PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: Wilkinson

First name: Louise   Other name/s: Anne

Abbreviation for degree as given in the University calendar: MSc

School: Mathematics and Statistics   Faculty: Science

Title: Exchangeability and G-measures

Abstract 350 words maximum: (PLEASE TYPE)

Exchangeability lies at the heart of the Bayesian school of thought in statistics. A sequence is exchangeable if its distribution is unchanged after being permuted. De Finettis' Theorem states that every infinite exchangeable sequence of random variables is mixed independently and identically distributed (i.i.d.). A more refined version of this theorem is that if an infinite sequence is exchangeable then it is conditionally i.i.d. .

We give two proof of de Finetti's Theorem. The first proof is by Glasner (2003). The second proof uses the theory of martingales.

Following on from the work of Riesz (1927) and Keane (1972), Brown and Dooley (1991) introduced the notion of a G-measure on the circle. Throughout this thesis we consider the circle as an infinite product space $Z = \prod_{i=1}^{\infty} \mathbb{Z}_l$ (usually with $l = 2$). Brown and Dooley examined measures on the circle which are quasi-invariant under the action of the group Γ of finite coordinate changes on this infinite product space. In this thesis we replace this by the action of the symmetric group of finite permutations.

We define a G-measure here for the symmetric group as a normalised, compatible family, $\{G_n\}$. We give examples of g-functions, G-functions and G-measures, including an example of a G-measure which is not ergodic.

We reproduce some of the ideas from the construction in Brown and Dooley (1991) for a general group of measurable transformations on a measure space. In particular, we show that finite groups of transformations always provide an averaging operator which is a conditional expectation operator. For the group of finite coordinate changes this bounded linear operator converges uniformly to a constant if and only if there is a unique G-measure (which is therefore ergodic). This is not the case for the bounded linear operator for the symmetric group. However, this may not converge to a constant and there may not be a unique G-measure.

EXCHANGEABILITY AND G-MEASURES

Louise Anne Wilkinson

Supervisors: Associate Professor Ian Doust

Professor Anthony Dooley

School of Mathematics and Statistics,

The University of New South Wales.

March 31, 2016

A thesis submitted for the degree of Master of Science

at The University of New South Wales, March 2016.

ORIGINALITY STATEMENT

‘I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.’

Signed ……………………………………………………...........

Date ……………………………………………………...........

COPYRIGHT STATEMENT

AUTHENTICITY STATEMENT

Contents

CHAPTER 1

Introduction

Ergodic theory is the study of ergodic measures or systems. Walters [1982] gives a broad definition of ergodic theory as the study of the qualitative properties of actions of groups on spaces. The space has some structure which we suppose in this thesis to be that of a measure space (other alternatives in the literature are a topological space or smooth manifold). Ergodic theory originated from the study of statistical mechanics by Boltzmann, Maxwell and others in the 19th century. Boltzmann was trying to understand the gas problem, that is, to understand the state of the particles in a gas after a long time. To solve this problem, Boltzmann introduced the 'ergodic hypothesis', which was the assumption that in the long run, the system would pass through every dynamical state which was consistent with the equations of energy. As Lebowitz and Penrose [1973] note, like good physicists, "they assumed that everything was or could be made all right mathematically and went on with the physics." The word ergodic comes from the Greek word *ergon,* meaning work, and *odos,* meaning path.

As stated, the ergodic hypothesis is not true. Fixing this required the development of measure theory in the early 20th century. The measure theoretic study of ergodic theory began with the ergodic theorems of von Neumann [1932a], von Neumann [1932b] and Birkhoff [1931]. For a discussion of the history of ergodic theory see Walters [1982] and Halmos [1956].

To illustrate some of the concepts in ergodic theory we introduce a couple of standard dynamical systems. Two important transformations in ergodic theory are the one-sided and two-sided Bernoulli shift transformations. A one-sided Bernoulli shift transformation is the map $T : Z \to Z$ where $(x_1, x_2, \ldots) \in Z$ defined by $(T(x)_n) = x_{n+1}$. Note that this map simply shifts each element of Z to the left. A one-sided Bernoulli shift is a measure-preserving transformation, that is the measure of the inverse image is equal to the measure on the set. A two-sided Bernoulli shift is defined similarly to the one-sided shift, the difference is that for the two-sided shift $Z = \mathbb{R}^{\mathbb{Z}}$ instead of $Z = \mathbb{R}^{\mathbb{Z}^+}$.

Ergodicity is the concept of irreducibility of non-singular transformations. A non-singular transformation is a transformation such that the measure of the inverse image is zero if and only if the measure on the set is zero. The two-sided Bernoulli shift transformation is ergodic but the one-sided shift is not. This clearly holds since the one-sided shift is not invertible as it is not one-to-one, therefore it cannot be non-singular. For a proof that the two-sided Bernoulli shift is ergodic see [Walters, 1982, pp. 32-33].

Ergodic theory overlaps several other branches of mathematics (to name a few, these include probability theory, harmonic analysis, number theory, topological groups and Hilbert spaces).

There are two main types of problems in ergodic theory. The first involves understanding measure-preserving transformations and when they are isomorphic. The second gives applications of measure-theoretic ergodic theory. These consider how the theory can be applied to problems in other branches of mathematics and physics.

'Entropy is a measure of randomness or disorder.' [Petersen, 1983, p. 227] In this thesis we will not be looking at the concept of entropy. However, we think that it is an important area of ergodic theory which requires some discussion to complete a general discussion of ergodic theory. Entropy is a categorical concept, therefore variations of this concept exist in other areas of mathematics. The concept of entropy was introduced to ergodic theory by Kolmogorov [1958]. There are two important theorems about entropy in ergodic theory. One of these, introduced by Kolmogorov [1958, 1959] and reformulated by Sinaǐ [1959], is called the Kolmogorov-Sinaǐ Theorem on entropy. This states that the entropy of a transformation can be calculated by finding its entropy with respect to a generator. Consequently, the entropy can sometimes be calculated and it can be concluded that the Bernoulli shifts with different entropies are not isomorphic. For example, [Walters, 1982, p. 102] show that the 2-sided $\left(\frac{1}{2}, \frac{1}{2}\right)$-shift has entropy $\log 2$ and the 2-sided$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$-shift has entropy $\log 3$, therefore these transformations are not isomorphic.

Another important theorem by Ornstein [1974] states that two Bernoulli shifts are isomorphic if and only if they have the same entropy. Thus, for example, any two measure-preserving transformations of a finite space have zero entropy and are therefore isomorphic (see Walters [1982]). Note that it is not always possible to calculate the mean entropy of a transformation, for example the limit used to calculate the mean entropy of a non-singular system which is not measure-preserving does not, in general, exist.

Michael Keane [1972]), studied Gibbs measures in statistical mechanics and introduced the notion of a g-function, where g is a $C^1\,[0,1]$ function which is the normed space of continuously differentiable functions on $[0,1]$ (see [Kreyszig, 1989, p. 110]. G-measure formalism was first introduced by Brown and Dooley [1991], where the notion of G-measures was used to generalise certain Riesz product measures. The discussion here is based on Dooley [2007].

Every quasi-invariant probability measure is a G-measure for the group of finite coordinate changes. We begin our discussion of G-measures with a simple example of a measure, a Markov measure. A Markov measure is a special case of a G-measure. In standard statistical theory a stochastic process is called a Markov process if the conditional distribution of any future state on the past and current states depends only on the conditional distribution of the current state.

Another important example of a G-measure is a Riesz product measure. This is not a product measure nor a Markov measure.

Riesz [1927] introduced a construction of a measure on the circle (that is it is on $\mathbb{T}$ or equivalently on the infinite product space $\mathbb{Z}_l^{\mathbb{Z}^+}$ where l is an integer such that $l \geq 2$), which has since become known as a Riesz product. This measure is originally from classical Harmonic Analysis. Riesz products in their modern form were introduced by Sidon in his study of lacunary trigonometric series in his theorems of 1927 and 1937 (for a more detailed discussion of this see Kahane [2006]). These are in general singular with respect to Lebesgue measure and are weak-$*$ limits of measures of the form

$$d\mu = \lim_{n\to\infty} \prod_{i=1}^{n} \left(1 + a_i \cos 2\pi \left(3^i x + \phi_i\right)\right) \, dx.$$

Note that $x \in [0, 1)$, $\{a_i\}$ is a sequence of real numbers with $-1 \leq a_i \leq 1$ and $\{\phi_i\}$ is a sequence of phases in $[0, 1)$. Riesz showed that for any choice of x, $\{a_i\}$, and $\{\phi_i\}$ there exists a unique weak-$*$ limit measure μ, which is often singular with respect to Haar measure. Note that Haar [1933] introduced the notion of a Haar measure. A Haar measure is as a translation invariant measure on any locally compact topological group. In this thesis, we let $\{\mu_n\}_{n\in\mathbb{Z}^+}$ be a sequence of probability measures on $(X, \mathcal{B}, \mu)$. Recall that the sequence μ_n

converges weakly to a measure μ, written $\mu_n \xrightarrow{w} \mu$ if

$$\lim_{n\to\infty} \int_X f \, d\mu_n \to \int_X f \, d\mu,$$

for all $\mathcal{C}(X)$, the space of all bounded continuous functions on X.

Work on Riesz products by Brown and Moran [1974] showed that they were absolutely continuous with respect to Lebesgue measure if and only if $\sum a_i^2 < \infty$, and otherwise they were singular.

Brown [1978] showed that if we consider the action of the triadic rationals on the circle, that is $\gamma : x \mapsto x + \frac{p}{3^n}$ for values of n and p (which are relatively prime to 3), then $\mu \circ \gamma \sim \mu$ and the measure μ is ergodic for the action of the group of triadic rationals.

G-measure formalism originated from generalising the description of the Riesz product measure. Here, the functions g_i are a given family of functions on the circle. We can think of the terms in the Riesz product as

$$g_i\left(3^i x\right) = \left(1 + a_i \cos\left(2\pi \cdot 3^i x\right)\right)$$

for $i \geq 1$, $x \in [0, 1)$ and

$$G_n(x) = \prod_{i=1}^{n} g_i\left(3^i x\right).$$

The functions G_n are related to the Radon-Nikodým derivative of the measure with respect to its translation by triadic rationals. We can see from this definition that Brown and Dooley [1991] generalise Keane [1972] and Riesz [1927]. Brown and Dooley [1991] generalised both Keane's construction and that of the classical Riesz products in the notion of G-measures. The term g-function is used for the case when all the g_n's are the same function.

