) =& \sum_{j=1}^{q-3} \mu\left(\left[\frac{2}{q}, \frac{q-j}{q}\right)\right), \\ =& \sum_{j=1}^{q-3} \frac{q-j-2}{q}, \\ =& \sum_{k=1}^{q-3} \frac{k}{q}, \\ =& \frac{(q-2)(q-3)}{2q}.\end{aligned}$$

Again by the argument above note that for $j \geq q-2$ is equivalent to having $P(0^n) = 0$ for $n > q$ so $\widetilde{P}(0^\infty) = 0$ and formula (7.5) will apply. By using the formula (7.5) the standard deviation of f over $U = \left[0, \frac{2}{q}\right)$ will be

$$\begin{aligned}\sqrt{-q^2/4 + 3q/2 - 2 + \frac{(q-3)(q-2)}{2}} =& \sqrt{-q^2/4 + 3q/2 - 2 + q^2/2 - 5q/2 + 3} \\ =& \sqrt{q^2/4 - q + 1} \\ =& \sqrt{\frac{1}{4}(q-2)^2} \\ =& \frac{1}{2}(q-2).\end{aligned}$$

Case III: Take $U = \left[0, \frac{1}{q}\right) \cup \left[\frac{2}{q}, \frac{3}{q}\right)$ and $\mu(U) = 2/q$.
So,

$$U^c = \left[\frac{1}{q}, \frac{2}{q}\right) \cup \left[\frac{3}{q}, 1\right).$$

Now,

$$f^{-1}(U^c) = \left[0, \frac{1}{q}\right) \cup \left[\frac{2}{q}, \frac{q-1}{q}\right).$$

So,

$$\begin{aligned}U^c \cap f^{-1}(U^c) =& \left(\left[\frac{1}{q}, \frac{2}{q}\right) \cup \left[\frac{3}{q}, 1\right)\right) \cap \left(\left[0, \frac{1}{q}\right) \cup \left[\frac{2}{q}, \frac{q-1}{q}\right)\right) \\ =& \left[\frac{3}{q}, \frac{q-1}{q}\right).\end{aligned}$$

Hence,

$$\begin{aligned}
U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) =& U^c \cap f^{-1}\Big(U^c \cap f^{-1}(U^c)\Big), \\
=& \left(\left[\frac{1}{q}, \frac{2}{q}\right) \cup \left[\frac{3}{q}, 1\right)\right) \cap f^{-1}\left(\left[\frac{3}{q}, \frac{q-1}{q}\right)\right), \\
=& \left(\left[\frac{1}{q}, \frac{2}{q}\right) \cup \left[\frac{3}{q}, 1\right)\right) \cap \left[\frac{2}{q}, \frac{q-2}{q}\right), \\
=& \left[\frac{3}{q}, \frac{q-2}{q}\right),
\end{aligned}$$

and so on. In general,

$$U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) \cap \ldots \cap f^{-j}(U^c) = \left[\frac{3}{q}, \frac{q-j}{q}\right), \text{ for } j \leq q-3.$$

If $j \geq q-3$ then

$$U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) \cap \ldots \cap f^{-(q-3)}(U^c) = \left[\frac{3}{q}, \frac{3}{q}\right) = \emptyset.$$

Hence, in particular

$$U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) \cap \ldots \cap f^{-(q-4)}(U^c) = \left[\frac{3}{q}, \frac{4}{q}\right).$$

Thus,

$$\begin{aligned}
\sum_{j=1}^{\infty} \mu\Big(U^c \cap f^{-1}(U^c) \cap f^{-2}(U^c) \cap \ldots \cap f^{-j}(U^c)\Big) =& \sum_{j=1}^{q-4} \mu\left(\left[\frac{3}{q}, \frac{q-j}{q}\right)\right), \\
=& \sum_{j=1}^{q-4} \frac{q-j-3}{q}, \\
=& \sum_{k=1}^{q-4} \frac{k}{q}, \\
=& \frac{(q-4)(q-3)}{2q}.
\end{aligned}$$

From above $P(0^n) = 0$ if $j \geq q-3$ so $\widetilde{P}(0^\infty) = 0$, in the context of (7.5). Hence, after applying formula (7.5), the standard deviation of f over $U = \left[0, \frac{1}{q}\right) \cup \left[\frac{2}{q}, \frac{3}{q}\right)$ will be

$$\sqrt{-q^2/4 + 3q/2 - 2 + q\Big(\frac{(q-4)(q-3)}{2q}\Big)} = \sqrt{\frac{1}{4}\Big((q-4)^2\Big)}$$
$$= \frac{(q-4)}{2}, \text{ for } q \geq 4.$$

We will generalize the preceding examples and calculate the standard deviation for f on a more general set U as follows.

Let $q \in \mathbb{N}, q > 2$ and $0 \leq x < 1$. Let $f : [0,1) \longrightarrow [0,1)$ be given by

$$f(x) = frac\left(x + \frac{1}{q}\right).$$

For $j = 1, 2, \ldots, r$ let $k_j, m_j \in \{0, 1, \ldots, q\}$ be such that

$$0 \leq k_1 < m_1 < k_2 < m_2 < \ldots < k_r < m_r.$$

Then, for $j = 1, 2, \ldots, r$ put

$$J_j = \left[\frac{k_j}{q}, \frac{m_j}{q}\right) \subseteq [0,1)\,.$$

The intervals $J_1, J_2, \ldots, J_r$ are disjoint (see Figure 7.9) and

$$\mu(J_j) = \frac{m_j - k_j}{q}.$$

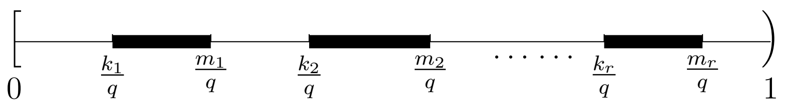

Figure 7.9

Figure 7.9. The Figure shows the interval $\Big[0, 1\Big)$ with a finite number of disjoint intervals with different lengths where $\left[\frac{k_1}{q}, \frac{m_1}{q}\right) \cup \ldots \cup \left[\frac{k_r}{q}, \frac{m_r}{q}\right) \subseteq \Big[0, 1\Big)$.

We put $V = \bigcup_{j=1}^{r} J_j = \bigcup_{j=1}^{r} \left[\frac{k_j}{q}, \frac{m_j}{q}\right)$ and we have

$$\mu(V) = \sum_{j=1}^{r} \mu(J_j) = \frac{1}{q} \sum_{j=1}^{r} m_j - k_j \leq 1.$$

Now put $L_j = m_j - k_j \geq 1$ so that

$$\mu(J_j) = \frac{L_j}{q} \text{ and } \mu(V) = \frac{1}{q} \sum_{j=1}^{r} L_j.$$

Case one: If $k_1 \geq 1$, then

$$f^{-1}(V) = \bigcup_{j=1}^{r} \left[\frac{k_j - 1}{q}, \frac{m_j - 1}{q}\right),$$

and

$$V_1 = V \cap f^{-1}(V) = \bigcup_{j=1}^{r} \left[\frac{k_j}{q}, \frac{m_j - 1}{q}\right). \tag{7.6}$$

Note that V_1 has the same form as V except that the right hand end point of each of the j intervals has reduced by $1/q$. Also, if $k_j = m_j - 1$, then the j^{th} interval is empty.
Now put

$$\begin{aligned} V_2 =& V_1 \cap f^{-1}(V_1) \\ =& V \cap f^{-1}(V) \cap f^{-1}(V \cap f^{-1}(V)) \\ =& V \cap f^{-1}(V) \cap f^{-2}(V). \end{aligned}$$

The same type of calculation as above gives

$$V_2 = \bigcup_{j=1}^{r} \left[\frac{k_j}{q}, \frac{m_j - 2}{q}\right).$$

Now, the interval j is empty when $L_j \leq 2$.

In general, we can use induction and get

$$V_{s+1} = V_s \cap f^{-1}(V_s) = V \cap f^{-1}(V) \cap \ldots \cap f^{-(s+1)}(V),$$

and we get

$$V_s = \bigcup_{j=1}^{r} \left[\frac{k_j}{q}, \frac{m_j - s}{q}\right),$$

and we see $V_s = \bigcup_{j=1}^{r} \left[\frac{k_j}{q}, \frac{m_j-s}{q}\right)$, where an interval $\left[\frac{k_j}{q}, \frac{m_j-s}{q}\right)$ is empty if $L_j \leq s$. Then,

$$\mu(V_s) = \frac{1}{q}\sum_{j=1}^{r} \max\{L_j - s, 0\}.$$

Case two: If $k_1 = 0$ and $m_r < q$ then,

$$f^{-1}(V) = \left[\frac{q-1}{q}, 1\right) \cup \left[0, \frac{m_1 - 1}{q}\right) \cup \bigcup_{j=2}^{r} \left[\frac{k_j - 1}{q}, \frac{m_j - 1}{q}\right),$$

so,

$$\begin{aligned} V \cap f^{-1}(V) =& \left[0, \frac{m_1 - 1}{q}\right) \cup \bigcup_{j=2}^{r} \left[\frac{k_j}{q}, \frac{m_j - 1}{q}\right) \\ =& \bigcup_{j=1}^{r} \left[\frac{k_j}{q}, \frac{m_j - 1}{q}\right), \end{aligned}$$

and this is the same formula as (7.6) so $\mu(V_s) = \frac{1}{q}\sum_{j=1}^{r} \max\{L_j - s, 0\}$, as in case one.

Case three: If $k_1 = 0$ and $m_r = q$ then, $L_1 + L_r = m_r + q - k_r = L_1^*$, where this is a definition of L_1^*.
Hence,

$$V \cap f^{-1}(V) = \left[0, \frac{m_1 - 1}{q}\right) \cup \bigcup_{j=2}^{r-1} \left[\frac{k_j}{q}, \frac{m_j - 1}{q}\right) \cup \left[\frac{k_r}{q}, 1\right).$$

Then, if $s < m_1$,

$$V \cap f^{-1}(V) \cap \ldots \cap f^{-(s)}(V) = \left[0, \frac{m_1 - s}{q}\right) \cup \bigcup_{j=2}^{r-1} \left[\frac{k_j}{q}, \frac{m_j - s}{q}\right) \cup \left[\frac{k_r}{q}, 1\right). \tag{7.7}$$

The sum of the lengths of the first interval plus the last interval in (7.7) will be

$$\frac{1}{q}\Big[m_1 + q - k_r - s\Big] = \frac{(L_1^* - s)}{q}.$$

Now if $s = m_1$, the first interval of (7.7) is empty, and

$$V \cap f^{-1}(V) \cap \ldots \cap f^{-(s)}(V) = \bigcup_{j=2}^{r-1}\left[\frac{k_j}{q}, \frac{m_j - s}{q}\right) \cup \left[\frac{k_r}{q}, \frac{m_1 + q - s}{q}\right).$$

So, here the lengths of the first (empty) interval plus the last interval will be again $(L_1^* - s)/q$. In this case three, if we identify 0 and 1 as length the same point then the first interval has left end point 0 and the last interval has right point 1. Together they really form one interval of length $L_1 + L_r$.

If we do that then V is a union of $(r-1)$ intervals and V_s is also a union of $(r-1)$ intervals with length decreased until 0.