This thesis extends the method and formalism of G-measures to the setting of the infinite permutation group. A measure associated to a normalised compatible family, $\{G_F\}$, is called a G-measure. In this thesis we will consider G-measures for the symmetric group. A family of functions G_F is compatible if for finite subsets $F_1 \subseteq F_2 \in \mathbb{Z}^+$

$$G_{F_1}(x)\, G_{F_2}(\sigma x) = G_{F_1}(\sigma x)\, G_{F_2}(x),$$

for all σ in the symmetric group, Σ_{F_1}. It is normalised if for any finite set $F \subseteq \mathbb{Z}^+$

$$\frac{1}{|\Sigma_F|} \sum_{\sigma \in \Sigma_F} G_F(\sigma x) = 1.$$

We will give conditions under which G-measures exist.

The general contents of each chapter are described below.

We begin Chapter two by outlining some simple concepts in ergodic and measure theory. These include definitions of different types of maps, ergodicity and invariant and quasi-invariant measures. In the section on measure theory we discuss infinite product measure spaces and whether two measures are mutually singular, absolutely continuous or equivalent. We then discuss the important concepts of the Radon-Nikodým Theorem and the Radon-Nikodým derivative. We also outline Kakutani's Theorem which gives conditions for when two measures are mutually singular or absolutely continuous.

The main topic in Chapter two is exchangeability and de Finetti's Theorem. A sequence is exchangeable if its distribution is unchanged after being permuted. That is, an infinite sequence $x = (x_1, x_2, \ldots)$, is exchangeable if

$$(x_1, x_2, \ldots) \stackrel{d}{=} \left(x_{\sigma(1)}, x_{\sigma(2)}, \ldots\right) \tag{1.0.1}$$

for σ an element of the symmetric group, where $\stackrel{d}{=}$ denotes equality in distribution. Note that if $x_1, x_2, \ldots$ are independently and identically distributed i.i.d. , then they are exchangeable, but the converse is not true. A sequence or other collection of random variables is i.i.d. if each random variable has the same probability distribution as the others and all are mutually independent.

De Finetti's Theorem says that every infinite, exchangeable sequence of random variables is mixed i.i.d. , that is, the distribution can be written as a weighted average of i.i.d. sequences. A more refined version of this theorem is that if an infinite sequence is exchangeable then it is conditionally i.i.d. . Using the results that have been developed in the literature since the first proof of de Finetti's Theorem allows this more succinct version of this theorem. The conditions above are for an infinite sequence. It is possible to derive a similar condition for finite exchangeable distributions.

We then discuss Aldous and Pitman's Theorem of exchangeability. Let $(X, \mathcal{B}, \mu)$ be a probability space, then the exchangeable σ-algebra is the collection of events $B \in \mathcal{B}$ where B is a measurable subset of sequence space which is invariant under the permutations of finitely many coordinates. This theorem states a necessary condition for the exchangeable σ-algebra to be ergodic. We then give examples to illustrate this theorem.

We define permutations and the symmetric group since this is the main group of interest in this thesis.

We give two different proofs of de Finetti's Theorem. The first proof is by Glasner [2003]. The other proof uses the notion of G-measures which we discuss in more detail in Chapter four.

Chapter two draws on the literature and we begin to introduce our own work in Chapter three. The major topic we discuss in Chapter three, is the concept of G-measures. Firstly, as already discussed above both G-measures and g-functions are weak-$*$ limits of measures which are often singular (with respect to Haar measure). Secondly, the functions, G_n, are products of the g_i functions such that $G_n = g_1 \cdots g_n$. For the special case of G-measures all the g_i are the same function (equal to g). Original content in this thesis includes the discussion in Chapter three of G-measures for the symmetric group.

The construction of G-measures from a G-family of functions uses a certain family of averaging operators. Note that for the group of finite coordinate changes this bounded linear operator converges uniformly to a constant if and only if there is a unique G-measure (which is therefore ergodic). This is not the case for the bounded linear operator for the symmetric group. This may not converge to constant and there may not be a unique G-measure. This part of the thesis is original work. In Chapter four we look more closely at these operators and show that they are conditional expectation operators with respect to suitable measures. This leads to a consideration of martingales.

We give another proof in Chapter four of de Finetti's Theorem using the theory of martingales. This version of de Finetti's Theorem states that if an infinite sequence is exchangeable then it is conditionally i.i.d. . We show how normalised compatible families are naturally associated to martingales. This link between the proof of de Finetti's Theorem using the theory of martingales and G-measures is our own work.

Note that the Radon-Nikodým derivative of any probability measure which is quasi-invariant for the finite coordinate changes form a normalised compatible family. We adapt this approach to the infinite permutation group acting on an infinite product space.

CHAPTER 2

Ergodic Theory and de Finetti's Theorem

Ergodic theory is the study of ergodic measures or ergodic systems. Note that it is not just the study of dynamics on a measure space since non-ergodic systems are included in dynamical systems. In this chapter we begin by outlining some basic concepts in ergodic and measure theory. We then go on to discuss ergodicity and invariant and quasi-invariant measures. This is followed by the Hewitt-Savage zero-one law which gives conditions under which a measure is ergodic (for a transformation). We also define permutations and the symmetric group for both the finite and infinite case.

In the section on measure theory we begin by discussing infinite product measure spaces. We then go on to discuss the important concept of the Radon-Nikodým Theorem and the Radon-Nikodým derivative. This is followed by a discussion of Kakutani's Dichotomy Theorem which gives the criteria under which two measures are mutually singular or absolutely continuous.

The starting point for the theory on exchangeability is the following example. Suppose that we have n balls labelled $1, \dots n$ in a bag which are drawn out randomly from the bag, *without replacement*. Let x_i denote the random variable giving the label of the ith ball. The sequence $x = (x_i)$ just gives the ordered list of labels of the balls, and since the balls are equally likely to have been pulled out in any order, the distribution here is just the uniform distribution supported on the permutations of $[n] = \{1, \dots n\}$. The sequence of random variables x_i, although identically distributed, is clearly not independent.

Note in particular, that if $\sigma : [n] \to [n]$ is any permutation, and $x_\sigma = \left(x_{\sigma(i)}\right)$ is the corresponding permutation of the sequence of random variables, then the distribution of x_σ is identical to that of x. In cases such as this we say that x is an **exchangeable** sequence of random variables.

In general, if $x = (x_i)$ is an i.i.d. sequence of random variables, then x is clearly exchangeable. As the above example shows, independence is not necessary for exchangeability.

For a second example, suppose that we have n bags, and that the ith bag contains one blue ball and i red balls. Choose one ball from each bag, and let y_i be the random variable which is one if the ith ball is blue and zero otherwise. The sequence $y = (y_i)$ consists of random variables which are independent, but which are not identically distributed. They are certainly not exchangeable. Certain events however, do not depend on the ordering of the bags. Consider for example

$$E_1 = \{(a_i) \in \{0,1\}^n : a_1 = 1\}$$

$$E_2 = \left\{ (a_i) \in \{0,1\}^n : \sum_{i=1}^{n} a_i = 3 \right\}.$$

Suppose $\sigma : [n] \to [n]$ is a permutation. Unless $\sigma(1) = 1$, it is clear that $\operatorname{Prob}(y \in E_1) \neq \operatorname{Prob}(y_\sigma \in E_1)$. However, $\operatorname{Prob}(y \in E_2) = \operatorname{Prob}(y_\sigma \in E_2)$. Events such as E_2 whose probabilities are invariant under permutations of the sequence are called **exchangeable** events.

De Finetti's theorem characterises infinite exchangeable sequences of random variables. It says that an infinite sequence of random variables is exchangeable if and only if it is a "mixture" of i.i.d. sequences of random variables.

If $x = (x_i)$ is a sequence of i.i.d. (and hence exchangeable) real-valued random variables each with distribution m and A is a suitable subset of $\mathbb{R}^\infty$ then

$$\operatorname{Prob}(x \in A) = m^\infty (A),$$

is given by the infinite product measure determined by m. This will be defined more formally in Section 2.2.

We say that a sequence $x = (x_i)$ of random variables is a mixture of i.i.d. random variables if there exists a set T of parameters equipped with a measure ν such that for each $t \in T$, m_t is a probability measure on $\mathbb{R}$ and so that

$$\operatorname{Prob}(x \in A) = \int_T m_t^\infty (A) \, d\nu (t),$$

for suitable $A \subseteq \mathbb{R}^\infty$. That is, $x = (x_1, x_2, \ldots)$ is a type of average of i.i.d. sequences with distributions m_t, $t \in T$. This will be made more precise in Sections 2.6 and 2.7.

As shown by the example above, de Finetti's Theorem may fail for finite exchangeable sequences, $x = (x_1, \ldots, x_n)$. In Section 2.8 we will state conditions for finite sequences to be exchangeable.

We will give two proofs of de Finetti's Theorem. Let $\mathfrak{P}(Z)$ be the set of probability measures on Z. The first proof, shows that $m_t^\infty \in \mathfrak{P}(Z)$ if and only if m_t^∞ is Σ_∞-ergodic. The second proof uses martingales and is given in Chapter four.

2.1. Introduction to Ergodic Theory

This section introduces some basic concepts in ergodic theory. We summarise some of the background material which will be used in this thesis. Most of this can be found in Aaronson [1997] (or even wikipedia).

Ergodic theory is the study of ergodic measures or systems. We define here what it means for a transformation to be measurable, non-singular and measure-preserving.

Throughout this section we shall let $(X, \mathcal{B}, \mu)$ denote a measure space. That is X is a nonempty set, $\mathcal{B}$ is a σ-algebra of subsets of X and μ is a non-negative measure on $\mathcal{B}$. We shall say that $(X, \mathcal{B}, \mu)$ is a probability space, and that μ is a probability measure if $\mu(X) = 1$.

Definition 2.1.1. Royden [1969, p.238]. Let $(X, \mathcal{B})$ be a fixed measurable space and μ and ν be two measures on $(X, \mathcal{B})$.