We have the same formula for $\mu(V_s)$ as in case one except that there are only $(r-1)$ intervals, and L_1 is replaced by $L_1^* = L_1 + L_r$.
Putting

$$m = \max\{L_1, L_2, \ldots, L_r\} \leq q,$$

we note $V_s = \emptyset$ if $s \geq m$. Now consider

$$\begin{aligned}\sum_{s=1}^{\infty} \mu(V_s) =& \frac{1}{q}\sum_{s=1}^{m}\sum_{j=1}^{r}\max\{L_j - s, 0\}\\ =& \frac{1}{q}\sum_{j=1}^{r}\sum_{s=1}^{m}\max\{L_j - s, 0\}.\end{aligned}$$

Now

$$\begin{aligned}\sum_{s=1}^{m}\max\{L_j - s, 0\} =& (L_j - 1) + (L_j - 2) + \cdots + 2 + 1\\ =& \frac{L_j(L_j - 1)}{2}.\end{aligned}$$

Hence,

$$\begin{aligned}\sum_{s=1}^{\infty}\mu(V_s) =& \frac{1}{q}\sum_{j=1}^{r}\frac{L_j(L_j - 1)}{2}\\ =& \frac{1}{2q}\sum_{j=1}^{r}L_j^2 - \frac{1}{2}\mu(V).\end{aligned}$$

Now by using the formula (7.5) with $U^c = V$ and $p = \mu(U) = 1 - \mu(V)$. We have

$$\begin{aligned}\sigma_U =& \sqrt{-\frac{1}{p^2} + \frac{3}{p} - 2 + \frac{2}{p}\sum_{s=1}^{\infty}\mu(V_s)} \\ =& \sqrt{-\frac{1}{p^2} + \frac{3}{p} - 2 - \frac{(1-p)}{p} + \frac{1}{pq}\sum_{j=1}^{r} L_j^2} \\ =& \sqrt{-\frac{1}{p^2} + \frac{2}{p} - 1 + \frac{1}{pq}\sum_{j=1}^{r} L_j^2}.\end{aligned}$$

Thus,

$$\sigma_U = \sqrt{-\frac{1}{p^2} + \frac{2}{p} - 1 + \frac{1}{pq}\sum_{j=1}^{r} L_j^2}. \tag{7.8}$$

Now, we apply the formula (7.8) above in the case III of example 7.3.3. Observe that

$$U = \left[0, \frac{1}{q}\right) \cup \left[\frac{2}{q}, \frac{3}{q}\right),$$

$p = \mu(U) = 2/q$ and

$$V = U^c = \left[\frac{1}{q}, \frac{2}{q}\right) \cup \left[\frac{3}{q}, 1\right).$$

Consider

$$J_1 = \left[\frac{1}{q}, \frac{2}{q}\right) \text{ and } J_2 = \left[\frac{3}{q}, 1\right)$$

That is, $\mu(J_1) = 1/q$ so $L_1 = 1$ and $\mu(J_2) = (q-3)/q$ so $L_2 = q - 3$. Now, by using the formula (7.8), we get noting that $pq = 2$,

$$\begin{aligned}\sigma_V =& \sqrt{-q^4/4 + q - 1 + \frac{1}{2}\Big(1 + (q-3)^2\Big)} \\ =& \sqrt{-q^4/4 + q - 1/2 + q^2/2 - 3q + 9/2} \\ =& \sqrt{q^4/4 - 2q + 4} \\ =& \sqrt{\frac{1}{4}(q-4)^2} \\ =& \frac{(q-4)}{2},\end{aligned}$$

exactly the same result we got before.

Chapter 8

Recurrence and standard deviation in a finite dynamical system

8.1 Introduction

The purpose of this Chapter is to discuss the variation of recurrence times in a finite dynamical system. This is related to the predictability and uncertainty of the recurrence times. In the preceding Chapter, one dynamical system considered was $\Big([0,1), f\Big)$, where $f : [0,1) \longrightarrow [0,1)$ was the transformation $x \longmapsto frac(x + 1/q)$, where q was a given element of $\mathbb{N}, q > 1$. That system can be regarded more abstractly as a system (S, f), where S is a finite set and f is a cyclic permutation of S. In this Chapter, we obtain more extensive results in such systems, relating to Kac's formula and the variation of recurrence times for a set A in relation to the "geometry" and distribution of the points of A in S. Let (S, f) be a dynamical system where S is a finite set and f is a cyclic permutation of S. It should be noted that every dynamical system (T, g), where T is a finite set, contains a dynamical system of this type. In any such dynamical system (T, g), we have

$$T \supseteq g(T) \supseteq g^2(T) \supseteq \dots .$$

If T is finite this sequence of inclusions must terminate with equality, so there is a least value $n \in \mathbb{N}$ such that

$$g^n(T) = g^{n+1}(T).$$

If we put $Y = g^n(T)$ we see that $g : Y \longrightarrow Y$ and that

$$g(Y) = g(g^n(T)) = g^{n+1}(T) = Y.$$

However, as Y is finite, so g is a permutation on Y. But every permutation is a composition of cycles, so g acts as a cyclic permutation on some subset of Y (note that this set may consist of a single point).

8.2 Preliminaries

If A is a finite set, $|A|$ will denote the number of elements in A. For the finite set S, we assume that $|S| \geq 2$. We introduce a probability P on the algebra of all subsets of $S = \{u_0, u_1, \dots, u_{|S|-1}\}$ by putting

$$P(A) = \frac{|A|}{|S|}, \text{ for all } A \subseteq S.$$

Then, for $x \in S, P(\{x\}) = 1/|S|$ and so $P(\{x\})$ is the same for all $x \in S$. Note that for $A \subseteq S, P\Big(f^{-1}(A)\Big) = P(A)$, so that f is length-preserving. Let $f : S \longrightarrow S$ be the cyclic permutation on S given by

$$f(u_j) = \begin{cases} u_{j+1}, & \text{for } 0 \leq j \leq |S|-2, \\ u_0, & \text{for } j = |S|-1. \end{cases} \tag{8.1}$$

Note that we choose a point $u_0 \in S$, then

$$u_1 = f(u_0), u_2 = f(u_1), \dots, u_{|S|-1} = f(u_{|S|-2}),$$

and note that $f(u_{|S|-1}) = u_0$.

Now, S may be visualised as consisting of q points, where $q = |S|$, on the unit circle $\mathbb{T}$ in the complex plane, starting with 1 and, proceeding anticlockwise, equally spaced around the circumference of $\mathbb{T}$. That is, S may be identified with the q^{th} roots of unity, with $u_j = e^{2\pi ij/q}$ for $j = 0, 1, \dots, q-1$. In this visualisation, f corresponds to an anticlockwise rotation through $2\pi/q$.

Definition 8.2.1.

A subset A of S is called an *arc* if it is of the form $\{u_j, u_{j+1}, \dots, u_k\}$ for some $0 \leq j \leq k \leq q-1$, or of the form $\{u_j, \dots, u_{q-1}, u_0, \dots, u_k\}$ for some $0 \leq k \leq j \leq q-1$. In each of these cases we call u_j the *beginning point* of the arc, and u_k the *endpoint* of the arc.

Definition 8.2.2.

Given two arcs J and K, we say that K is *consecutive* to J if f maps the endpoint of J to the beginning point of K.

Definition 8.2.3.

We say that two arcs J, K of S are *separated* if their union is not an arc. This notion of an arc in S coincides with the usual meaning when we regard S as a subset of $\mathbb{T}$ consisting of equally spaced consecutive points as we move in an anti-clockwise direction around $\mathbb{T}$ (see Figure 8.1). Distance is meaning in an anti-clockwise direction. Then the distance from J to K is not the same the distance from K to J.

Definition 8.2.4.

Any non-empty subset A of S may be written uniquely in the form

$$A = \bigcup_{j=1}^{t} J_j, \tag{8.2}$$

where $t \in \{1, 2, \ldots, q\}$, and $J_1, J_2, \ldots, J_t$ are separated non-empty arcs. The sets $J_1, J_2, \ldots, J_t$ are also called the *components* of A. The expression of A in the form (8.2) is called the *decomposition* of A. Also, provided both A and A^c are non-empty, the complement A^c has a decomposition into t disjoint arcs, let's say

$$A^c = \bigcup_{j=1}^{t} K_j. \tag{8.3}$$

The arcs J_j and K_j in (8.2) and (8.3) may be numbered so that the arcs $J_1, K_1, J_2, K_2, \ldots, J_t, K_t, J_1$ are consecutive. That is, in the circle interpretation, starting with J_1 as we proceed anti-clockwise around the circle $\mathbb{T}$, we encounter $J_1, K_1, J_2, K_2, \ldots, J_t, K_t$ and J_1 in that order. We will always assume that the arcs have been numbered in this way.

Definition 8.2.5.

We call K_j in (8.3) the *gap* between J_j and J_{j+1} for $j = 1, 2, \ldots, t-1$ and also we call K_t the *gap* between J_t and J_1. The *length* of the gap is taken to be $|K_j|$.

In Figure 8.1 below, the set on the left is decomposed into nine separated arcs, each of which is a single point; while the set on the right is decomposed into three separated arcs, having respectively three, four and five points.

Note that since each arc contains at least one point and if a subset A of S has t components, then $t \leq |A|, t + |A| \leq |S|$ and $t \leq |S|/2$.

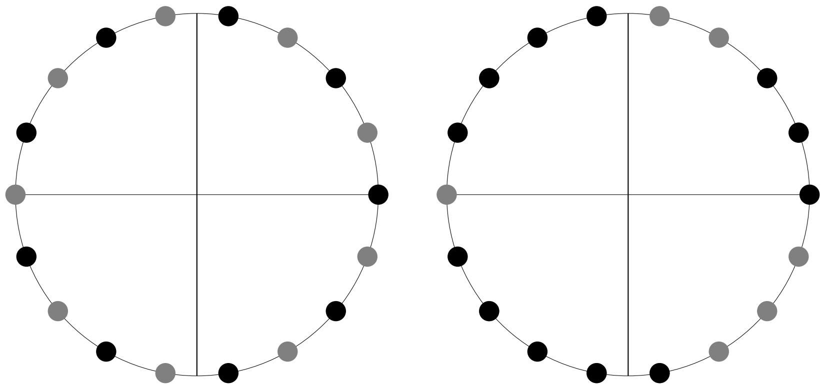

Figure 8.1. In each of the left and the right, the black and grey points together comprise the set S, in this case the 18^{th} roots of unity. The transformation f corresponds to an anticlockwise rotation through 20^o. On the left, the black points define a subset A of S with $|A| = 9$, and the decomposition of A is into 9 arcs. The recurrence time for every point of A is 2, because the points of A are equally spaced. So, for the set A, the recurrence times are completely predictable. On the right, the black points define a subset B of S, with $|B| = 12$ and the decomposition of B is into 3 arcs. The points of B have a recurrence time of 1, 2, 3 or 4. So, if all we know is that a point is in B, its recurrence time is not predictable.

8.3 Recurrence times and the average in a finite dynamical system

Given a non-empty subset A of S, and given $x \in A$, there is $n \in \mathbb{N}$ such that $f^n(x) \in A$, and we consider the recurrence times as described in Section 5.3:

$$r_A(x) = \min\{n \in \mathbb{N} \text{ and } f^n(x) \in A\}, \tag{8.4}$$

where f is a cyclic permutation of S as in described (8.1). Note that for $x \in A, f(x) \in f(A)$ and

$$r_{f(A)}\Big(f(x)\Big) = r_A(x).$$

Thus, A and $f(A)$ have the same recurrence properties, and we say that recurrence phenomena are *invariant* under f.

Now the recurrence times of points in a subset A of S are determined by the lengths of the arcs in the decomposition of S, and by the gaps between the arcs. Thus, if $A = \bigcup_{j=1}^{t} J_j$ is the decomposition of A as in (8.2), and if $x \in J_j$ but x is not the endpoint of A, then $r_A(x) = 1$; but if x is the endpoint of $J_j, r_A(x)$ is 1 plus the length of the gap between x and the beginning point of the next arc in the decomposition. The following gives us an idea of the possibilities for the recurrence times, by expressing in (8.5) and (8.6) below estimates for the proportion of points having a given recurrence time.

Theorem 8.3.1. *Let S be a finite set with $|S| \geq 2$, let $A \subseteq S$ and let f be permutation on S as in (8.1). Then, the following statements hold.*
(i) The recurrence times for all points in A are in $\{1, 2, \ldots, |S|\}$ and for all $k \in \{2, 3, \ldots, |S|\}$,

$$\frac{1}{|A|}\Big|\Big\{x : x \in A \text{ and } r_A(x) = k\Big\}\Big| \leq \frac{1}{k-1} \cdot \frac{|A^c|}{|A|}. \tag{8.5}$$

If there is a point of A with recurrence time $|S|$, and we put $k = |S|$ in (8.5), then equality holds.
(ii) The inequality (8.5) is sharp, in the sense that for any given $k \in \mathbb{N}$, the set S may be chosen together with a subset A of S such that equality will hold.
(iii) If the subset A has a decomposition with t components, then

$$\frac{1}{|A|}\Big|\Big\{x : x \in A \text{ and } r_A(x) = 1\Big\}\Big| = 1 - \frac{t}{|A|}, \tag{8.6}$$

and the maximum possible recurrence time for a point in A is $|S| - |A| - t + 2$. This is attained when there is a gap of length 1 between $t - 1$ components of the decomposition of A and the remaining gap is of length $|S| - |A| - t + 1$.

Proof. (i) As $f^{|S|}(x) = x$ for all $x \in S$, it is clear that $r_A(x) \in \{1, 2, \ldots, |S|\}$. If there is a point of A with recurrence time $|S|$, we see that $|A| = 1$ and the result holds for $k = |S|$ with both sides of (8.5) being equal to 1.
(ii) Let $k \in \{2, 3, \ldots, |S|\}$ and suppose there are ℓ elements in A such that $r_A(x) = k$. Then, from earlier remarks we have $|A^c| \geq \ell(k-1)$, from which (8.5) follows. To see that the estimate in (8.5) is sharp, let $k \in \mathbb{N}$ be given and let $|S|$ be divisible by k. Let A be a subset of S in which every element has recurrence time k. Then, equality will hold in (8.5).
(iii) When A has t components, $\Big|\Big\{x : x \in A \text{ and } r_A(x) = 1\Big\}\Big| = |A| - t$ and (8.6) follows.