(i) We say that μ and ν are **mutually singular** if there are disjoint sets A and B in $\mathcal{B}$ such that $X = A \cup B$ and $\mu(A) = \nu(B) = 0$. We write this as $\mu \perp \nu$.

(ii) We say that μ is **absolutely continuous** with respect to ν if $\nu(E) = 0$ implies $\mu(E) = 0$ for all $E \in \mathcal{B}$. This is written as $\mu \ll \nu$.

(iii) If both $\mu \ll \nu$ and $\mu \gg \nu$ then μ and ν are called **equivalent**, this is written as $\mu \sim \nu$. It is easy to see that $\sim$ is an equivalence relation.

Definition 2.1.2. We let $\mathfrak{P}(\mathrm{X})$ be the set of probability measures on X and ν is a probability measure on the Borel sets of X.

Definition 2.1.3. Suppose that P is a property that elements of X may satisfy. This may be encoded by a True/False valued function π defined on X, where of course $\pi(x)$ is true if x satisfies P. We shall say that P holds **almost everywhere (a.e.)** or $\pi(x)$ is true almost everywhere if

$$\mu(\{x \in X : \pi(x) \text{ is false}\}) = 0.$$

Here we give definitions used to classify transformations.

Definition 2.1.4. [Walters, 1982] Let $(X_1, \mathcal{B}_1, \mu_1)$ and $(X_2, \mathcal{B}_2, \mu_2)$ be probability spaces. A transformation $T : X_1 \to X_2$ is

(1) **Measurable** if $T^{-1}B_2 \in \mathcal{B}_1$ for all $B_2 \in \mathcal{B}_2$.
(2) **Measure-preserving** if T is a measurable map and $\mu_1(T^{-1}(B_2)) = \mu_2(B_2)$ for all $B_2 \in \mathcal{B}_2$.
(3) An **invertible measure-preserving** transformation if T is a measure-preserving, bijective, and T^{-1} is also measure-preserving.
(4) **Non-singular** if it is a measurable transformation, such that $\mu_1(T^{-1}(B_2)) = 0$ if and only if $\mu_2(B_2) = 0$ for all $B_2 \in \mathcal{B}_2$.

If $T : X \to X$ is a non-singular transformation, then we shall call the quadruple $(X, \mathcal{B}, \mu, T)$ an **abstract dynamical system**.

Definition 2.1.5. Suppose that $(X, \mathcal{B}, \mu, T)$ is an abstract dynamical system.

(1) We say that μ is **invariant** under T if for every measurable set $B \in \mathcal{B}$, $\mu(T^{-1}(B)) = \mu(B)$. That is, μ is an invariant measure under T, if T is measure-preserving, or equivalently, that $\mu \circ T^{-1} = \mu$.

(2) We say that μ is a **quasi-invariant** under T if μ is equivalent to the measure $\mu \circ T^{-1}$. This is equivalent to requiring that $\mu(B) = 0$ if and only if $\mu\left(T^{-1}(B)\right) = 0$ for $B \in \mathcal{B}$.

We have now given enough definitions to explain the concept of ergodicity. We can think of ergodicity as the concept of irreducibility of non-singular transformations. The following definition is from Aaronson [1997, p. 51]. This uses the set symmetric difference, denoted by $\triangle$.

DEFINITION 2.1.6. Suppose that $(X, \mathcal{B}, \mu)$ is a probability space, and that $\mathcal{G}$ is a locally compact, second countable topological group. A non-singular action of $\mathcal{G}$ on $(X, \mathcal{B}, \mu)$ is a map

$$\mathcal{G} \times X \quad \rightarrow \quad X$$

defined by

$$(\mathrm{g}, x) \mapsto T_{\mathrm{g}} x,$$

where each $T_{\mathrm{g}} : X \rightarrow X$ is a non-singular transformation and

$$T_{\mathrm{g}} \circ T_{\mathrm{h}} = T_{\mathrm{gh}}$$

for all $\mathrm{g}, \mathrm{h} \in \mathcal{G}$. We say that this action is **ergodic** if $A \in \mathcal{B}$, $\mu(A \triangle T_{\mathrm{g}} A) = 0$ for all $\mathrm{g} \in \mathcal{G}$, implies that $\mu(A) = 0$ or $\mu(A^c) = 0$.

Note that $\mu(A \triangle T_{\mathrm{g}} A) = 0$ if and only if $\mu\left(A \triangle T_{\mathrm{g}}^{-1} A\right) = 0$. The set S is a.e. invariant if $\mu(S \triangle T_{\mathrm{g}} S) = 0$ for all $\mathrm{g} \in \mathcal{G}$. It is ergodic if every invariant set has measure zero or its complement has measure zero.

Definition 2.1.7. Let $(X, \mathcal{B}, \mu, T)$ be an abstract dynamical system. We shall say that μ is an **ergodic measure** for T if μ is a T-invariant probability measure and the only T-invariant sets have measure 0 or 1. That is, for $A \in \mathcal{B}$, $T(A) = A$ implies $\mu(A) = 0$ or $\mu(A) = 1$.

2.2. Measure Theory and Ergodicity

This section outlines some definitions from measure theory. We begin by discussing **infinite product measure spaces**, see Bogachev [2007, pp. 187-188]. This is followed by the Radon-Nikodým theorem and the concept of the Radon-Nikodým derivative, drawing heavily on Halmos [1950] and Capiński and Kopp [2004].

Definition 2.2.1. Let $(X_i, \mathcal{B}_i, \mu_i)$ denote a family of probability measure spaces indexed by elements of $i \in \mathbb{Z}^+$. Let X denote the Cartesian product of these spaces

$$X = \prod_{i \in \mathbb{Z}^+} X_i = \left\{ (x_i)_{i=1}^{\infty} : x_i \in X_i \right\}.$$

A (measurable) cylinder set in X is a set of the form

$$A = A_1 \times A_2 \times \cdots \times A_n \times X_{n+1} \times X_{n+2} \times \cdots$$

where $n \in \mathbb{Z}^+$ and $A_i \in \mathcal{B}_i$ for $i = 1, 2, \ldots, n$.

One can define a map μ on a cylinder set by

$$\mu(A) = \prod_{i=1}^{n} \mu_i(A_i).$$

The cylinder sets do not form a σ-algebra. If one denotes by $\mathcal{B}$ the smallest σ-algebra containing the cylinder sets, then it can be shown in [Walters, 1982,

p. 5] that μ can be extended to a measure on $(X, \mathcal{B})$. In particular, if the spaces $(X_i, \mathcal{B}_i, \mu_i)$ are all probability spaces, then so is $(X, \mathcal{B}, \mu)$. We shall write $\mathcal{B} = \otimes_{i=1}^{\infty} B_i$ and $\mu = \otimes_{i=1}^{\infty} \mu_i$.

The same construction applies (with fewer technical difficulties) for finite Cartesian products, $X = \prod_{i=1}^{n} X_i$, $\mathcal{B} = \otimes_{i=1}^{n} B_i$ and $\mu = \otimes_{i=1}^{n} \mu_i$.

The most standard example of this (which we shall return to often) is as follows.

Example 2.2.2. Let $X_i = \{0, 1, \ldots, l-1\}$ where l is an integer such that $l \geq 2$ and μ_i is a normalised counting measure for each $i = 1, 2, \ldots$. Then $X = \prod_{i \in \mathbb{Z}^+} X_i$ can be identified (up to sets of measure zero) with $[0, 1)$ via the map $(x_1, x_2, \ldots) \mapsto \sum_{i=1}^{\infty} x_i l^{-i}$. In this case $\mu = \otimes_{i=1}^{\infty} \mu_i$ is just Lebesgue measure on $\mathcal{B} = \otimes_{i=1}^{\infty} B_i$ which comprises the Borel subsets of $[0, 1]$.

We shall say that two sequences $x = (x_i)$ and $y = (y_i)$ are eventually equal if there exists n such that $x_i = y_i$ for all $i \geq n$.

Theorem 2.2.3. *(Hewitt-Savage 0-1 Law) Let $(X, \mathcal{B}, \mu)$ be the infinite product space generated by the sequence of spaces $(X_i, \mathcal{B}_i, \mu_i)_{i=1}^{\infty}$. Suppose that $Y \in \mathcal{B}$ has the property that $x = (x_i)$ is in Y if and only if every sequence $y = (y_i)$ which is eventually equal to x is also in Y. Then $\mu(Y)$ is either* 0 *or* 1.

Suppose that $\sigma : \mathbb{Z}^+ \to \mathbb{Z}^+$ is a finite permutation of $\mathbb{Z}^+$, that is, σ fixes all but finitely many of the elements of $\mathbb{Z}^+$. Let $X = \prod_{i=1}^{\infty} X_i$ where all the X_i are the same. Define $T_\sigma : X \to X$ by $T_\sigma((x_i)) = (x_\sigma(i))$. If $A \in \mathcal{B}$ is invariant under all T_σ then the Hewitt-Savage 0-1 Law implies that $\mu(A)$ is either 0 or 1. This means that μ is ergodic for $(T_\sigma)_{\sigma \in \Sigma_\infty}$. The set of such finite permutations forms a group Σ_∞ which we will formally introduce in Section 2.4. The measure μ is therefore ergodic for the action of the group Σ_∞.

Let $(X, \mathcal{B}, \mu, T)$ be a probability space. For every non-negative integrable real function $f : X \to \mathbb{R}$, it is straightforward to show that the set function

$$E \mapsto \nu(E) = \int_E f \, d\mu$$

defines a measure ν on $(X, \mathcal{B})$, for example, see Capiński and Kopp [2004, §7.2]. In this section we discuss how the Radon-Nikodým Theorem answers the question of which measures ν can be constructed using this method.

Definition 2.2.4. We say that a measure space $(X, \mathcal{B}, \mu)$ is **σ-finite** if there exist sets $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{B}$ with $X = \bigcup_{i=1}^{\infty} A_i$ and $\mu(A_i) < \infty$ for each i.