The final statement is immediate on realising that the maximum recurrence time comes from having a gap of maximum possible length in the decomposition, and then the calculating the length of the gap and the corresponding recurrence time. □

Here, because the system is finite, we can give a very quick derivation based on counting the recurrence times of the elements of the subset to obtain the average recurrence time. The following result gives a version of Kac's recurrence formula for a finite dynamical system (for more discussion see [16]).

Theorem 8.3.2. *Let (S, f) be a finite dynamical system and suppose that f acts as a cyclic permutation on S. Let A be a non-empty subset of S. Then the average of the recurrence time of r_A over A equals $|S|/|A|$. That is,*

$$\frac{1}{|A|}\sum_{x\in A} r_A(x) = \frac{|S|}{|A|}.$$

Proof. We write the decomposition of A as

$$A = \bigcup_{j=1}^{t} J_j,$$

where $J_1, J_2, \ldots, J_t$ are arcs as in (8.2). The complement A^c also has a decomposition into t arcs as in (8.3), say

$$A^c = \bigcup_{j=1}^{t} K_j.$$

If we renumber the arcs $K_1, K_2, \ldots, K_t$, if necessary, we can see that for $x \in J_j, r_A(x) = 1$ or $r_A(x) = |K_j| + 1$, and the latter occurs for a single $x \in J_j$. Thus, the average of $r_A(x)$ over A equals

$$\frac{1}{|A|}\sum_{x\in A} r_A(x) = \frac{1}{|A|}\Big[\sum_{j=1}^{t} |J_j| - 1 + |K_j| + 1\Big] = \frac{1}{|A|}\Big[\sum_{j=1}^{t} |J_j| + |K_j|\Big] = \frac{|S|}{|A|}.$$

□

Note that the formula above shows that the average of the recurrence time depends only upon the number of points in the set A, and not upon how the points are distributed in S. However, when considering the variation of recurrence times, the distribution of the points of A in S affects the amount of variation.

8.4 The standard deviation of the recurrence time

We consider a finite dynamical system (S, f) where f is acting as a cyclic permutation. If A is a non-empty subset of S and $r_A : A \longrightarrow \mathbb{N}$ is the function giving the recurrence time for points in A, then the standard deviation (as in Section 6.2) of $r_A(x)$ over A is

$$\sigma_A = \sqrt{\frac{1}{|A|}\sum_{x\in A}\left(r_A(x) - \frac{|S|}{|A|}\right)^2}. \tag{8.7}$$

Recall that the standard deviation is a measure of how much the recurrence times $r(x)$ vary from their average value $|S|/|A|$. Note that if $A = S$, then $r_A = 1$ and $\sigma_A = 0$. So, our discussion assumes that A is a non-empty but proper subset of S and we assume that its decomposition consists of t arcs. Then, the decomposition of A^c also consists of t arcs, say $K_1, K_2, \ldots, K_t$. Using (8.7) we see that

$$\sigma_A^2 = \frac{1}{|A|}\left[\sum_{j=1}^{|A|-t}\left(1 - \frac{|S|}{|A|}\right)^2 + \sum_{j=1}^{t}\left(|K_j| + 1 - \frac{|S|}{|A|}\right)^2\right],$$

so that

$$\sigma_A = \sqrt{\left(1 - \frac{t}{|A|}\right)\left(1 - \frac{|S|}{|A|}\right)^2 + \frac{1}{|A|}\sum_{j=1}^{t}\left(|K_j| + 1 - \frac{|S|}{|A|}\right)^2}. \tag{8.8}$$

Note that σ_A depends upon the lengths of the arcs in the decomposition of A^c. Equivalently, σ_{A^c} depends upon the lengths of the arcs in the decomposition of A. Note that $|S|$ is even if $|A| = |A^c| = |S|/2$.

Now, let us compare the formula (8.8) above with the formula (7.8) in Chapter 7. We have from (8.8) that

$$\sigma_A = \sqrt{\left(1 - \frac{t}{|A|}\right)\left(1 - \frac{|S|}{|A|}\right)^2 + \frac{1}{|A|}\sum_{j=1}^{t}|K_j|^2 + \frac{2}{|A|}\left(1 - \frac{|S|}{|A|}\right)\sum_{j=1}^{t}|K_j| + \frac{t}{|A|}\left(1 - \frac{|S|}{|A|}\right)^2}. \tag{8.9}$$

We mentioned in the introduction that the set-up in this Chapter is an extension of a particular example considered in Chapter 7. We will now check that formula (8.9) is consistent with the formula (7.8) in Chapter 7 .

In Chapter 7, we considered the system $\Big([0,1), f\Big)$, where

$$f(x) = frac(x + 1/q),$$

for $0 \leq x < 1$, and $q \in \mathbb{N}$ is given with $q > 1$. We considered a union of subintervals U of $[0,1)$ whose length was a multiple of $1/q$ and we put $\mu(U) = p$, so in (7.8) we put

$$|S| = q, |A| = \frac{\mu(U)}{1/q} = pq, t = r, |K_j| = L_j, \text{ and } \sum_{j=1}^{t} |K_j| = q - pq.$$

Now, we put terms the above into the formula (8.9) above for the standard deviation to get

$$\begin{aligned}
\sigma_A &= \sqrt{\left(1 - \frac{r}{pq}\right)\left(1 - \frac{1}{p}\right)^2 + \frac{1}{pq}\sum_{j=1}^{r} L_j^2 + \frac{2}{pq}\left(1 - \frac{1}{p}\right)(q - pq) + \frac{r}{pq}\left(1 - \frac{q}{pq}\right)^2} \\
&= \sqrt{\left(1 - \frac{1}{p}\right)^2\left[1 - \frac{r}{pq} + \frac{r}{pq}\right] + \frac{2(q - pq)}{pq}\left(1 - \frac{1}{p}\right) + \frac{1}{pq}\sum_{j=1}^{r} L_j^2} \\
&= \sqrt{\left(1 - \frac{1}{p}\right)^2 + \frac{2(q - pq)}{pq}\left(1 - \frac{1}{p}\right) + \frac{1}{pq}\sum_{j=1}^{r} L_j^2} \\
&= \sqrt{\left(1 - \frac{1}{p}\right)\left[1 - \frac{1}{p} + \frac{2}{p} - 2\right] + \frac{1}{pq}\sum_{j=1}^{r} L_j^2} \\
&= \sqrt{\left(1 - \frac{1}{p}\right)\left(\frac{1}{p} - 1\right) + \frac{1}{pq}\sum_{j=1}^{r} L_j^2} \\
&= \sqrt{-\frac{1}{p^2} + \frac{2}{p} - 1 + \frac{1}{pq}\sum_{j=1}^{r} L_j^2},
\end{aligned}$$

which is exactly the formula (7.8) in Chapter 7.

The following result is essentially a special case of known results concerning extreme values of convex functions, (see [25, pages 122-126], for example). It is included here to make the exposition as complete and as elementary as possible.

Lemma 8.4.1. *Let $t \in \mathbb{N}, b \geq t$ and C be the subset of $\mathbb{R}^t$ given by*

$$C = \Big\{(x_1, x_2, \ldots, x_t) : x_1, x_2, \ldots, x_t \geq 1 \text{ and } \sum_{j=1}^{t} x_j = b\Big\}.$$

Let $a \in \mathbb{R}$. Let $f : C \longrightarrow [0, \infty)$ be the function given by

$$f(x_1, x_2, \ldots, x_t) = \sum_{j=1}^{t} (x_j - a)^2.$$

Then f attains a maximum and a minimum over C. The maximum occurs precisely at the t points of the form $(1, \ldots, 1, b - t + 1, 1, \ldots, 1)$ and equals $(b - t - a + 1)^2 + (t - 1)(a - 1)^2$. The minimum occurs at a unique point which is $(b/t, b/t, \ldots, b/t)$ and it equals $t(b/t - a)^2$.

Proof. Clearly, f is continuous on C and so attains a maximum and a minimum over C, since C is closed and bounded in $\mathbb{R}^t$ (see [21, page 114] or [26, pages 115 and 119]). A property of f we will use is the symmetry property: if $(x_1, x_2, \ldots, x_t) \in \mathbb{R}^t$ and if $(y_1, y_2, \ldots, y_t) \in \mathbb{R}^t$ is obtained by writing the coordinates $x_1, x_2, \ldots, x_t$ in a different order, then f has the same value at both $(x_1, x_2, \ldots, x_t)$ and $(y_1, y_2, \ldots, y_t)$.

Now, let's suppose f attains a maximum at $x = (x_1, x_2, \ldots, x_t) \in C$ with, say $x_1 > 1$ and $x_2 > 1$. Put $x' = (x_1 + x_2 - 1, 1, x_3, \ldots, x_t) \in C$. Then, $x' \in C$ and we have

$$\begin{aligned} f(x') - f(x) =& (x_1 + x_2 - 1 - a)^2 + (1 - a)^2 - (x_1 - a)^2 - (x_2 - a)^2 \\ =& x_1^2 + 2x_1x_2 + x_2^2 - 2x_1 - 2x_2 - 2ax_1 - 2ax_2 + 2 - x_1^2 + 2ax_1 - x_2^2 + 2ax_2 \\ =& 2x_1x_2 - 2x_1 - 2x_2 + 2 \\ =& 2(x_1x_2 - x_1 - x_2 + 1) \\ =& 2(x_1 - 1)(x_2 - 1) \\ >& 0, \end{aligned}$$

so that $f(x') > f(x)$. This contradicts the assumption that f has a maximum over C at x. Thus, by the symmetry property of f, we see that the points in C where f attains its maximum are precisely those of the form

$$(1, \ldots, 1, b - t + 1, 1, \ldots, 1),$$

where $b - t + 1$ is in position j for some $j \in \{1, 2, \ldots, t\}$. Also, we see that the maximum of f over C is the value of f at each of these points, which is $(b - t - a + 1)^2 + (t - 1)(a - 1)^2$, as stated.

In considering the minimum of f over C, we show first that this occurs at a unique point in C. Then, the symmetry property of f implies that all coordinates of the point where the minimum occurs must be equal. So, let's assume that f has a minimum value m at points $u, v \in C$ where $u \neq v$. We use the fact that for $\alpha, \beta \in \mathbb{R}$.

$$\frac{\alpha^2}{2} + \frac{\beta^2}{2} = \Big(\frac{\alpha+\beta}{2}\Big)^2 + \Big(\frac{\alpha-\beta}{2}\Big)^2.$$

Thus, putting $u = (u_1, \ldots, u_t)$ and $v = (v_1, \ldots, v_t)$, and noting that $f(u) = f(v) = m$, we have

$$\begin{aligned}
m =& \frac{1}{2}f(u) + \frac{1}{2}f(v) \\
=& \sum_{j=1}^{t}\left[\frac{(u_j-a)^2}{2} + \frac{(v_j-a)^2}{2}\right] \\
=& \sum_{j=1}^{t}\Big(\frac{u_j+v_j}{2} - a\Big)^2 + \sum_{j=1}^{t}\Big(\frac{u_j-v_j}{2}\Big)^2 \\
=& f\Big(\frac{u+v}{2}\Big) + \sum_{j=1}^{t}\Big(\frac{u_j-v_j}{2}\Big)^2 \\
>& f\Big(\frac{u+v}{2}\Big), \text{ since } u \neq v.
\end{aligned}$$

This contradicts the assumption that m is the minimum value of f over C. Thus, f assumes its minimum at a unique point, all coordinates of this point must be equal by symmetry, so the minimum is assumed at $(b/t, b/t, \ldots, b/t)$. We also see that the minimum value is $t(b/t - a)^2$.

□

We will use the preceding result to show that under the given conditions, the maximum of the standard deviation of the recurrence times, taken over a subset A with a given number of points and a given number of components, is attained when the components are as close together as possible, but with two components having a larger gap, in general. The corresponding minimum is attained when the components have the same number of points and are equally spaced.