Theorem 2.2.5. *Royden [1969, pp.238-40].* ***Radon-Nikodým Theorem.*** *Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space. Let ν be a measure on $\mathcal{B}$ which is absolutely continuous with respect to μ, that is $\nu \ll \mu$. Then there exists a non-negative measurable function f such that for each set $E \in \mathcal{B}$ we have*

$$\nu(E) = \int_E f \, d\mu.$$

The function f is unique up to sets of μ-measure zero. The function f is called the ***Radon-Nikodým derivative*** *of ν with respect to μ and can be denoted by $\frac{d\nu}{d\mu}$.*

THEOREM 2.2.6. *[Glasner, 2003, p. 80]. A dynamical measure-preserving system, $(X, \mathcal{B}, \mu, T)$, is ergodic if and only if for every f, $g \in L^2(\mu)$ (or $L^1(\mu)$),*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \int f \circ T^i \cdot \bar{g}\, d\mu = \int f\, d\mu \int \bar{g}\, d\mu.$$

2.3. Kakutani's Dichotomy Theorem

Kakutani's Dichotomy Theorem gives criteria for when two infinite product measures are mutually singular or absolutely continuous. Note that Brown and Dooley [1994] extend Kakutani's Dichotomy Theorem to G-measures for the group of finite coordinate changes. The following section is based on Hewitt and Stromberg [1965].

THEOREM 2.3.1. ***Kakutani's Dichotomy Theorem****. Let $X = \prod_{i=1}^{\infty} X_i$ be an infinite product space and for $i = 1, 2, 3, \ldots$, let μ_i and η_i be probability measures on X_i. Let $\mu = \otimes_{i=1}^{\infty} \mu_i$ and let $\eta = \otimes_{i=1}^{\infty} \eta_i$. Then either*

(i) $\eta \ll \mu$ or

(ii) $\eta \perp \mu$.

Let f_i be the Radon-Nikodým derivative $\frac{d\eta_i}{d\mu_i}$ as defined in Theorem 2.2.5. Then (i) holds if and only if

(iii) $\prod_{i=1}^{\infty} \left(\int_{X_i} f_i^{\frac{1}{2}}\, d\mu_i \right) > 0$,

and (ii) holds if and only if

(iv) $\prod_{i=1}^{\infty} \left(\int_{X_i} f_i^{\frac{1}{2}}\, d\mu_i \right) = 0$.

EXAMPLE 2.3.2. Hewitt and Stromberg [1965]. Let $X_i = \{0, 1\}$ for all $i \in \mathbb{Z}^+$, and let α be a sequence $\{\alpha_i\}_{i=1}^{\infty}$ such that $0 < \alpha_i < 1$. Let μ_α be the measure on $(X, \mathcal{B})$, that is the product of measures μ_i on $\{0, 1\}$ such that $\mu_i(\{0\}) = \alpha_i$, $\mu_i(\{1\}) = 1 - \alpha_i$. Suppose that α and β are any two such sequences. Then exactly one of the following assertions holds:

(i) $\mu_\alpha \ll \mu_\beta$ and $\mu_\beta \ll \mu_\alpha$, or

(ii) $\mu_\alpha \perp \mu_\beta$.

This follows directly from Kakutani's Theorem, Theorem 2.3.1. Property (i) holds if and only if

(iii) $\sum_{i=1}^{\infty} \left(1 - \alpha_i^{\frac{1}{2}} \beta_i^{\frac{1}{2}} - (1-\alpha_i)^{\frac{1}{2}} (1-\beta_i)^{\frac{1}{2}}\right) < \infty$,

and property (ii) holds if and only if this series in (iii) diverges.

2.4. Symmetric Groups

In this section we give definitions of the finite and infinite symmetric groups. The following definition can be found in Dooley and Fan [1997, pp. 113-114], Dummit and Foote [2004, pp. 28, 82] and Olshanski [2008].

Definition 2.4.1. Let A be any non-empty set. The set of the **permutations** of A, is a bijection from A to A, denoted by $\sigma : A \to A$. We denote the set of all such maps by Σ_A.

The set Σ_A is a group under function composition: $\circ$. This group is called the **symmetric group** on the set A. In the special case of permutations of the index set $[n] = \{1, 2, \ldots, n\}$ the symmetric group on A is denoted Σ_n, the symmetric group of degree n. The cardinality of this group is $|\Sigma_n| = n!$.

We can generalise this as follows. Suppose that Z is a set with symbols $\{x_1, \ldots x_n\}$, $\sigma \in \Sigma_n$ acts on Z via

$$\sigma : Z \to Z$$

such that

$$\sigma(x_i) = x_{\sigma(i)}.$$

If $\sigma \in \Sigma_n$, we let σ act on the positive integers by permuting $\{1, 2, \ldots, n\}$ and fixing integers greater than n. In this way we can consider Σ_n to be a subgroup of the infinite group $\Sigma_{\mathbb{Z}^+}$. We define the infinite symmetric group to be

$$\Sigma_\infty = \bigcup_{n=1}^{\infty} \Sigma_n.$$

Note that this is sometimes called the finitary symmetric group.

Obviously, Σ_∞ is a countable, locally finite group. Note that the symmetric group, Σ_n, is not commutative for $n > 2$.

We can compare the symmetric group to the group of finite coordinate changes.

Definition 2.4.2. (The **group of finite coordinate changes**, see Brown and Dooley [1991]) Let $l(i) \geq 2$ be a sequence of integers. Let $\Gamma = \oplus_{i=1}^{\infty} \mathbb{Z}_{l(i)}$ be the direct sum of the groups $\mathbb{Z}_{l(i)}$, consisting of all sequences $\gamma = (\gamma_i)$ with $\gamma_i \in \mathbb{Z}_{l(i)}$ and $\gamma_i = 0$ except for finitely many coordinates. The set Γ forms a group under elementwise addition mod $l(i)$ and this group acts on $Z = \prod_{i=1}^{\infty} \mathbb{Z}_{l(i)}$ by

$$\gamma x = (x_1 + \gamma_1, x_2 + \gamma_2, \ldots),$$

where each addition is done mod $l(i)$. The group Γ is called the group of finite coordinate changes. For each $n \in \mathbb{Z}^+$ and for all $i > n$, let

$$\Gamma_n = \{\gamma \in \Gamma : \gamma_i = 0, \forall\, i \notin n\}.$$

Clearly, $\Gamma = \bigcup_{n=1}^{\infty} \Gamma_n$.

2.5. Exchangeable Random Variables and Events

Aldous [1985] provides a survey of articles on exchangeability from a statistical view. The study of exchangeability began with the publications of de Finetti [1972b, 2011]. Lauritzen [2007] provides a summary of exchangeability and de Finetti's Theorem. This section draws heavily on [Olav, 2005].

Suppose that $(X_i, \mathcal{B}_i, \mu_i)$, $i \in \mathbb{Z}^+$ is a sequence of probability spaces. Let $X = \prod_{i=1}^{\infty} X_i$ be the infinite product of these spaces equipped with the product measure $\mu = \otimes_{i=1}^{\infty} \mu_i$ on $\mathcal{B}$, the smallest σ-algebra containing all the cylinder sets.

Definition 2.5.1. Suppose that $x = (x_i)$ is a sequence of random variables, with each x_i taking values in $(X_i, \mathcal{B}_i, \mu_i)$ where all X_i are equal. We say that x is exchangeable if for every finite permutation $\sigma \in \Sigma_\infty$, the distribution of $\sigma(x)_i = \left(x_{\sigma(i)}\right)$ is the same as that of x. That is, given an infinite sequence of random elements $x = (x_1, x_2, \ldots)$, we say that x is **exchangeable** if

$$(x_{k_1}, x_{k_2}, \ldots, x_{k_m}) \stackrel{d}{=} (x_1, x_2, \ldots, x_m) \tag{2.5.1}$$

for any distinct elements $k_1, k_2, \ldots, k_m$ of the index set, where $\stackrel{d}{=}$ denotes equality in distribution. For this to be the case, it is clearly enough to require that Equation 2.5.1 be satisfied for any finite permutation. Note that x is exchangeable if its distribution is invariant under finite permutations.

Example 2.5.2. If the random variables $x_i \in X_i$ are i.i.d. then they are certainly exchangeable.

Definition 2.5.3. Let $X = \prod_{i=1}^{\infty} Y$ for a fixed Y. The σ-algebra of symmetric events $\mathcal{B}$ is defined by Hewitt and Savage [1955, p. 474] to be the set of events $A \subseteq X$ which are invariant under finite permutations of the indices in the sequence (x_i). That is, if $x = (x_i)_{i=1}^{\infty} \in A$ then $\sigma(x) \in A$ for all $\sigma \in \Sigma_\infty$. This is referred to as the **exchangeable σ-algebra**.

[Hoffmann-Jorgensen et al., 2012] show that $\mu \in \operatorname{Prob}(X, \mathcal{B})$ is an **exchangeable probability measure** on $(X, \mathcal{B})$ if

$$\mu\left(\prod_{i=1}^{\infty} B_i\right) = \mu\left(B_{\sigma(1)} \otimes \ldots \otimes B_{\sigma(n)} \otimes B_{n+1} \otimes \ldots\right)$$

for all $\sigma \in \Sigma_n$ and $B_i \in \mathcal{B}$ with $i \geq 1$ and $n \geq 1$.

Example 2.5.4. In addition to the example of symmetric and non-symmetric events discussed in the introduction to this chapter, we give a simple example here of an exchangeable sequence. We can consider the example of Polya's Urn as discussed in Lauritzen [2007]. Consider an urn with b black balls and w white balls. Choose a ball at random. Add a balls of the same colour to the urn together with the withdrawn ball. Repeat the process. Let $x_i = 1$ if the ith ball is black and $x_i = 0$ otherwise. Let $X_i = \{0, 1\}$ and $\mu = \otimes_{i=1}^{n} \mu_i$. Here the x_i are not independent and they are not a Markov process. However,

$$\begin{aligned}
\operatorname{Prob}(1,1,0,1) &= \frac{b}{b+w} \cdot \frac{b+a}{b+w+a} \cdot \frac{w}{b+w+2a} \cdot \frac{b+2a}{b+w+3a} \\
&= \frac{w}{b+w} \cdot \frac{b}{b+w+a} \cdot \frac{b+a}{b+w+2a} \cdot \frac{b+2a}{b+w+3a} \\
&= \operatorname{Prob}(0,1,1,1).
\end{aligned}$$

It is relatively straightforward to generalise this special case to show that $x = (x_i)$ is exchangeable.