Theorem 8.4.1. *Let (S, f) be a finite dynamical system where f acts as a cyclic permutation on S and $|S| \geq 2$. Let $t, s \in \{1, 2, \ldots, |S|-1\}$ be given with $t \leq s$. Let $\mathcal{A}(t, s)$ consist of all subsets A of S such that $|A| = s$ and the decomposition of A consists of t arcs. Then,*

$$\max_{A\in\mathcal{A}(t,s)} \sigma_A$$
$$= \sqrt{\left(1-\frac{t}{s}\right)\left(1-\frac{|S|}{s}\right)^2+\frac{1}{s}\left(|S|-s-t-\frac{|S|}{s}+2\right)^2+\frac{t-1}{s}\left(-\frac{|S|}{s}+2\right)^2}. \tag{8.10}$$

Now, assume further that t divides $|S|-s$. Then,

$$\min_{A\in\mathcal{A}(t,s)} \sigma_A = \left(\frac{|S|}{s}-1\right)\sqrt{\frac{s}{t}-1}. \tag{8.11}$$

The maximum in (8.10) is attained at any set A in $\mathcal{A}(t,s)$ such that, for the arcs in the decomposition of A, the length of the gap between $t-1$ consecutive arcs is 1, while the length of the remaining gap is $|S|-t-s+1$. The minimum in (8.11) is attained at any set A in $\mathcal{A}(t,s)$ such that, for the arcs in the decomposition of A, the length of all the gaps between consecutive arcs is $(|S|-s)/t$.

Proof. Observe that we have $|S| = |A| + |A^c| = s + \sum_{j=1}^{t} |K_j|$, so that

$$\sum_{j=1}^{t} |K_j| = |S| - s. \tag{8.12}$$

The idea is to use Lemma 8.4.1. Observe that, because of (8.8) and (8.12) the problem is equivalent to finding the maxima and minima of the function $f : \mathbb{R}^t \longrightarrow [0,\infty]$ given by

$$f(x_1, x_2, \ldots, x_t) = \sum_{j=1}^{t}\left(x_j + 1 - \frac{|S|}{s}\right)^2, \tag{8.13}$$

subject to the conditions that

$$x_j \geq 1 \text{ for } j = 1,2,\ldots,t \text{ and } \sum_{j=1}^{t} x_j = |S| - s. \tag{8.14}$$

However, note that for the maxima and minima of the standard deviation over $\mathcal{A}(t,s)$ we need to have them occurring for positive integer values of $x_1, x_2, \ldots, x_t$.

Now, the conditions on $(x_1, x_2, \ldots, x_t)$ in (8.14) define a subset C of $\mathbb{R}^t$ and, in fact, we see that Lemma 8.4.1 now applies to f over C as given in (8.13) with

$$a = \frac{|S|}{s} - 1 \text{ and } b = |S| - s.$$

By Lemma 8.4.1, the maximum of f over C occurs at the points $e_1, e_2, \ldots, e_t$ where

$$e_j = (1, 1, \ldots, |S| - s - t + 1, 1, 1, \ldots, 1),$$

and $|S| - s - t + 1$ is in position j. Note that the coordinates of e_j are all positive integers. Also, this maximum value is

$$\Big(b-a-t+1\Big)^2 + (t-1)\Big(a-1\Big)^2 = \left(|S| - s - t + 2 - \frac{|S|}{s}\right)^2 + (t-1)\left(\frac{|S|}{s} - 2\right)^2.$$

Consequently, using (8.8) and (8.13) we see that the maximum value of σ_A^2 over $\mathcal{A}(t, s)$ is

$$\left(1 - \frac{t}{s}\right)\left(1 - \frac{|S|}{s}\right)^2 + \frac{1}{s}\left(|S| - s - t - \frac{|S|}{s} + 2\right)^2 + \frac{t-1}{s}\left(-\frac{|S|}{s} + 2\right)^2,$$

and so the conclusion (8.10) follows.

Lemma 8.4.1 applies also to considering the minimum, and we see that f has a minimum over C at the point

$$\left(\frac{|S| - s}{t}, \frac{|S| - s}{t}, \ldots, \frac{|S| - s}{t}\right),$$

and note that because we are assuming that t divides $|S| - s$, the coordinates of this point are positive integers and so will give a minimum for the standard deviation, not just a minimum for f. The value of this minimum is

$$t\left(\frac{b}{t} - a\right)^2 = t\left(\frac{|S| - s}{t} + 1 - \frac{|S|}{s}\right)^2.$$

Thus, again using (8.8) and (8.13) we see that the minimum value of σ_A^2 over $\mathcal{A}(t, s)$ is

$$\begin{aligned} &\left(1 - \frac{t}{s}\right)\left(1 - \frac{|S|}{s}\right)^2 + \frac{t}{s}\left(\frac{|S| - s}{t} + 1 - \frac{|S|}{s}\right)^2 \\ = \; &\left(1 - \frac{t}{s}\right)\left(1 - \frac{|S|}{s}\right)^2 + \frac{t}{s}\Big(\frac{s}{t} - 1\Big)^2\left(\frac{|S|}{s} - 1\right)^2 \end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{|S|}{s} - 1\right)^2 \left(1 - \frac{t}{s} + \frac{t}{s}(\frac{s}{t} - 1)^2\right) \\
&= \left(\frac{s}{t} - 1\right)\left(\frac{|S|}{s} - 1\right)^2.
\end{aligned}
$$

This proves (8.11). □

A special case of Theorem 8.4.1 is when $t = |S| - s$, which is when the lengths of the gaps in the decomposition of A all equal 1. In this case the standard deviation is the same for all elements of $\mathcal{A}(t, s)$, so the maximum and minimum in (8.10) and (8.11) should both equal to

$$
\sqrt{\left(\frac{|S|}{s} - 1\right)\left(2 - \frac{|S|}{s}\right)}.
$$

We now check this for the maximum. We have

$$
\begin{aligned}
\sigma_A = & \sqrt{\left(2 - \frac{|S|}{s}\right)\left(1 - \frac{|S|}{s}\right)^2 + \frac{1}{s}\left(2 - \frac{|S|}{s}\right)^2 + \frac{|S| - s - 1}{s}\left(2 - \frac{|S|}{s}\right)^2} \\
= & \sqrt{\left(2 - \frac{|S|}{s}\right)\left(1 - \frac{|S|}{s}\right)^2 + \left(2 - \frac{|S|}{s}\right)^2\left(\frac{1}{s} + \frac{|S|}{s} - 1 - \frac{1}{s}\right)} \\
= & \sqrt{\left(2 - \frac{|S|}{s}\right)\left(\frac{|S|}{s} - 1\right)^2 + \left(2 - \frac{|S|}{s}\right)^2\left(\frac{|S|}{s} - 1\right)} \\
= & \sqrt{\left(\frac{|S|}{s} - 1\right)\left(2 - \frac{|S|}{s}\right)\left(\frac{|S|}{s} - 1 + 2 - \frac{|S|}{s}\right)} \\
= & \sqrt{\left(\frac{|S|}{s} - 1\right)\left(2 - \frac{|S|}{s}\right)},
\end{aligned}
$$

which was obtained by applying the formula (8.10) for the case $t = |S| - s$. Also, the minimum in (8.11) will be

$$
\begin{aligned}
\sigma_A = & \sqrt{\left(\frac{|S|}{s} - 1\right)^2\left(\frac{2s - |S|}{|S| - s}\right)} \\
= & \sqrt{\left(\frac{|S|}{s} - 1\right)\left(2 - \frac{|S|}{s}\right)}.
\end{aligned}
$$

Note also that in Theorem 8.4.1 the expression for the maximum value looks complicated compared with the one for the minimum value. This is not too surprising, as the minimum is attained when the points in the minimising set are equally spaced, but for the maximising set the points are irregularly spaced (see Figure 8.2).

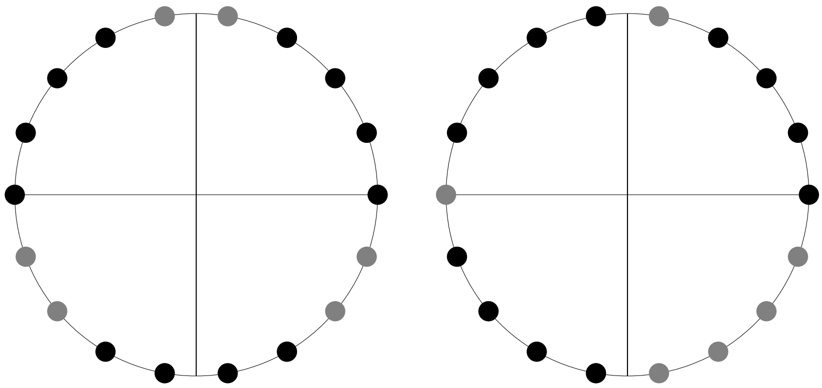

Figure 8.2. On the left and the right the set S consists of the 18^{th} roots of unity, as in Figure 8.1. On the left, the set A given by the black dots has 12 elements and 3 components, and we observe that $A \in \mathcal{A}_{3,12}$. Note that 3 divides into $18 - 12 = 6$. The components of A are equally spaced and so the set A minimises the standard deviation of the recurrence time over the sets in $\mathcal{A}_{3,12}$. On the right, the set B indicated by the black dots is also in $\mathcal{A}_{3,12}$. This time the components are placed as close together as possible, but of necessity leaving a larger gap between 2 of the components. So, the set B maximises the standard deviation of the recurrence time over the sets in $\mathcal{A}_{3,12}$.

Also, a case of special interest in Theorem 8.4.1 is when $t = s$. This restriction means that the components of the sets in $\mathcal{A}(t,s) = \mathcal{A}(t,t)$ consist of single points, and we must have $t \leq |S|/2$. When $t = s$ and $|S|$ is even, the minimum in (8.11) is 0, and arises from having $|S|/2$ points equally spaced around the unit circle, if we interpret the system as in Figure 8.1.

Now, for the maximum, (8.10) gives that the maximum of σ_A with $A \in \mathcal{A}(t,t)$ is

$$\sqrt{\frac{1}{t}\Big(|S| - 2t - \frac{|S|}{t} + 2\Big)^2 + \frac{t-1}{t}\Big(-\frac{|S|}{t} + 2\Big)^2},$$

which upon simplification is

$$\sqrt{t-1}\left(\frac{|S|}{t} - 2\right). \tag{8.15}$$

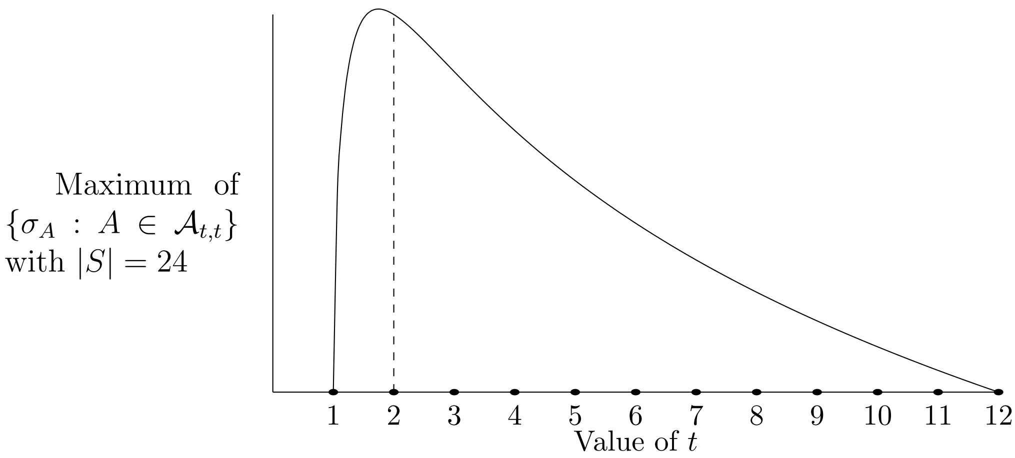

Figure 8.3. The graph illustrates how the maximum of the standard deviation for sets in $\mathcal{A}(s,t)$ varies when $|S| = 24$ and $t = s$. The latter condition is equivalent to requiring that the components of any set $A \in \mathcal{A}_{t,s}$ consist of single points. In accordance with (8.15), the graph is of the function

$$t \longmapsto \sqrt{t-1}\left(\frac{24}{t} - 2\right).$$

If we restrict ourselves to discrete values only, as needed in a discrete context, the function maximum is at $t = 2$.

Figure 8.3 illustrates how the maximum of the standard deviation varies when $|S| = 24$ and $t = s$. In the general case, still with $t = s$ but where $|S|$ is given but arbitrary, the maximum occurs again for $t = 2$. An intuitive interpretation is as follows. When $t = 1$, the recurrence time is constant and so the standard deviation is 0. When one point is added to give $t = 2$, one point has recurrence time 2, the other has recurrence time $|S| - 2$. As more points are added the recurrence times decrease, and there is “less room” to deviate from the average value. More precisely, the maximum of the function

$$t \longmapsto \sqrt{t-1}\Big(\frac{|S|}{t} - 2\Big),$$

occurs at $(-|S| + \sqrt{|S|^2 + 16|S|})/4$, which is less than 2 and approaches 2 as $|S| \longrightarrow \infty$. So, when considering only discrete values, the maximum must occur at $t = 2$ regardless of the value of $|S|$.

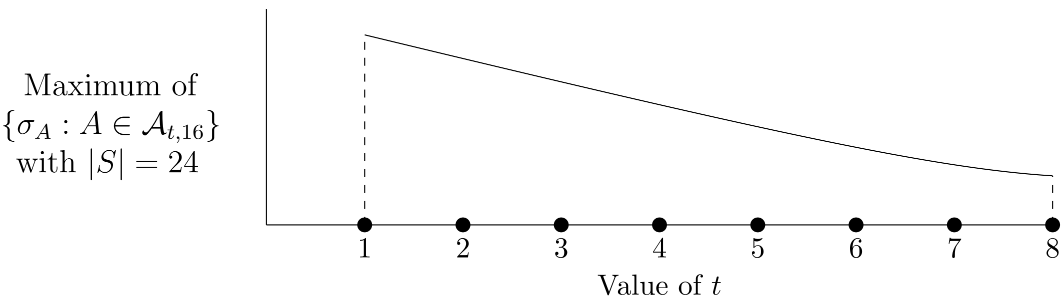

Figure 8.4. The graph is of the function

$$t \longmapsto \sqrt{\frac{1}{4}\Big(1-\frac{t}{16}\Big)+\frac{1}{16}\Big(\frac{17}{2}-t\Big)+\frac{1}{4}\Big(\frac{t-1}{16}\Big)},$$

which comes from (8.10) taking $|S| = 24$ and $s = 16$. It shows how the maximum of σ_A over subsets A with $|A| = 16$ and A having t gaps decreases as more gaps are allowed in the decomposition of the set A.

Now, let S be a given finite set, and let $s \in \{2, 3, \ldots, |S|\}$ be given. Considering (8.10) the maximum of σ_A over sets $A \in \mathcal{A}(t, s)$ is given by the function f where

$$f(t) = \sqrt{\Big(1-\frac{t}{s}\Big)\Big(1-\frac{|S|}{s}\Big)^2+\frac{1}{s}\Big(|S|-s-t-\frac{|S|}{s}+2\Big)^2+\frac{t-1}{s}\Big(-\frac{|S|}{s}+2\Big)^2}.$$

Allowing t for the moment to take on real values, a calculation of the derivative shows that

$$f'(t) = -\frac{1}{2sf(t)}\Big(1+2(|S|-s-t)\Big).$$

This implies that f is decreasing as a function of t, because $s+t \leq |S|$ and so $f'(t) \leq 0$. Intuitively, we can interpret this as follows. As t increases, there are more gaps in the decomposition of sets in $\mathcal{A}(t, s)$. This generally causes the recurrence times to decrease, and they have a narrower range of values which causes the standard deviation to decrease. Alternatively, as t increases the sets in $\mathcal{A}(t, s)$ become more “ spread out” in S so the recurrence times have a “less extreme” behavior. Even so, it may happen if $A \in \mathcal{A}(t, s)$ and $B \in \mathcal{A}(t+1, s), \sigma_A < \sigma_B$, as illustrated in the following example.

Example 8.4.1.

Let $A \in \mathcal{A}(2, 3)$ and $B \in \mathcal{A}(3, 3)$ with $|S| = 18$ (see Figure 8.5). The recurrence times of $A \in \mathcal{A}(2, 3)$ are $1, \ell_1 + 1, \ell_2 + 1$, hence $\ell_1 + \ell_2 + 3 = 18$ so

$\ell_1 + \ell_2 = 15$. Note that by using Theorem 8.3.2 the average of the recurrence time is 6. Thus, the standard deviation of r_A over A is

$$\begin{aligned} \sigma_A = & \sqrt{\frac{1}{3}\Big((1-6)^2 + (\ell_1+1-6)^2 + (\ell_2+1-6)^2\Big)} \\ = & \sqrt{\frac{1}{3}\Big(25 + (\ell_1-5)^2 + (\ell_2-5)^2\Big)}, \end{aligned}$$

if we take $\ell_1 = 7$ and $\ell_2 = 8$, then

$$\begin{aligned} \sigma_A = & \sqrt{\frac{1}{3}(25+4+9)} \\ = & \sqrt{\frac{38}{3}}. \end{aligned}$$

Now, the recurrence times of $B \in \mathcal{A}(3,3)$ are $2, 2, k+1$ (see Figure 8.5), hence $k + 2 + 3 = 18$ so $k = 13$. The average of the recurrence time is also 6. Thus, the standard deviation of r_B over B is

$$\begin{aligned} \sigma_B = & \sqrt{\frac{1}{3}\Big((2-6)^2 + (2-6)^2 + (14-6)^2\Big)} \\ = & \sqrt{\frac{1}{3}(32+64)} \\ = & 4\sqrt{2}. \end{aligned}$$

Therefore, $\sigma_A < \sigma_B$.

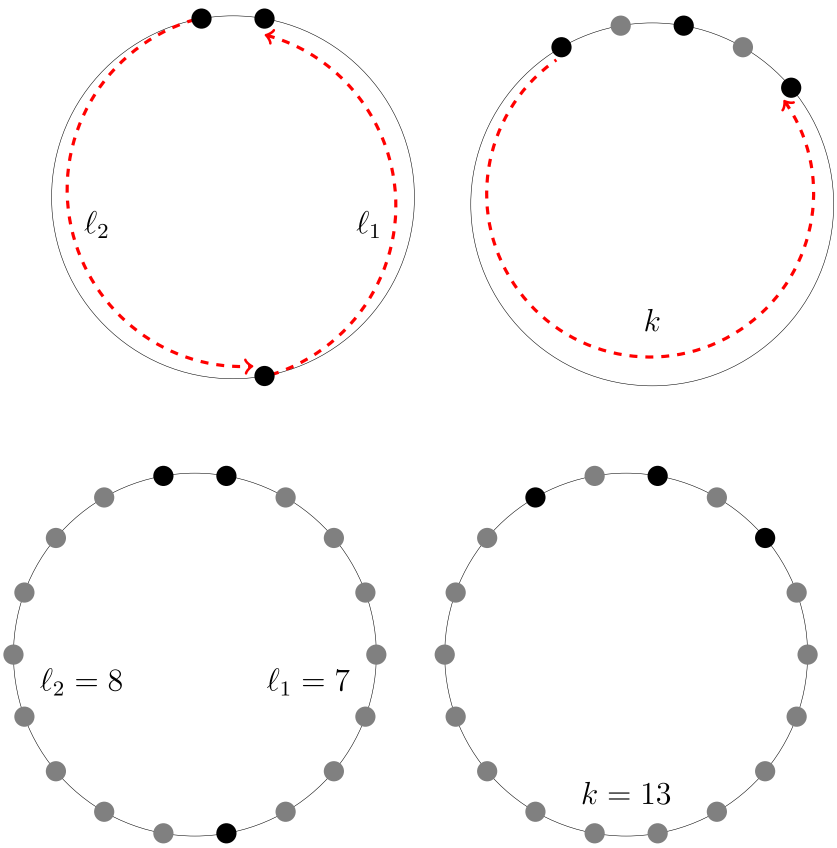

Figure 8.5. At the top left of the Figure, the black points define a subset A of S and $A \in \mathcal{A}(2,3)$. The recurrence times for points in A are $1, \ell_1+1, \ell_2+1$. At the top right, the black points define a subset $B \in \mathcal{B}(3,3)$ of S, whose points have a recurrence time of $2, 2, k+1$. However, the bottom left and right of the Figure illustrate the special cases where we take $\ell_1 = 7, \ell_2 = 8$ with $A \in \mathcal{A}(2,3)$ and $k = 13$ with $B \in \mathcal{B}(3,3)$.

Now, consider that happens when the number of gaps is kept fixed, but the number $|A|$ of elements in a subset A of S varies. Figure 8.6 illustrates what happens to the minimum of the standard deviation over set in $\mathcal{A}(6,s)$ when we consider the case where $|S| = 48$. In general, if t is given by (8.11) the minimum of σ_A taken over sets $A \in \mathcal{A}(t,s)$ is the minimum of the function

$$s \longmapsto \left(\frac{|S|}{s} - 1\right)\sqrt{\frac{s}{t} - 1},$$

taken over $t \leq s \leq |S| - t$.

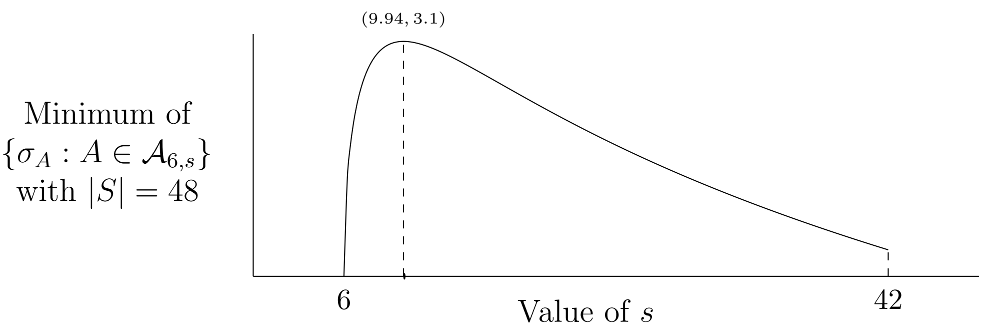

Figure 8.6 The graph is of the function

$$s \longmapsto \left(\frac{48}{s} - 1\right) \sqrt{\frac{s}{6} - 1},$$

which comes from (8.11) taking $|S| = 48$ and $t = 6$. For $s \in \{6, 12, 18, 24, 30, 36, 42\}$ the function value is the minimum of σ_A, over subsets $A \in \mathcal{A}_{6,s}$. So, the graph shows how the minimum of the standard deviation σ_A, taken over $A \in \mathcal{A}_{6,s}$, varies with s. Note that because we only consider subsets having 6 components, we must have $s \leq 42$.

A calculation shows the minimum occurs at

$$s_0(t) = \frac{4t|S|}{|S| + \sqrt{|S|^2 + 8|S|t}}. \tag{8.16}$$

A way of thinking about this is that if a set $A \in \mathcal{A}_{t,s}$ has a small number of elements and a small standard deviation, as points are added one-by-one to the set, the minimum possible standard deviation may increase for a while but, as points continue to be added, the set becomes more "crowded", the recurrence times are reduced regardless of how the points are added, and so beyond the point $s_0(t)$ in (8.16) the minimum standard deviation decreases.

Chapter 9

Standard deviation of recurrence times in infinite discrete systems

9.1 Introduction

In Chapter 8, we considered some questions concerning the standard deviation in finite dynamical systems. In this Chapter, we consider further a class of dynamical systems, arising from an infinite "sum" of finite systems. Whereas for finite systems the standard deviation of a recurrence time is finite, in the case considered here the standard deviation of a recurrence time may be infinite. The technique depends upon finding the standard deviation of a countable sum of finite systems in terms of the standard deviations in the individual systems. Any finite dynamical system that arising from a permutation on a finite set is a "sum" of subsystems on which the permutation acts cyclically. So, in this sense, the work in this Chapter encompasses the case of an arbitrary permutation of a finite set.

9.2 Finite sums of dynamical systems

In this Section, we motivate later work by considering a special case of two finite systems, and introduce some ideas for later use. Suppose (S_1, f_1) and (S_2, f_2) are two finite dynamical systems. We assume that the transformations $f_1 : S_1 \longrightarrow S_1$ and $f_2 : S_2 \longrightarrow S_2$ act as cyclic permutations on S_1 and S_2 respectively. Then, f_1 and f_2 are one-to-one and onto. We assume

$S_1 \cap S_2 = \emptyset$, and put $S = S_1 \cup S_2$. We define $f : S \longrightarrow S$ by

$$f(x) = \begin{cases} f_1(x), & \text{if } x \in S_1, \\ f_2(x), & \text{if } x \in S_2. \end{cases}$$

Then, (S, f) is a dynamical system. Note that f is not a cyclic permutation on S but is one-to-one and onto.

We define a probability P_1 on the algebra of all subsets of S_1 by putting

$$P_1(A) = \frac{|A|}{|S_1|}, \text{ for any } A \subseteq S_1,$$

where $|A|$ is the size of the set A. Note that $P(\{x\}) = 1/|S_1|$ for all $x \in S_1$. Using the fact that f_1 is a one-to-one and onto on S_1, we will show

$$P_1\Big(f_1^{-1}(A)\Big) = P_1(A), \text{ for all } A \subseteq S.$$

This is because

$$\begin{aligned} \Big|P_1\Big(f_1^{-1}(A)\Big)\Big| &= \frac{|f_1^{-1}(A)|}{|S_1|} \\ &= \frac{|A|}{|S_1|} \\ &= P_1(A). \end{aligned}$$

Then, f_1 is "measure preserving" on S_1 (a formal definition is below). Similarly, we define P_2 by

$$P_2(A) = \frac{|A|}{|S_2|}, \text{ for any } A \subseteq S_2,$$

and the corresponding property holds and f_2 is "measure preserving" on S_2.