Let (x_i) be a sequence of random variables taking values in Y and let μ_i be the probability measures defined on $(X, \mathcal{B})$ by $\mu_i(B) = \operatorname{Prob}(x_i \in B)$ for $B \in \mathcal{B}_i$. The following theorem gives a condition for $\mathcal{B}$ to be an exchangeable σ-algebra.

THEOREM 2.5.5. *[Aldous and Pitman, 1979, Theorem 1.6] A necessary condition for a σ-algebra $\mathcal{B}$ to be exchangeable is that for all $A \in \mathcal{B}$,*

$$\sum_{i=1}^{\infty} \min\left(\mu_i\left(A\right), \mu_i\left(A^c\right)\right),$$

is either 0 *or* ∞.

2.6. Examples of Exchangeabiliy

We give an example here of an application of de Finetti's Theorem as a mixture of i.i.d. sequences.

EXAMPLE 2.6.1. Suppose $t \in [0,1]$. Let x_t denote the sequence of outcomes from the toss of a coin C_t which satisfies $\mathrm{Prob}\left(C_t = 1\right) = t$. Then x_t is an example of an i.i.d. sequence. As we have already discussed, a sequence of random variables is an i.i.d. sequence if each random variable has the same probability distribution as the others and all are mutually independent. This is clearly the case here. As discussed in the introduction to Chapter two, while independence is not necessary for exchangeability, if x_t is an i.i.d. sequence it is clearly exchangeable. This explains the exchangeability of the outcomes from this coin toss.

Let ν be a probability measure on $[0,1]$. We can now define a random sequence $x = (x_1, x_2, \ldots)$ by choosing t at random (according to ν) and then letting x_i be the result of the ith toss of the coin C_t.

Let m_t denote the (Bernoulli) probability distribution of such a coin toss, so $m_t\left(\{1\}\right) = t$ and $m_t\left(\{0\}\right) = 1 - t$. Then for measurable $A \subseteq X = \prod_{i=1}^{\infty} \{0,1\}$

$$\mathrm{Prob}\left(x \in A\right) = \int_0^1 m_t^{\infty}\left(A\right)\, d\nu\left(t\right),$$

where $m^\infty = m \times m \times \cdots$ denotes the product distribution.

Therefore, x is exchangeable since it is a mixture of i.i.d. sequences.

In fact, any average of i.i.d. sequences will have the same property. De Finetti's Theorem states that all exchangeable sequences are of this form. Note that this does not depend on the measure ν.

2.7. De Finetti's Theorem

We consider several different versions of de Finetti's Theorem in this section.

THEOREM 2.7.1. ***De Finetti's Theorem***. *Every infinite, exchangeable sequence of random variables $x = (x_1, x_2, \ldots)$ is mixed independently identically distributed (i.i.d.).*

REMARK 2.7.2. While everyone agrees with de Finetti's Theorem as stated in Theorem 2.7.1. There are several alternatives to the theorem in the literature. This is because the concept of "a mixture of i.i.d. sequences" can be defined differently. The definitions and lemmas stated here are from Aldous [1985].

In Bayesian statistics, de Finetti's Theorem derives a general infinite exchangeable sequence (x_i) by choosing a distribution m at random from some prior. Equivalently, we let the (x_i) be i.i.d. variables with distribution m. De Finetti's Theorem says that we can associate to (x_i) a random distribution $\mu(\omega, \cdot)$ conditional on $\mu = m$.

We can define a sequence (x_i) by

(i) Select m at random from a probability distribution on $\mathbb{R}$ given by $\{m_1, \ldots, m_k\}$ such that

$$\text{Prob}\,(m = m_i) = p_i,$$

where $p_1, \ldots, p_k > 0$ and $\sum_{i=1}^k p_i = 1$;

(ii) Let (x_i) be i.i.d. with distribution m.

As an alternative to (i) we have

(i') Select m at random from the distribution ν.

Here we give the idea from Bayesian statistics that (x_i) are i.i.d. variables with distribution m, where m has the prior distribution ν. Similar to our discussion in the introduction to this chapter, we can write this as

$$\text{Prob}\,(x \in A) = \int_{m \in \mathfrak{P}(Z)} m^{\infty}(A)\, \nu\,(dm)\,, \tag{2.7.1}$$

where $A \subset \mathbb{R}^{\infty}$ and that is the $x = (x_1, x_2, \ldots)$ are random variables with values in $\mathbb{R}^{\infty}$. Let $m^{\infty} = m \times m \times \cdots$ denote the distribution of an i.i.d. sequence on $\mathbb{R}^{\infty}$, based on the measure m. Therefore, we have defined the distribution of a sequence which is a mixture of i.i.d. sequences.

The following discussion of de Finetti's Theorem is from Aldous [1985].

Definition 2.7.3. Suppose that $(\Omega, \mathcal{S})$ and $(Y, \mathcal{A})$ are measurable spaces. A **probability kernel** with source $(\Omega, \mathcal{S})$ and target $(Y, \mathcal{A})$ is a map $\mu : \Omega \times \mathcal{A} \to [0, 1]$ such that

1. $A \mapsto \mu\,(\omega, A)$ is a probablity measure on $\mathcal{A}$, for each $\omega \in \Omega$.
2. $\omega \mapsto \mu\,(\omega, A)$ is a random variable, for each $A \subset \mathcal{A}$.

Some authors use the terms random measure, Markov kernel or stochastic kernel rather than probability kernel.

A very simple example would be with $\Omega = Y = [0,1]$ and $\mathcal{S} = \mathcal{A}$ being the Borel subsets of $[0,1]$. The map

$$\mu(\omega, A) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A, \end{cases}$$

is a probability kernel. For each $\omega \in [0,1]$, the measure $A \mapsto \mu(\omega, A)$ is just the Dirac measure at ω. For each $A \in \mathcal{A}$, the function $\omega \mapsto \mu(\omega, A)$ is just the characteristic function χ_A of A.

Suppose now that $(\Omega, \mathcal{S}, \nu)$ is a probability space and that $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{S}$. For each $f \in L^1(\Omega, \mathcal{S}, \nu)$ the Radon-Nikodým theorem ensures that there exists an $\mathcal{F}$-measurable function g on Ω such that

$$\int_S f \, d\nu = \int_S g \, d\nu$$

for all $S \in \mathcal{S}$. The function g is unique up to sets of ν-measure zero and is called the **conditional expectation** of f with repect to $\mathcal{F}$, written $\mathbb{E}(f|\mathcal{F})$. The operator $\mathbb{E}(\cdot|\mathcal{F})$ is in fact a norm 1 projection of $L^1(\Omega, \mathcal{S}, \nu)$ onto $L^1(\Omega, \mathcal{F}, \nu)$.

Suppose now that $x : (\Omega, \mathcal{S}, \nu) \to (Y, \mathcal{A})$ is a random variable. Each $A \in \mathcal{A}$ determines a function $x_A : \Omega \to \{0,1\}$ by

$$x_A(\omega) = \begin{cases} 1, & x(\omega) \in A, \\ 0, & x(\omega) \notin A. \end{cases}$$

Given a sub-σ-algebra $\mathcal{F}$ of $\mathcal{S}$ there exists a probability kernel μ with source $(\Omega, \mathcal{F})$ and target $(Y, \mathcal{A})$ such that for every $A \in \mathcal{A}$

$$\mu(\omega, A) = \mathbb{E}(x_A|\mathcal{F})(\omega) \tag{2.7.2}$$

for almost all $\omega \in \Omega$. This probability kernel is called a **regular conditional distribution (r.c.d.)** of x given $\mathcal{F}$.

Remark 2.7.4. [Sun, 2014] Let $x : (\Omega, \mathcal{S}, \nu) \to (Y, \mathcal{A})$ be as defined in Definition 2.7.3. If Y is a complete separable metric space with Borel σ-field $\mathcal{A}$, then there exists a r.c.d. $(\mu(\omega, \cdot))_{\omega \in \Omega}$ for x given $\mathcal{F}$.

Example 2.7.5. Let $\Omega = [0,1]$, $\mathcal{S}$ be the Borel subsets of $[0,1]$ and let ν be the Lebesgue measure on $[0,1]$. Let $Y = \prod_{i=1}^{\infty} \mathbb{Z}_2$ with $\mathcal{A}$ the natural product σ-algebra. Let $x : \Omega \to Y$ be

$$x(t) = (t_1, t_2, \ldots), \qquad \text{where } t = \sum_{k=1}^{\infty} \frac{t_k}{2^k}.$$

Let $\mathcal{F} = B \times B \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \ldots$ where B is the σ-algebra for the power set, $\mathscr{P}(\mathbb{Z}_2)$, on $\mathbb{Z}_2$, so $\mathcal{F}$ is generated by 4 'atoms'. (If we identify Y with $[0,1)$ these atoms are just $\left[0, \frac{1}{4}\right), \left[\frac{1}{4}, \frac{1}{2}\right), \left[\frac{1}{2}, \frac{3}{4}\right)$and $\left[\frac{3}{4}, 1\right)$.) Given $\omega \in \Omega$, let D_ω be the atom containing $x(\omega)$. A small calculation then shows that for any $\omega \in \Omega$ and $A \in \mathcal{A}$.

$$\mathbb{E}(x_A | \mathcal{F})(\omega) = \frac{\nu(A \cap D_\omega)}{\nu(D_\omega)}.$$

One can check that setting $\mu(\omega, A) = \mathbb{E}(x_A | \mathcal{F})(\omega)$ does define a probability kernel and so is a regular conditional distribution of x given $\mathcal{F}$.

Note that in general, 'the' regular conditional distribution of x given $\mathcal{F}$ is not uniquely determined.

Remark 2.7.6. An alternative version of de Finetti's Theorem can be given without directly referring to a random measure. Here we give the definition of a conditionally i.i.d. sequence.

DEFINITION 2.7.7. Let (x_i) be random variables and $\mathcal{F}$ a σ-algebra. The standard definition of a **conditionally i.i.d**. sequence (x_i) given $\mathcal{F}$ is given by the properties below. We say that (x_i) is a mixture of i.i.d.'s directed by μ if $\mu(\omega)_{i=1}^{\infty}$ is a r.c.d. for (x_i) given $\mathcal{F}$. This condition is equivalent to

(i) the (x_i) are conditionally i.i.d. given $\mathcal{F}$. We can write this as

$$\mathrm{Prob}\left(x_i \in A|\mathcal{F}\right) = \mathrm{Prob}\left(x_j \in A|\mathcal{F}\right) \text{ a.e., for each } A \subseteq \mathbb{R},\ i \neq j.$$

and,

$$\mathrm{Prob}\left(x_i \in A, 1 \leq i \leq n|\mathcal{F}\right) = \prod_{i=1}^{n} \mathrm{Prob}\left(x_i \in A|\mathcal{F}\right).$$

(ii) the conditional distribution of (x_i) given $\mathcal{F}$ is μ. Using our notation here gives Equation 2.7.2, that is

$$\mu\left(\omega, A\right) = \mathbb{E}\left(x_A|\mathcal{F}\right)(\omega).$$

The following lemma allows us to define a mixture of i.i.d. sequences if we have a conditionally i.i.d. sequence.

LEMMA 2.7.8. *[Aldous, 1985, Lemma 2.12]. Let (x_i) be conditionally i.i.d. given $\mathcal{F}$ and μ be a r.c.d. for (x_i) given $\mathcal{F}$. Then*

(i) A mixture of i.i.d. 's directed by μ is given by (x_i).

(ii) Given μ, (x_i) and $\mathcal{F}$ are conditionally independent.

We give an example of conditionally i.i.d. sequences here.

EXAMPLE 2.7.9. This example is based on Lauritzen [2007]. The sequence of outcomes from the toss of a coin are as described in Example 2.6.1 where we let $(X, \mathcal{F})$ be a measure space. A binary sequence x_i is $\{0,1\}$-valued and exchangeable if and only if there exists a distribution function ν such that for all n,

$$\text{Prob}\left(x_1 = c_1, \ldots, x_n = c_n\right) = \int_0^1 p^{t_n}\left(1-p\right)^{n-t_n}\, d\nu\left(p\right),$$

where $t_n = \sum_{i=1}^{n} c_i$. Note that as defined here, ν is the distribution of the limiting frequency

$$y = \limsup_{n\to\infty} \sum_{i=1}^{n} \frac{x_i}{n} = p, \tag{2.7.3}$$

where

$$\text{Prob}\left(y \leq z\right) = \nu\left(z\right) \tag{2.7.4}$$

That is, the sample average converges to the expected value, p, where p is the probability of tossing a head. This can be defined by the conditional probability

$$\text{Prob}\left(x_1 = c_1, \ldots, x_n = c_n | y = p\right) = p^{t_n}\left(1-p\right)^{n-t_n}.$$

That is, for a single toss,

$$\text{Prob}\left(x_i = t | y = p\right) = \begin{cases} p & \text{if } t = 1 \\ (1-p) & \text{if } t = 0 \end{cases}$$

Conditionally on $y = p$, $x = (x_1, \ldots, x_n, \ldots)$ are independent and Binomial with parameter p.

This example gives a conditionally i.i.d. sequence which by Lemma 2.7.8 above is a mixture of i.i.d. sequences. We can say that (x_i) is a mixture of i.i.d. sequences if and only if the (x_i) are conditionally i.i.d. given the σ-algebra $\mathcal{F}$ generated by the random variable y. The (x_i) in this example are a mixture of i.i.d. sequences which are conditionally independent given $y = p$. Hence by de Finetti's Theorem (Theorem 2.7.1) and Lemma 2.7.8, exchangeable sequences are of this form.

2.8. Finite case of de Finetti's Theorem

De Finetti's Theorem may fail for finite sequences $x = (x_1, \ldots, x_n)$. There are many examples of this in the literature. See, for example, Diaconis [1977] which uses a geometric interpretation of independence and exchangeability resulting in an understanding of the failure of de Finetti's Theorem for a finite exchangeable sequence. [Olav, 2005, p. 30] notes that while de Finetti's Theorem may fail for finite sequences we may derive a general representation formula for exchangeable distributions that resembles that for the infinite case.

EXAMPLE 2.8.1. Let $X = \{0,1\} \times \{0,1\}$. Define $x = (x_1, x_2)$ as a pair of random variables with $x_i \in \{0,1\}$ with distribution

		X_1	
		0	1
X_2	0	0	$\frac{1}{2}$
	1	$\frac{1}{2}$	0

Let $\hat{x} = (x_2, x_1)$ be the non-trivial permutation of x. Then clearly the distribution of x amd $\hat{x}$ are the same so x is exchangeable (but not an i.i.d. pair).

An i.i.d. pair $y = (y_1, y_2)$ would have distribution

$$\text{Prob}\,(y_1 = t_1, y_2 = t_2) = p^t\,(1-p)^{2-t}\,,$$

where $t = t_1 + t_2$. If x is a mixture of i.i.d. sequences then

$$\text{Prob}\,(x_1 = t_1, x_2 = t_2) = \int_0^1 p^t\,(1-p)^{2-t}\,d\mu\,(p)\,, \tag{2.8.1}$$

for some measure μ on $[0,1]$.

Diaconis and Freedman, 1980a show that for finite exchangeable sequences $(x_1, x_2, \ldots, x_n)$, Equation 2.8.1 need not hold. This is illustrated by the following example. Consider the following finite exchangeable sequence where we let

$$\text{Prob}\,(x_1 = 0,\; x_2 = 0) = \text{Prob}\,(x_1 = 1,\; x_2 = 1) = 0.$$

Here x_1 and x_2 are exchangeable, but if a representation like Equation 2.8.1 holds then we get the following for

$$\text{Prob}\,(x_1 = 0,\; x_2 = 0) = 0$$

we have $t_1 = t_2 = 0$ gives $\int_0^1 p^0\,(1-p)^{2-0}\,\mu\,(dp) = \int_0^1 (1-p)^2\,\mu\,(dp) = 0$. Similar working gives for

$$\text{Prob}\,(x_1 = 1,\; x_2 = 1) = 0$$

we have $t_1 = t_2 = 1$ hence $\int_0^1 p^2\,(1-p)^0\,\mu\,(dp) = \int_0^1 p^2\mu\,(dp) = 0$. Therefore, if the finite representation applies we have

$$\int_0^1 p^2\mu\,(dp) = \int_0^1 (1-p)^2\,\mu\,(dp) = 0.$$

This implies that μ puts the probability of one at both 0 and 1, which is impossible.

Therefore, this shows that de Finetti's Theorem may not apply for finite sequences.

REMARK 2.8.2. Although, de Finetti's Theorem potentially fails for finite sequences we may be able to derive a general representation formula for exchangeable distributions that is similar to that for the infinite case. We are still able to derive a similar property for exchangeable distributions to that of the conditional i.i.d. property in the infinite case. Suppose that $X_1 = X_2 = \cdots = X_n = X$, $Z = \prod_{i=1}^n X_i$ and $x = (x_1, \ldots, x_n) \in \prod_{i=1}^n X_i$. We can define the Dirac measure[1] by $\mu_x = \sum_{k=1}^n \delta_{x_k}$, this is a measure on X. The associated factorial measure $\mu_x^{(n)}$ on Z is defined by

$$\mu_x^{(n)} = \sum_{\sigma \in \Sigma_n} \delta_{\sigma \circ x}.$$

The measures μ_x and $\mu_x^{(n)}$ are clearly unchanged if the order of the elements $x_1, \ldots, x_n$ is permuted. An **urn sequence**, $z = (z_1, \ldots, z_n)$, is obtained by successive drawing without replacement from a finite set. The measure $\mu_x^{(n)}/n!$ arises as the distribution of the urn sequence $x = (x_1, \ldots, x_n)$. Every finite exchangeable sequence is a mixture of urn sequences. For a proof see Olav [2005, pp. 31-32]. Note that $\mu_x^{(n)}$ has the one-dimensional marginals μ.

EXAMPLE 2.8.3. We can give a simple example of a Dirac and a factorial measure as follows. Let $n = 2$, $X_i = \{0, 1, 2\}$, $x = (0, 2)$ and $y = (y_1, y_2) \subseteq Z$.

[1]Note that this is not a Dirac measure but a sum of Dirac measures.

The factorial measure $\mu_x^{(n)}$ on Z is given by

Z	0	1	2
0	0	0	1
1	0	0	0
2	1	0	0
μ_x	1	0	1

which has μ_x as its marginal distribution.

REMARK 2.8.4. This section builds up to an important theorem by Diaconis and Freedman [1980a, p. 746] for finite exchangeable sequences by providing the definitions used in the theorem.

Let $x_1, x_2, \ldots, x_k, x_{k+1}, \ldots x_n$ be exchangeable random variables taking values in a set $S = \{s_1, \ldots s_l\}$. Let S be a finite set of cardinality l and S^k be the set of k-tuples of elements of S. That is, $x = (x_i)_{i=1}^n$ takes values $x_i \in S$ where $x \in X = \prod_{i=1}^n S_i$ and $\nu \in \mathfrak{P}(S)$, the set of probabilities on S.

De Finetti's Theorem shows that each exchangeable sequence of random variables $x = (x_i)_{i=1}^\infty$ on an infinite product space occurs as a mixture of i.i.d. random variables. This is not true in general for finite exchangeable sequences, [Diaconis and Freedman, 1980a] however showed that for a finite exchangeable sequence $x_1, \ldots, x_n$, the distribution of each subsequence $x_1, \ldots, x_k$ is, if $k \leq n$, close in a certain sense to a mixture of i.i.d. random variables. More precisely, if each x_i, takes values in the finite set $S = \{s_1, \ldots s_l\}$ then the variation distance between the distribution of $x_1, \ldots, x_k$ and the closest mixture of i.i.d. random variables is at most $\frac{2kl}{n}$. The variation distance between two measures μ and ν on a measure space $(X, \mathcal{B})$ is (following [Diaconis and Freedman, 1980a][2])

$$\|\mu - \nu\| = 2 \sup |\mu(A) - \nu(A)|$$

where the supremum is taken over all $A \in \mathcal{B}$.