We now define a probability P on the family of all subsets of S by putting, for $A \subseteq S$,

$$P(A) = \frac{1}{2}\Big[P_1(A \cap S_1) + P_2(A \cap S_2)\Big].$$

Note that $P(S) = 1$. Observe that, for $A \subseteq S$,

$$f_1^{-1}(A \cap S_1) \subseteq S_1 \text{ and } f_2^{-1}(A \cap S_2) \subseteq S_2,$$

and so as f_1 and f_2 are "measure preserving",

$$P_1\Big(f_1^{-1}(A \cap S_1)\Big) = P_1(A \cap S_1),$$

and similarly for P_2,

$$P_2\Big(f_2^{-1}(A \cap S_2)\Big) = P_2(A \cap S_2).$$

Also,

$$f^{-1}(A) = f_1^{-1}(A \cap S_1) \cup f_2^{-1}(A \cap S_2).$$

Hence,

$$\begin{aligned} P\Big(f^{-1}(A)\Big) &= \frac{1}{2}\Big[P_1\Big(f_1^{-1}(A \cap S_1)\Big) + P_2\Big(f_2^{-1}(A \cap S_2)\Big)\Big] \\ &= \frac{1}{2}\Big[P_1(A \cap S_1) + P_2(A \cap S_2)\Big], \text{ as } f_1 \text{ and } f_2 \text{ are "measure preserving"}, \\ &= P(A), \text{ by using the definition of } P. \end{aligned}$$

We now consider $P(0^m)$ for $m \geq 1$. By definition this is $P(B_m)$ where

$$B_m = \{x : x \in A^c, f(x) \in A^c, \ldots, f^{m-1}(x) \in A^c\}.$$

Now,

$$P(B_m) = \frac{1}{2}\Big[P_1(B_m \cap S_1) + P_2(B_m \cap S_2)\Big].$$

Observe that if $A = \emptyset, B_m = S$ for all $m \in \mathbb{N}$. If $A = S, A^c = \emptyset$ and we see that $B_m = \emptyset$ for all $m \in \mathbb{N}$. So, assume that $A \neq S$ and $A \neq \emptyset$. Assume that $A \cap S_1 \neq \emptyset$ and let $x \in A \cap S_1$. As f_1 acts as a cyclic permutation on S_1, there is $0 \leq k \leq |S_1| - 1$ such that $f^k(x) \in A \cap S_1$. Then, $x \notin B_{k+1}$. Hence, if $x \in A \cap S_1$ we see that $x \notin B_{|S_1|-1}$. So, if $m \geq |S_1| - 1$, then

$$S_1 \cap B_m = \emptyset. \tag{9.1}$$

Similarly, we find that if $A \cap S_2 \neq \emptyset$ and $m \geq |S_2| - 1$, then

$$S_2 \cap B_m = \emptyset. \tag{9.2}$$

Now, if $A \subseteq S_1$ and $A \neq \emptyset$, as above we have, for $m \geq |S_1|-1, S_1 \cap B_m = \emptyset$. However, if $x \in S_2$, then

$$x \notin A, f_2(x) \notin A, \ldots, f_2^k(x) \notin A, \ldots,$$

and we see that

$$S_2 \subseteq B_m, \text{ for all } m \in \mathbb{N}. \tag{9.3}$$

Similarly, if $A \subseteq S_2$ and $A \neq \emptyset$, we have $S_2 \cap B_m = \emptyset$, for $m \geq |S_2| - 1$, and

$$S_1 \subseteq B_m, \text{ for all } m \in \mathbb{N}. \tag{9.4}$$

Summarizing, the above may be immediately deduced from (9.1), (9.2), (9.3), and (9.4):

$$\text{if } A = \emptyset, \text{ then } P(B_m) = 1, \text{ for all } m \in \mathbb{N},$$

or

$$\text{if } A = S, \text{ then } P(B_m) = 0, \text{ for all } m \in \mathbb{N},$$

or

$$\text{if } A \neq \emptyset, \text{ and } A \subseteq S_1 \text{ with } m \geq |S_1| - 1, \text{then } P(B_m) = 1/2, \text{ for all } m \in \mathbb{N},$$

or

$$\text{if } A \neq \emptyset, \text{ and } A \subseteq S_2, \text{ with } m \geq |S_2| - 1, \text{ then } P(B_m) = 1/2, \text{ for all } m \in \mathbb{N},$$

or if $A \neq \emptyset$, and $A \cap S_1 \neq \emptyset$ and $A \cap S_2 \neq \emptyset$, then

$$P(0^m) = P(B_m) = 0, \text{ for all } m \geq \max\{|S_1| - 1, |S_2| - 1\}.$$

A similar argument applies to any finite collection of finite dynamical systems. When we consider a finite number of dynamical systems, an investigation along the preceding lines gives Theorem 9.2.1 below. We first give a formal definition of a measure preserving transformation for our context.

Definition 9.2.1.

Let S be set and suppose P is a probability on the algebra of all subsets of S. In this case, we may simply say that P is a *probability* on S. Let $f : S \longrightarrow S$ be a transformation on S. Then, f is called *measure preserving* if

$$P\Big(f^{-1}(A)\Big) = P(A), \text{ for all } A \subseteq S.$$

This is a special case of the definition in Halmos [14, page 164], for example.

Note that if A is a finite set, if P is a probability on S, and $f : S \longrightarrow S$ acts as cyclic permutation on S, then f is measure preserving on S, and we must have

$$P\Big(\{x\}\Big) = \frac{1}{|S|}, \text{ for all } x \in S.$$

Note that this idea of measure preserving is exactly the same as the previous notion of *P-invariant* as saying that P is *P-invariant* on the algebra of all subsets of S, as in Definition 3.4.4 in Chapter 3.

Theorem 9.2.1. *Suppose (S_j, f_j) are finite dynamical systems for $j = 1, \ldots, n$ such that the sets S_j are disjoint and each f_j is a cyclic permutation on S_j. Suppose P_j is a probability on S_j such that f_j is measure preserving. Put $S = S_1 \cup S_2 \cup \cdots \cup S_n$, and define $f : S \longrightarrow S$ by*

$$f(x) = f_j(x), \text{ for all } x \in S_j.$$

Suppose $\alpha_1, \alpha_2, \ldots, \alpha_n$ are positive numbers with $\alpha_1 + \alpha_2 + \ldots + \alpha_n = 1$, and for any set $A \subseteq S$ put

$$P(A) = \sum_{j=1}^{n} \alpha_j P_j(A \cap S_j).$$

Then, (S, f) is a finite dynamical system, P is a probability on the family of all subsets S, and f is measure preserving on S. Further, if $m \geq \max\{|S_1| - 1, |S_2| - 1, \ldots, |S_n| - 1\}$, then

$$P(0^m) = \sum_{A \cap S_j = \emptyset} \alpha_j.$$

Proof. We give a proof of the last statement. We have

$$\begin{aligned}
P(0^m) &= P\Big(\{x : x \in A^c, f(x) \in A^c, \ldots, f^{m-1}(x) \in A^c\}\Big) \\
&= P\Big(\bigcup_{j=1}^{n}\{x : x \in S_j, x \in A^c, f(x) \in A^c, \ldots, f^{m-1}(x) \in A^c\}\Big) \\
&= \sum_{j=1}^{n} \alpha_j P_j\Big(\{x : x \in S_j, x \in A^c, f(x) \in A^c, \ldots, f^{m-1}(x) \in A^c\}\Big) \\
&= \sum_{j=1}^{n} \alpha_j P_j\Big(\{x : x \in S_j, x \in S_j \cap A^c, f(x) \in S_j \cap A^c, \ldots, f^{m-1}(x) \in S_j \cap A^c\}\Big).
\end{aligned} \tag{9.5}$$

Now, by above if $S_j \cap A \neq \emptyset$ and $m \geq |S_j| - 1$, then

$$P_j\Big(\{x : x \in S_j, x \in S_j \cap A^c, f(x) \in S_j \cap A^c, \ldots, f^{m-1}(x) \in S_j \cap A^c\}\Big) = \emptyset.$$

From (9.5) we see that if $m \geq \max\{|S_1| - 1, |S_2| - 1, \ldots, |S_n| - 1\}$,

$$\begin{aligned}
P(0^m) &= \sum_{A \cap S_j = \emptyset} \alpha_j P_j(S_j) \\
&= \sum_{A \cap S_j = \emptyset} \alpha_j, \text{ by using the fact that } P_j(S_j) = 1.
\end{aligned}$$

□

9.3 Countable sums of finite dynamical systems

We now look at the case of an infinite but countable collection of systems.

Definition 9.3.1.

Let S_j be a non-empty finite set, for each $j \in \mathbb{N}$, and let f_j be a cyclic permutation on the whole of S_j. Suppose further that $S_i \cap S_j = \emptyset$ for all $i \neq j$, let $S = \bigcup_{j\in\mathbb{N}} S_j$, and define $f : S \longrightarrow S$ by $f(x) = f_j(x)$ for all $x \in S_j$. Then, (S, f) is called the *sum* of the dynamical systems (S_j, f_j). Note that each system (S_j, f_j) is of the type considered in the previous Chapter.

Definition 9.3.2.

Suppose we are given a sequence (w_j), with $w_j > 0$ for all j and $\sum_{j\in\mathbb{N}} w_j = 1$. We defined a *weighted* probability P on S by putting

$$P(A) = \sum_{j\in\mathbb{N}} w_j P_j(A \cap S_j), \text{ for } A \subseteq S,$$

where $P_j(A \cap S_j) = |A \cap S_j|/|S_j|$. It is easy to see that P has the properties expected of a probability, in this context. That is we have:

(i) If $A \subseteq S$, then $0 \leq P(A) \leq 1$.

(ii) If $A_1, A_2, \ldots$ are disjoints subsets of S then $P\Big(\bigcup_{j=1}^{\infty} A_j\Big) = \sum_{j=1}^{\infty} P(A_j)$.

(iii) $P(S) = 1$.

Note that for all points $x \in S_j$, $P\Big(\{x\}\Big)$ has the same value. That is,

$$P\Big(\{x\}\Big) = \frac{w_j}{|S_j|}, \text{ for all } x \in S_j. \tag{9.6}$$

Now we will show that f is measure preserving. Note that

$$f^{-1}(A \cap S_j) \subseteq S_j, \text{ and } f^{-1}(A) = \bigcup_{j=1}^{\infty} f^{-1}(A \cap S_j),$$

which is a disjoint union. We have

$$\begin{aligned}
f^{-1}(A) \cap S_j &= f^{-1}(A \cap S_j), \text{ so that} \\
P\Big(f^{-1}(A)\Big) &= \sum_{j \in \mathbb{N}} w_j P_j\Big(f^{-1}(A) \cap S_j\Big) \\
&= \sum_{j \in \mathbb{N}} w_j P_j\Big(f_j^{-1}(A \cap S_j)\Big) \\
&= \sum_{j \in \mathbb{N}} w_j P_j(A \cap S_j), \text{ as } f_j \text{ is measure preserving on } S_j, \\
&= P(A), \text{ by using the definition.}
\end{aligned}$$

Thus, f is measure preserving. □

Note that for $x \in A$ we define the recurrence time $r_A(x)$ as in (8.4). That is,

$$r_A(x) = \min\{n \in \mathbb{N} \text{ and } f^n(x) \in A\}.$$

Also, note that the *expectation* or *average* of r_A over A is given by

$$E_A(r_A) = \frac{1}{P(A)} \sum_{x \in A} r_A(x) P\Big(\{x\}\Big), \tag{9.7}$$

as in [11, page 207].

The following result demonstrates the validity of Kac's formula in the case of finite and infinite sums of finite dynamical systems.