It is more convenient here to work directly with the probability distributions rather than the random variables. Suppose then that $\mathbb{P}$ is a probability distribution on $S^n = \prod_{i=1}^{n} S$ where S has l elements. For each $k \leq n$, this determines a probability distribution $\mathbb{P}_k$ on S^k by taking the corresponding marginal distributions. That is

$$\mathbb{P}_k(A) = \mathbb{P}\left(A \otimes S^{n-k}\right), \qquad A \subseteq S^k.$$

As in Section 2.6, each mixture of i.i.d. distributions on S^k is of the form

$$\mathbb{P}_{\nu k}(A) = \int_{m \in \mathfrak{P}(S)} m^k(A)\, \nu(dm), \qquad A \subseteq S^k \tag{2.8.2}$$

where $\mathfrak{P}(S)$ is the set of probability distributions on S, m^k is the k-fold product measure on S^k generated by $m \in \mathfrak{P}(S)$ and ν is a measure on the Borel subsets of $\mathfrak{P}(S)$.

THEOREM 2.8.5. *Diaconis and Freedman [1980a, p. 746]. Let S be a finite set with l elements and suppose that $\mathbb{P}$ is an exchangeable probability on S^n and S^k is the set of k-tuples of elements of S . Then there exists a probability*

[2] Many authors omit the factor 2.

ν on the Borel subsets of $\mathfrak{P}(S)$ such that for all $k \leq n$,

$$\|\mathbb{P}_k - \mathbb{P}_{\nu k}\| \leq \frac{2kl}{n}. \tag{2.8.3}$$

In other words, if $\mathbb{P}$ is an exchangeable probability distribution on S^n, then there is a probability distribution $\mathbb{Q} = \mathbb{Q}_\nu$ on S^n which is a mixture of i.i.d. distributions and for which all of the projections satisfy the bounds $\|\mathbb{P}_k - \mathbb{Q}_k\| \leq \frac{2kl}{n}$. [Diaconis and Freedman, 1980a] give examples to show that the upper bound here is close to optimal.

The proof of Theorem 2.8.5 uses the idea of an 'urn sequence'. Consider an urn U which contains n balls, each of which is labelled with one of the l elements of S. (Of course, for a given n and l, the set $\mathfrak{U}(n,l)$ of urns of this type is always finite.) Given $k \leq n$, the urn generates two random variables $x = (x_1, \dots x_k)$ and $y = (y_1, \dots y_k)$ taking values in S^k, resulting from recording the sequence of labels of k balls drawn at random from the urn either without, or with replacement. The distribution of x will be a hypergeometric distribution $H_{U,k}$ and that of y will be a multinomial distribution $M_{U,k}$. Both random variables are clearly exchangeable, but $x_1, \dots x_n$ are not independent. The important point is that every exchangeable distribution $\mathbb{P}$ on S^n is a weighted average of finitely many hypergeometric distributions

$$\mathbb{P} = \sum_{j=1}^{J} c_j H_{U_j,n}$$

with $0 \leq c_j \leq 1$ and $\sum_j^J c_j = 1$ and each $U_j \in \mathfrak{U}(n,l)$. The distribution $\mathbb{Q} = \sum_{j=1}^J c_j M_{U_j,n}$ is then a mixture of i.i.d. random variables giving

$$\|\mathbb{P}_k - \mathbb{Q}_k\| = \left\| \sum_{j=1}^{J} c_j H_{U_j,k} - \sum_{j=1}^{J} c_j M_{U_j,k} \right\| \leq \sum_{j=1}^{J} c_j \left\| H_{U_j,k} - M_{U_j,k} \right\|.$$

The bound in Theorem 2.8.5 then reduces to finding a bound on $\left\|H_{U_j,k} - M_{U_j,k}\right\|$ for $U \in \mathfrak{U}(n,l)$.

[Diaconis and Freedman, 1980a] also use these ideas to determine when an exchangeable distribution on S^k is the projection of one on S^n. That is, suppose that $\mathbb{P}$ is an exchangeable distribution on S^k. If there exists an exchangeable distribution $\hat{\mathbb{P}}$ on S^n so that $\mathbb{P} = \hat{\mathbb{P}}_k$ we shall say that $\mathbb{P}$ can be extended from k-tuples to n-tuples. The distribution in Example 2.8.1 is one which cannot be extended. What is noted by [Diaconis and Freedman, 1980a] is that an exchangeable distribution on S^k can be extended precisely when it is of the form $\mathbb{P} = \sum_{j=1}^{J} c_j H_{U_j,k}$.

Example 2.8.6. The above discussion explains what finite exchangeable sequences taking values in a finite set need to look like, and allows one to construct illustrative examples.

Fix a large n and let $S = \{\text{red, blue}\}$. Take two urns

- Urn U_1 with one red ball and $n-1$ blue balls
- Urn U_2 with $n-1$ red balls and one blue ball.

For $j = 1, 2$ and $k \leq n$, let $H_{U_j,k}$ and $M_{U_j,k}$ be the distributions from taking k random draws from the urns, without and with replacement. Then, for $0 \leq c \leq 1$, $\mathbb{P}_k = cH_{U_1,k} + (1-c)\, H_{U_2,k}$ is an exchangeable probability distribution on S^k and $\mathbb{Q}_k = cM_{U_1,k} + (1-c)\, M_{U_2,k}$ is the corresponding mixture of i.i.d. distributions.

For small values of k one can easily explicitly calculate the various probabilities. For $c = \frac{1}{2}$ and $k = 3$ these are given in the table below.

x	$H_{U_1,3}(x)$	$M_{U_1,3}(x)$	$H_{U_2,3}(x)$	$M_{U_2,3}(x)$	$\mathbb{P}_3$	$\mathbb{Q}_3$
RRR	0	$\frac{1}{n^3}$	$\frac{n-3}{n}$	$\frac{(n-1)^3}{n^3}$	$\frac{n-3}{2n}$	$\frac{1+(n-1)^3}{2n^3}$
RRB	0	$\frac{n-1}{n^3}$	$\frac{1}{n}$	$\frac{(n-1)^2}{n^3}$	$\frac{1}{2n}$	$\frac{n-1}{2n^2}$
RBR	0	$\frac{n-1}{n^3}$	$\frac{1}{n}$	$\frac{(n-1)^2}{n^3}$	$\frac{1}{2n}$	$\frac{n-1}{2n^2}$
RBB	$\frac{1}{n}$	$\frac{(n-1)^2}{n^3}$	0	$\frac{n-1}{n^3}$	$\frac{1}{2n}$	$\frac{n-1}{2n^2}$
BRR	0	$\frac{n-1}{n^3}$	$\frac{1}{n}$	$\frac{(n-1)^2}{n^3}$	$\frac{1}{2n}$	$\frac{n-1}{2n^2}$
BRB	$\frac{1}{n}$	$\frac{(n-1)^2}{n^3}$	0	$\frac{n-1}{n^3}$	$\frac{1}{2n}$	$\frac{n-1}{2n^2}$
BBR	$\frac{1}{n}$	$\frac{(n-1)^2}{n^3}$	0	$\frac{n-1}{n^3}$	$\frac{1}{2n}$	$\frac{n-1}{2n^2}$
BBB	$\frac{n-3}{n}$	$\frac{(n-1)^3}{n^3}$	0	$\frac{1}{n^3}$	$\frac{n-3}{2n}$	$\frac{1+(n-1)^3}{2n^3}$

The variation distance between $\mathbb{P}_3$ and $\mathbb{Q}_3$ in this case is $\frac{3}{n^2}$ which is rather better than the general bound $\frac{12}{n}$ provided by Theorem 2.8.5. More complicated examples could be constructed by increasing the number of colours (the size of S), increasing the number of different urns, or by combining the urn distributions less symmetrically.

Theorem 2.8.7. *Diaconis and Freedman [1980a, p. 746] and [Freedman, 1977]. Let S be a set of infinite cardinality and suppose that $\mathbb{P}$ is an exchangeable probability on S^n and S^k is the set of k-tuples of elements of S. Then there exists a probability ν on the Borel subsets of $\mathfrak{P}(S)$ for all $k \leq n$, such that*

$$\|\mathbb{P}_k - \mathbb{P}_{\nu k}\| \leq \frac{k\,(k-1)}{n}.$$

2.9. Choquet Theory and Other Background to de Finetti's Theorem

There are many different proofs of de Finetti's Theorem in the literature. For a discussion of finite forms of de Finetti's Theorem on exchangeability, see, for example, de Finetti [1972a, p. 213], Ericson [1973], Diaconis [1977], Diaconis and Freedman [1980a] and Kerns and Székely [2006]. For the infinite case, some proofs include Diaconis and Freedman [1980a] and Olav [2005].

We give two proofs of de Finetti's Theorem in this thesis. The first is by Glasner [2003]. The second proof which uses martingales is given in Chapter four. The following definitions are used in the first proof.

Definition 2.9.1. [Makarenkov, p. 22]. Let $(X, \mathcal{B})$ and $(Y, \mathcal{F})$ be two measure spaces, suppose that $\pi : X \to Y$ is a measurable map and μ is a measure on X. We define the **push-forward measure** $\pi_*\mu$ on $(Y, \mathcal{F})$ by the formula $\pi_*\mu\,(A) := \mu\,(\pi^{-1}\,(A))$ for every $A \in \mathcal{F}$.

Definition 2.9.2. Suppose that V is a vector space. Then $\{v_1, \ldots, v_k\} \subseteq V$ are **affinely independent** if and only if for each collection of m distinct points $\{w_1, \ldots, w_m\} \in \{v_1, \ldots, v_k\}$ there is no m-dimensional affine subspace containing $\{w_1, \ldots, w_m\}$. Equivalently, $\{v_1, \ldots, v_k\}$ are affinely independent if the vectors $v_2 - v_1, v_3 - v_1 \ldots, v_k - v_1$ are linearly independent in $\mathbb{R}^n$.