Theorem 9.3.1. *Let Q denote either the set $\mathbb{N}$ or a finite set. For each $j \in Q$, let S_j be a non-empty finite set and let f_j be a cyclic permutation on S_j such that $S_i \cap S_j = \emptyset$ for all $i \neq j$. Let (S, f) be the sum of the dynamical systems $(S_j, f_j)_{j \in Q}$. Let weights $(w_j)_{j \in Q}$ be given as in the definition above, and let the probability of a subset A of S be given by*

$$P(A) = \sum_{j \in Q} w_j P_j(A \cap S_j).$$

Then, if $A \subseteq S$ is such that $A \cap S_j \neq \emptyset$, for all $j \in Q$, the average $E_A(r_A)$ of r_A over A is given by

$$E_A(r_A) = \frac{1}{P(A)}.$$

Further, if $A \subseteq S$ is such that $A \cap S_n = \emptyset$, for some $n \in Q$, then

$$E_A(r_A) < \frac{1}{P(A)}.$$

Proof. If $A \subseteq S$ and $A \neq \emptyset$, put $Y = \{j : j \in Q \text{ and } A \cap S_j \neq \emptyset\}$. Then, by (9.7),

$$\begin{aligned} E_A(r_A) &= \frac{1}{P(A)} \sum_{x \in A} r_A(x) P\Big(\{x\}\Big) \\ &= \frac{1}{P(A)} \sum_{j \in Y} \sum_{x \in A \cap S_j} r_A(x) P\Big(\{x\}\Big) \\ &= \frac{1}{P(A)} \sum_{j \in Y} \sum_{x \in A \cap S_j} r_A(x) \frac{w_j}{|S_j|}, \text{ by using (9.6),} \end{aligned}$$

$$\begin{aligned} &= \frac{1}{P(A)} \sum_{j \in Y} \frac{w_j}{|S_j|} \Bigg(\sum_{x \in A \cap S_j} r_{A \cap S_j}(x) \Bigg) \\ &= \frac{1}{P(A)} \sum_{j \in Y} w_j, \text{ by using Theorem 8.3.2.} \end{aligned} \tag{9.8}$$

Now, there are two possibilities. If $A \cap S_j \neq \emptyset$, for all $j \in Q$, then in this case $\sum_{j \in Y} w_j = 1$, so (9.8) implies that $E_A(r_A) = 1/P(A)$. However, if $A \cap S_n = \emptyset$, for some $n \in Q$, then (9.8) gives $E_A(r_A) < 1/P(A)$. Thus, Kac's formula remains true if $A \cap S_j \neq \emptyset$, for all $j \in Q$, otherwise it fails.

□

9.4 Recurrence and standard deviation in a sum of systems

Let Q denote either the set $\mathbb{N}$ or a finite set. For each $j \in Q$, let S_j be a non-empty finite set and let f_j be a cyclic permutation on S_j. Assume that $S_i \cap S_j = \emptyset$ for all $i, j \in Q$ with $i \neq j$. Let (S, f) be the sum of the dynamical systems $(S_j)_{j \in Q}$. Let $w_j > 0$ for all $j \in Q$, such that $\sum_{j \in Q} w_j = 1$. Then, for $A \subseteq S$, put

$$P(A) = \sum_{j \in Q} w_j \frac{|A \cap S_j|}{|S_j|} \text{ and } P_j(A) = \frac{|A \cap S_j|}{|S_j|}.$$

Let r_A denote the recurrence time over A in the system (S, f). The *standard deviation* of r_A in (S, f) is defined by σ_A, where

$$\sigma_A^2 = \frac{1}{P(A)} \sum_{x \in A} \Big| r_A(x) - E_A(r_A) \Big|^2 P\Big(\{x\}\Big), \tag{9.9}$$

as in [11, page 213], where $E_A(r_A)$ given by (9.7). We have seen in (9.8) that if $P(A) > 0$,

$$E_A(r_A) = \frac{1}{P(A)} \sum_{A\cap S_j \neq \emptyset} w_j.$$

So, (9.9) gives

$$\sigma_A^2 = \frac{1}{P(A)} \sum_{x\in A} \left| r_A(x) - \frac{1}{P(A)} \sum_{k, A\cap S_k \neq \emptyset} w_k \right|^2 P\Big(\{x\}\Big). \tag{9.10}$$

It follows from (9.10) that if $A \subseteq S$ and $j \in Q$ with $A \cap S_j \neq \emptyset$, then in the system (S, f)

$$\begin{aligned}
\sigma_{A\cap S_j}^2 &= \frac{1}{P(A\cap S_j)} \sum_{x\in A\cap S_j} \left| r_A(x) - \frac{|S_j| w_j}{w_j |A\cap S_j|} \right|^2 \frac{w_j}{|S_j|} \\
&= \frac{|S_j|}{w_j |A\cap S_j|} \left(\sum_{x\in A\cap S_j} \Big(r_A(x) - \frac{|S_j|}{|A\cap S_j|} \Big)^2 \right) \frac{w_j}{|S_j|} \\
&= \frac{1}{|A\cap S_j|} \sum_{x\in A\cap S_j} \Big(r_A(x) - \frac{|S_j|}{|A\cap S_j|} \Big)^2.
\end{aligned} \tag{9.11}$$

Now, note that if $j \in Q$ is given, the system (S_j, f_j) is of type (S, f) where S is finite and f is a cyclic permutation of S. Note in this case that if $A \subseteq S$ and $A \cap S_j \neq \emptyset$, then

$$r_A(x) = r_{A\cap S_j}(x), \text{ for all } x \in A \cap S_j.$$

Using the definition in (9.9), now for $A \cap S_j$ in place of A and (S_j, f_j) in place of (S, f) we now find that if $A \cap S_j \neq \emptyset$, and denoting the expectation and standard deviation of $r_{A\cap S_j}$ in (S_j, f_j) by $E_{A\cap S_j}(r_{A\cap S_j})$ and $\sigma_{A\cap S_j}$ respectively,

$$E_{A\cap S_j}(r_{A\cap S_j}) = \frac{1}{P_j(A\cap S_j)} = \frac{|S_j|}{|A\cap S_j|} = \frac{w_j}{P(A\cap S_j)}, \tag{9.12}$$

and

$$\begin{aligned}
\sigma^2_{A\cap S_j} &= \frac{1}{P_j(A\cap S_j)} \sum_{x\in A\cap S_j} \left| r_A(x) - \frac{|S_j|}{|A\cap S_j|}\right|^2 P_j\Big(\{x\}\Big) \\
&= \frac{|S_j|}{A\cap S_j}\frac{1}{|S_j|} \sum_{x\in A\cap S_j} \left| r_A(x) - \frac{|S_j|}{|A\cap S_j|}\right|^2 \\
&= \frac{1}{|A\cap S_j|} \sum_{x\in A\cap S_j} \left| r_A(x) - \frac{|S_j|}{|A\cap S_j|}\right|^2 . \qquad (9.13)
\end{aligned}$$

Now, (9.8) and (9.12) show that the expectation of r_A over $A \cap S_j$ is the same in either system (S, f) or (S_j, f_j). Consequently, the expressions $E_{A\cap S_j}(r_A)$ has no ambiguity.

Also, (9.11) and (9.13) show that the expression $\sigma^2_{A\cap S_j}$ is the same, in either system (S, f) or (S_j, f_j). Then, $\sigma^2_{A\cap S_j}$ has no ambiguity.

The following result expresses the relationship between the standard deviation σ_A of recurrence times in the system (S, f), which is a sum of systems (S_j, f_j), in terms of the standard deviations $\sigma_{A\cap S_j}$ of the component systems (S_j, f_j).

Theorem 9.4.1. *Let Q denote either the set $\mathbb{N}$ or a finite set. For each $j \in Q$, let S_j be a non-empty finite set and let f_j be a cyclic permutation on the whole of S_j such that $S_i \cap S_j = \emptyset$ for all $i \neq j$. Let (S, f) be the sum of the dynamical systems $(S_j, f_j)_{j\in Q}$. Let weights $(w_j)_{j\in Q}$ be given as in Definition 9.3.2, and let the probability of a subset A of S be given by*

$$P(A) = \sum_{j\in Q} w_j P_j(A\cap S_j).$$

Then, if $A \subseteq S$ is such that $A \cap S_j \neq \emptyset$ for all $j \in Q$, the standard deviation of r_A over A is given by σ_A, where

$$\sigma^2_A = \frac{1}{P(A)}\left(\sum_{j\in Q} w_j \frac{|A\cap S_j|}{|S_j|}\sigma^2_{A\cap S_j}\right) + \frac{1}{P(A)}\sum_{j\in Q} w_j \frac{|S_j|}{|A\cap S_j|} - \frac{1}{P(A)^2}. \qquad (9.14)$$

Proof. By using (9.9) and Theorem 9.3.1 we have

$$\begin{aligned}
\sigma_A^2 &= \frac{1}{P(A)}\sum_{x\in A}\left|r_A(x)-\frac{1}{P(A)}\right|^2 P\Big(\{x\}\Big)\\
&= \frac{1}{P(A)}\sum_{x\in A}\left(r_A^2(x)-\frac{2r_A(x)}{P(A)}+\frac{1}{P(A)^2}\right)P\Big(\{x\}\Big)\\
&= \frac{1}{P(A)}\sum_{x\in A}r_A^2(x)P\Big(\{x\}\Big)-\frac{2}{P(A)^2}\sum_{x\in A}r_A(x)P\Big(\{x\}\Big)+\frac{1}{P(A)^3}\sum_{x\in A}P\Big(\{x\}\Big)\\
&= \frac{1}{P(A)}\sum_{x\in A}r_A^2(x)P\Big(\{x\}\Big)-\frac{2}{P(A)^2}+\frac{1}{P(A)^2},\ \text{by using Theorem 9.3.1, (9.6), and (9.7),}\\
&= \frac{1}{P(A)}\sum_{x\in A}r_A^2(x)P\Big(\{x\}\Big)-\frac{1}{P(A)^2}\\
&= \frac{1}{P(A)}\sum_{j\in Q}\left(\sum_{x\in A\cap S_j}\frac{w_j r_A^2(x)}{|S_j|}\right)-\frac{1}{P(A)^2}\\
&= \frac{1}{P(A)}\sum_{j\in Q}\frac{w_j}{|S_j|}\left(\sum_{x\in A\cap S_j}r_A^2(x)\right)-\frac{1}{P(A)^2}\\
&= \frac{1}{P(A)}\sum_{j\in Q}\frac{w_j}{|S_j|}\left(\sum_{x\in A\cap S_j}\left[\left(r_A(x)-\frac{|S_j|}{|A\cap S_j|}\right)^2\right]\right.\\
&\qquad\left.+\frac{2|S_j|}{|A\cap S_j|}\sum_{x\in A\cap S_j}r_A(x)-\sum_{x\in A\cap S_j}\frac{|S_j|^2}{|A\cap S_j|^2}\right)-\frac{1}{P(A)^2}\\
&= \frac{1}{P(A)}\sum_{j\in Q}\frac{w_j}{|S_j|}\left(\sum_{x\in A\cap S_j}\left[\left(r_A(x)-\frac{|S_j|}{|A\cap S_j|}\right)^2\right]+\frac{2|S_j|^2}{|A\cap S_j|}-\frac{|S_j|^2}{|A\cap S_j|}\right)-\frac{1}{P(A)^2}\\
&= \frac{1}{P(A)}\sum_{j\in Q}\frac{w_j}{|S_j|}\left(\sum_{x\in A\cap S_j}\left[\left(r_A(x)-\frac{|S_j|}{|A\cap S_j|}\right)^2\right]+\frac{|S_j|^2}{|A\cap S_j|}\right)-\frac{1}{P(A)^2}\\
&= \frac{1}{P(A)}\left(\sum_{j\in Q}w_j\frac{|A\cap S_j|}{|S_j|}\sigma^2_{A\cap S_j}\right)+\frac{1}{P(A)}\sum_{j\in Q}w_j\frac{|S_j|}{|A\cap S_j|}-\frac{1}{P(A)^2}.
\end{aligned}$$

□

Note that by Theorem 9.3.1 if $A\neq\emptyset$ the average of r_A over A is finite. However, the standard deviation may be infinite, as we shall see.

Theorem 9.4.2. *For each $j \in \mathbb{N}$, let (S_j, f_j) be a finite dynamical system where f_j acts as a cyclic permutation on S_j. Assume that the sets S_j are disjoint and let (S, f) be the sum of the systems (S_j, f_j). Let w_j be a sequence of positive numbers such that $\sum_{j=1}^{\infty} w_j = 1$, and let P be the probability on S given by*

$$P(A) = \sum_{j=1}^{\infty} w_j \frac{|A \cap S_j|}{|S_j|}, \text{ for all } A \subseteq S_j.$$

Let $A \subseteq S$ be such that $|A \cap S_j| = 1$, for all $j \in \mathbb{N}$. Put, for all $m = 1, 2, \dots$

$$0^m = \{x : x \in A^c, f(x) \in A^c, \dots, f^{m-1}(x) \in A^c\}.$$

Then, $\lim_{m \longrightarrow \infty} P(0^m) = 0$, and

$$\sum_{m=1}^{\infty} P(0^m) = -\frac{1}{2} + \frac{1}{2} \sum_{m=1}^{\infty} w_j |S_j|.$$

Also,

$$\sigma_A^2 = \frac{1}{P(A)} \sum_{j \in \mathbb{N}} w_j |S_j| - \frac{1}{P(A)^2}, \text{ for all } A \subseteq S_j. \tag{9.15}$$

Proof. We have, as $S = \bigcup_{j=1}^{\infty} S_j$

$$P(0^m) = \sum_{j=1}^{\infty} P(0^m \cap S_j). \tag{9.16}$$

Now, $A \cap S_j$ consists of a single point, so we see that

$$|0^1 \cap S_j| = |S_j| - 1, |0^2 \cap S_j| = |S_j| - 2, \dots, |0^m \cap S_j| = |S_j| - m.$$

In general,

$$|0^m \cap S_j| = \max\{|S_j| - m, 0\}.$$

Hence, by (9.16)

$$P(0^m) = \sum_{m=1}^{\infty} w_j \frac{\max\{|S_j| - m, 0\}}{|S_j|}. \tag{9.17}$$

Using (9.17) we have

$$\begin{aligned} P(0^m) &= \sum_{j, |S_j| \geq m} \frac{w_j}{|S_j|} \big(|S_j| - m\big) \\ &= \sum_{j, |S_j| \geq m} w_j \big(1 - \frac{m}{|S_j|}\big) \\ &\leq \sum_{j, |S_j| \geq m} w_j. \end{aligned} \tag{9.18}$$

Now, let $\varepsilon > 0$ and let M_0 be such that

$$\sum_{j=M_0+1}^{\infty} w_j < \varepsilon. \tag{9.19}$$

Put

$$K_m = \{j : j \in \mathbb{N} \text{ and } |S_j| \geq m\}.$$

Then, there is m_0 such that

$$m \geq m_0 \Longrightarrow K_m \subseteq \{M_0 + 1, M_0 + 2, \ldots\}.$$

Consequence, if $m \geq m_0$,

$$\sum_{j,|S_j|\geq m} w_j = \sum_{j \in K_m} w_j \leq \sum_{j=M_0+1}^{\infty} w_j < \varepsilon, \text{ by using (9.19).}$$

We deduce from (9.18) that $\lim_{m \longrightarrow \infty} P(0^m) = 0$.

Also, from (9.17)

$$\begin{aligned}
\sum_{m=1}^{\infty} P(0^m) &= \sum_{j=1}^{\infty} \sum_{m=1}^{\infty} w_j \frac{\max\{|S_j| - m, 0\}}{|S_j|} \\
&= \sum_{j=1}^{\infty} \frac{w_j}{|S_j|} \Bigg(\sum_{m=1}^{|S_j|-1} \big(|S_j| - m\big) \Bigg) \\
&= \sum_{j=1}^{\infty} \frac{w_j}{|S_j|} \sum_{k=1}^{|S_j|-1} k \\
&= \sum_{j=1}^{\infty} \frac{w_j}{|S_j|} \frac{|S_j|}{2} \big(|S_j| - 1\big) \\
&= \frac{1}{2} \sum_{j=1}^{\infty} w_j \big(|S_j| - 1\big) \\
&= -\frac{1}{2} \sum_{j=1}^{\infty} w_j + \frac{1}{2} \sum_{j=1}^{\infty} w_j |S_j| \\
&= -\frac{1}{2} + \frac{1}{2} \sum_{j=1}^{\infty} w_j |S_j|.
\end{aligned}$$

Now, consider when we have finite disjoint systems $(S_j, f_j), 1 \leq j \leq n$ and when $|S_j| = m_j$ for all $1 \leq j \leq n$ and $|A \cap S_j| = 1$ for all $j = 1, 2, \ldots, n$.

Then, $r_A(x) = m_j$ for all $x \in A \cap S_j$, and

$$\begin{aligned}\sigma^2_{A\cap S_j} &= \frac{1}{|A \cap S_j|}\sum_{x\in A\cap S_j}\left|r_A(x) - \frac{|S_j|}{|A\cap S_j|}\right|^2 \\ &= \sum_{x\in A\cap S_j}(m_j - m_j)^2 \\ &= 0.\end{aligned}$$

Now using (9.14) of Theorem 9.4.1 we get in the system (S, f),

$$\sigma^2_A = \frac{1}{P(A)}\sum_{j=1}^{n} w_j m_j - \frac{1}{P(A)^2}.$$

□

Note that this should be compared with the corresponding condition in Section 6.3 of Chapter 6.

Example 9.4.1.

A special case of Theorem 9.4.2 is when $|S_j| = j$, in which case we have

$$\sigma^2_A = \frac{1}{P(A)}\sum_{j=1}^{n} j w_j - \frac{1}{P(A)^2}.$$

In this example because the system (S, f) is finite the average of r_A over A is finite and also the standard deviation of r_A over A is finite.

Example 9.4.2.

Consider when we have an infinite sum of finite systems $(S_j, f_j)_{j\in\mathbb{N}}$, where the sets S_j are disjoint. Let (S, f) be the sum of the systems $(S_j, f_j)_{j\in\mathbb{N}}$. Let $A \subseteq S$ be such that $|A \cap S_j| = 1$ for all $j \in \mathbb{N}$. Then, $r_A(x) = |S_j|$ in (S, f), whenever $x \in A \cap S_j$. By using (9.13) we have

$$\sigma_{A\cap S_j} = 0 \text{ for all } j \in \mathbb{N}.$$

Then, by using (9.15) of Theorem 9.4.2 we have, as above,

$$\sigma^2_A = \frac{1}{P(A)}\sum_{j\in\mathbb{N}} w_j |S_j| - \frac{1}{P(A)^2}.$$

Now, if $w_j = 1/2^j$ and $|S_j| = j$, then $\sum_{j\in\mathbb{N}} w_j|S_j| < \infty$, that is $\sigma_A < \infty$. However, if we take $w_j = 1/2^j$ and $|S_j| = 2^j$, then $\sum_{j\in\mathbb{N}} w_j|S_j| = \infty$, that is $\sigma_A = \infty$. Thus, in this example, we see that in an infinite sum of finite systems, the standard deviation of the recurrence time over a subset A of S may be infinite. For this example, $P(0^m) \longrightarrow 0$ as $m \longrightarrow \infty$ but $\sum_{j=1}^{\infty} P(0^m)$ diverges. Also, in the case of an infinite sum, we do not have a counting measure but a measure with an infinite number of weights.

Chapter 10

Conclusion

Summary. In this work we have considered various aspects of recurrence times in stochastic processes and dynamical systems. The work up to and including Chapter 6 was set in the context of zero-one stochastic processes. Note that by a zero-one stochastic process is meant a sequence (X_n) of functions on a given set S where each function X_n takes values in $\{0, 1\}$. The discussion was primarily concerned with stationary processes and was a rigourous and, in some aspects, a more general discussion of some work of Kasteleyn [18]. A connection between notions of recurrence in stationary zero-one stochastic processes and dynamical systems admitting an invariant probability was established. Chapters 7, 8 and 9 presented, almost in their entirety, new results in some special dynamical systems. These results were mainly to do with the standard deviation of recurrence times.

Conclusion details. The approach in chapters 2 to 6 was based on the work of Kasteleyn, but the intention was to put Kasteleyn's work on a more rigourous and systematic mathematical basis, as well as to give some generalizations coming out of this approach. Thus, the assumptions made at various points that are needed in the proofs were identified and specific definitions made. Chapters 2 and 3 primarily set out this more structured approach, introducing such notions as probabilities, probability functions, stationary processes and the connection between dynamical systems and zero-one processes, especially in relation to recurrence phenomena. The connection between recurrence in stationary processes and in dynamical systems with an invariant probability was discussed and made explicit in a way that seems to be currently lacking in the literature.

Chapter 4 contained basic identities and preparatory material which in some cases are more general than the approach of Kastelyn (See for example, Lemma 4.2.1) and its subsequent application.) Chapter 5 developed some of the results concerning recurrence times in certain types of zero-one stochastic processes. The treatment aimed to be systematic, spelling out precise assumptions and adhering to strict standards of proof. Chapter 6 investigated the standard deviation of the recurrence times in zero-one processes. The question of the standard deviation of recurrence times seems to be little-studied in the literature and was looked at in specific detail in some dynamical systems in subsequent chapters.

Chapter 7 presented some new results and estimates for the standard deviation of recurrence times in dynamical systems (S, f), where S was a unit interval of real numbers and f was a certain type of piecewise linear transformation on S. The new results here can be compared with some of the results and discussion in Chapter 4 of [20].

Chapter 8 presented new results on the standard deviation of recurrence times in a finite dynamical system, upon which a transformation f acts as a cyclic permutation. These results show up some of the subtleties of what might be expected in more complex systems and were open, in some cases, to clarifying intuitive interpretations.

In Chapter 9, the results of Chapter 8 were used to obtain new results in discrete but infinite dynamical systems obtained as a "sum" of finite systems. Some of these results clarified the conditions needed for the standard deviation to be finite, and related such conditions to the stochastic processes context in earlier chapters, illuminating the condition $\widetilde{P}(0^\infty) = 0$ in Theorem 6.2.1. In a sense, for the systems described in this chapter, this section clarifies the relationship between the finiteness of the standard deviation and condition (4.38) in [20, page 275], by realizing condition (4.38) in a very concrete way.

Bibliography

[1] R. B. Ash. *Real analysis and probability.* Academic Press, New York 1972.

[2] J. Banks, V. Dragan, and A. Jones. *Chaos: a mathematical introduction.* Cambridge University Press, Cambridge, 2003.

[3] R. N. Bhattacharya and E. C. Waymire. *Stochastic processes with applications.* Wiley, New York 1990.

[4] J. R. Blum and J. I. Rosenblatt. On the moments of recurrence time. *Journ. Math. Sci.(Delhi)*, 2:1–6, 1967.

[5] L. Breiman. *Probability.* Addison-Wesley Pub. Co, The United States of America , 1968.

[6] J. R. Brown. *Ergodic Theory and Topological Dynamics.* Academic Press, New York 1976.

[7] W. Th. F. den Hollander. Mixing properties for random walk in random scenery. *The Annals of probabilty*, 16:1788–1802, 1988.

[8] W. Th. F. den Hollander and P. W. Kasteleyn. Random walks on lattices with points of two coloures. *Journal Phys.*, 117A:179–188, 1983.

[9] R. L. Devaney. *A first course in chaotic dynamical systems: theory and experiment.* Addison-Wesley, Massachusetts, 1992.

[10] J. L. Doob. *Stochastic processes.* Wesley, The United States of America, 1953.

[11] W. Feller. *An introduction to probability theory and its application.* Volume I ,Wiley, London 1950.

[12] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971.

[13] H. Furstenburg. *Recurrence in Ergodic Theory and combinatorial number theory.* Princeton University Press, Princeton 1981.

[14] P. R. Halmos. *Measure Theory.* Van Nostrand Princeton, New York 1950.

[15] P. R. Halmos. *Lectures on Ergodic Theory.* Chelsea, New York 1956.

[16] M. Kac. On the notion of recurrence in discrete stochastic processes. *Bull. Amer. Math. soc.*, 53:1002–1010, 1947.

[17] S. Karlin and H. M. Taylor. *A first course in stochastic processes.* Second Edition, Academic Press, New York 1975.

[18] P. W. Kasteleyn. Variations on a theme by marc kac. *Journal Stat. Phys.*, 46:811–827, 1987.

[19] M. Keane and W. Th. F. Den Hollander. Ergodic properties of color records. *Journal Phys.*, 138A:183–193, 1986.

[20] R. Nillsen. *Randomness and Recurrence in Dynamical Systems.* Mathematical Association of America, 2011.

[21] Peter V. O'Neil. *Advanced Calculus: pure and applied.* Macmillan, New York, 1975.

[22] Y. Pesin and V. Climenhaga. *Lectures on fractal geometry and dynamical systems.* America Mathematical Society, The United States of America, 2009.

[23] K. Petersen. *Ergodic Theory.* Cambridge Studies in advanced mathematics 2, Cambridge University Press, 1983.

[24] A. G. Postnikov. *Ergodic problems in the theory of congruences and of diophantine approximations.* Proceedings of the Steklov Institute of Mathematics 82, American Mathematical Society, 1967.

[25] A. W. Roberts and D. E. Varberg. *Convex functions.* Academic Press,New York, 1973.

[26] G. F. Simmons. *Introduction to topology and modren analysis.* McGraw-Hill, New York, 1963.

[27] K. R. Stromberg. *An Introduction to classical real analysis.* Wadsworth, Belmont, 1981.

[28] J. Wolfowitz. The moments of recurrence times. *Proc. Amer. Math. Soc.*, 18:613–614, 1967.