[Basener, 2006] and [Taylor, 2009] define the **convex hull** of $\{v_1, \ldots, v_k\} \subseteq V$ to be the smallest convex set containing $\{v_1, \ldots, v_k\}$. It is denoted $CH\{v_1, \ldots, v_k\}$. Thus

$$\begin{aligned} CH\{v_1, \ldots, v_k\} &= \Big\{ w \in \mathbb{R}^n : \text{there exists } c_1, \ldots c_k \in [0,1] \text{ such that} \\ &\qquad w = \sum c_i v_i \text{ and } c_1 + c_2 + \cdots + c_k = 1 \Big\}. \end{aligned}$$

Since all k-dimensional simplices are homeomorphic we shall talk about the standard k-dimensional simplex and denote this by Δ^k. The vertices $\{v_1, \ldots, v_k\}$ give the set

$$\Delta^k = \{c_1 v_1 + c_2 v_2 + \cdots + c_k v_k \in \mathbb{R}^n : c_1 + c_2 + \cdots + c_k = 1 \text{ and } c_1, \ldots c_k \in [0,1]\}.$$

The standard k-dimensional **simplex** or unit k-simplex is the convex hull of k affinely independent points. That is, a simplex is the smallest convex set which contains the given vertices.

Definition 2.9.3. Rudin [1966, p. 251] A point, $x \in K$, in a convex set K is called an **extreme point** if x does not lie in the interior of any line segment joining two points of K.

Theorem 2.9.4. *[Elliott et al., 1999, p. 194] and [Phelps, 2001]. Let X be a compact convex subset of a locally convex topological space E. Then X is the closed convex hull, denoted $\overline{CH}$, of its extreme points.*

Definition 2.9.5. A **Choquet simplex** is a compact convex subset of a locally convex topological space which is also a simplex.

Haydon [1975, p. 97]. The **extreme boundary** of a compact convex set is the set of all extreme points of that set. [Elliott et al., 1999, p. 195] A Choquet simplex with closed extreme boundary is called a **Bauer simplex**.

The following discussion of Prokhorov's theorem is based on Capiński and Kopp [2004], Tao [2009] and Lauriničikas [2012].

We use Prokhorov's Theorem to prove that there is a measure. We begin by defining when a sequence of Borel probability measures is tight.

DEFINITION 2.9.6. A topological space is **σ-compact** if it is the union of countably many compact subspaces.

Let X be a locally compact metric space which is σ-compact and let μ_n be a sequence of Borel probability measures on X. We call a sequence μ_n **tight** if for every $\varepsilon > 0$ there exists a compact set K such that $\mu_n (X \setminus K) \leq \varepsilon$ for all n.

DEFINITION 2.9.7. A family, $\mathfrak{P}$, of probability measures on X, is **relatively compact** if every sequence of elements contains a weakly convergent subsequence.

THEOREM 2.9.8. ***Prokhorov's Theorem****. If a family of probability measures on $(X, \mathcal{B})$ is tight then it is relatively compact.*

That is, if μ_n is a tight sequence in $\mathfrak{P}(X)$ then there exists a subsequence μ_{n_k} and a probability measure $\mu \in \mathfrak{P}(X)$ such that

$$\mu_{n_k} \xrightarrow{w} \mu.$$

2.10. Proof of de Finetti's Theorem

In this section we shall give a proof of a special case of de Finetti's theorem which characterises exchangeable random variables taking values in an infinite product of finite spaces.

Suppose then that $X = \{x_1, \dots x_n\}$ is a finite set, equipped with the power set σ-algebra, and let $Z = \prod_{k=1}^{\infty} X$, with the usual product σ-algebra $\mathcal{B}$. Note that in this case $\mathfrak{P}(\mathrm{X})$ is just an n-simplex. Let $\mathfrak{J}$ denote the set of all exchangeable probability measures on $(Z, \mathcal{B})$.

$$\mathfrak{J} = \{\mathbb{P} \in \mathfrak{P}\,(\mathrm{Z}) : \mathbb{P}\,(\sigma \mathrm{A}) = \mathbb{P}\,(\mathrm{A}) \text{ for all } \mathrm{A} \in \mathcal{B} \text{ and } \sigma \in \Sigma_\infty\}\,.$$

Our aim is to show that each $\mathbb{P} \in \mathfrak{J}$ has a unique representation

$$\mathbb{P}\,(A) = \int_{m \in \mathfrak{P}(X)} m^\infty\,(A)\,\nu\,(dm)\,, \quad A \in \mathcal{B} \tag{2.10.1}$$

for some measure $\nu \in \mathfrak{P}\,(\mathfrak{P}(\mathrm{X}))\,.$

THEOREM 2.10.1. ***De Finetti's Theorem**: Every exchangeable measure on Z is an average of product measures. That is, if $\mathbb{P} \in \mathfrak{J}$, then there exists $\nu \in \mathfrak{P}\,(\mathfrak{P}(\mathrm{X}))$ such that*

$$\mathbb{P} = \phi\,(\nu) = \int_{m \in \mathfrak{P}(X)} m^\infty\, d\nu\,(m)\,.$$

Note first that given any $m \in \mathfrak{P}\,(X)$ and $A \in \mathcal{B}$, $m^\infty\,(A) = m^\infty\,(A)^2$, so for any $\nu \in \mathfrak{P}\,(\mathfrak{P}(\mathrm{X}))$, the integral $\phi\,(\nu) = \int_{m \in \mathfrak{P}(X)} m^\infty \nu\,(dm)$ does define an exchangeable measure $(Z, \mathcal{B})$. Our aim then is to show that the map ϕ : $\mathfrak{P}\,(\mathfrak{P}(\mathrm{X})) \to \mathfrak{J}$ is onto.

The proof of this depends on the simplicial structure of the sets $\mathfrak{P}\,(\mathfrak{P}(\mathrm{X}))$ and $\mathfrak{J}$. Since $\mathfrak{P}(\mathrm{X})$ is a complete metric space, it follows (from Prokhorov's

Theorem, for example) that the set $\mathfrak{P}(\mathfrak{P}(\mathrm{X}))$ is compact (in the usual weak topology on the space of signed Borel measures on $\mathfrak{P}(\mathrm{X})$). It is easy to see that $\mathfrak{P}(\mathfrak{P}(\mathrm{X}))$ is convex and hence that it is a Choquet simplex. Indeed the extreme points of $\mathfrak{P}(\mathfrak{P}(\mathrm{X}))$ are precisely the Dirac measure δ_m for $m \in \mathfrak{P}(\mathrm{X})$ (see for example, [Simon, 2011, Example 8.16 in Chapter 8]). This set of extreme points is closed and so $\mathfrak{P}(\mathfrak{P}(\mathrm{X}))$ forms a Bauer simplex.

The space $\mathfrak{I}$ also shares this structure. Again it is easy to check that $\mathfrak{I}$ is a closed subset of the weakly compact set of probability measures on Z and hence is compact, and that $\mathfrak{I}$ is convex.

FACT 2.10.2. *[Glasner, 2003, p. 167] Two Bauer simplices are affinely homeomorphic if and only if their closed sets of extreme points are homeomorphic compact spaces.*

Recall that $A \in \mathcal{B}$ is μ**-a.e.** Σ_∞**-invariant** if $\mu(\sigma A \triangle A) = 0$ for all $\sigma \in \Sigma_\infty$, and that the measure μ is Σ_∞**-ergodic** if every μ-a.e. Σ_∞-invariant set A has $\mu(A)$ either zero or one.

In light of Fact 2.10.2, it is sufficient that we check that ϕ maps the extreme points of $\mathfrak{P}(\mathfrak{P}(\mathrm{X}))$ bi-continuously onto the extreme points of $\mathfrak{I}$. In particular, $\mathfrak{I}$ is a simplex, the simplex of Σ_∞-invariant Borel probability measures on the compact space Z. Therefore, as we prove below a measure $\mathbb{P} \in \mathfrak{I}$ is an extreme point if and only if it is a Σ_∞-ergodic measure in $\mathfrak{I}$.

Let $\mathfrak{I} = \{\mu \in \mathfrak{P}(Z) : \mu \text{ is } \Sigma_\infty\text{-invariant}\}$.

LEMMA 2.10.3. *A measure μ on $(Z, \mathcal{B})$ is Σ_∞-ergodic if and only if it is an extreme point of $\mathfrak{I}$.*

PROOF. Suppose first that $\mu \in \mathfrak{I}$ is *not* ergodic. Then there exists $\sigma \in \Sigma_\infty$ and a μ-a.e.invariant set $A \in \mathcal{B}$ with $0 < \mu(A) < 1$. Let $B = Z \setminus A$. Define the disjoint probability measures μ_A and μ_B by

$$\mu_A(F) = \frac{\mu(A \cap F)}{\mu(A)}, \quad \mu_B(F) = \frac{\mu(B \cap F)}{\mu(B)}.$$

Now, if $F \in \mathcal{B}$ and $\sigma \in \Sigma_\infty$, then

$$\begin{aligned}
\mu_A\left(\sigma^{-1}F\right) &= \frac{\mu\left(A \cap \sigma^{-1}F\right)}{\mu(A)} \\
&= \frac{\mu\left(\sigma^{-1}A \cap \sigma^{-1}F\right)}{\mu(A)} \quad \text{(as } A \text{ is } \mu\text{-invariant)} \\
&= \frac{\mu\left(\sigma^{-1}(A \cap F)\right)}{\mu(A)} \\
&= \frac{\mu(A \cap F)}{\mu(A)} \quad (\text{as } \mu \in \mathfrak{J}) \\
&= \mu_A(F).
\end{aligned}$$

Thus μ_A (and obviously also μ_B) is in $\mathfrak{J}$. Now it is easy to check that

$$\mu = \mu(A)\,\mu_A + \mu(B)\,\mu_B = \mu(A)\,\mu_A + (1 - \mu(A))\,\mu_B$$

and hence that $\mu \notin \operatorname{Ext}(\mathfrak{J})$.

Suppose conversely that $\mu \in \mathfrak{J}$ is not an extreme point of $\mathfrak{J}$. That is

$$\mu = \alpha\mu_1 + (1 - \alpha)\,\mu_2$$

for some $\alpha \in (0, 1)$ and distinct μ_1, $\mu_2 \in \mathfrak{J}$. The following theorem is similar to Theorem 2.2.6. Since $\mu \in \mathfrak{J}$, the ergodic theorem says that if $\sigma \in \Sigma_\infty$ and $f \in L^1(Z, \mathcal{B}, \mu)$, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f\left(\sigma^i x\right) = \int_Z f\,d\mu$$

for μ-almost all x. Of course μ_1, $\mu_2 \in \mathfrak{I}$, so for each j,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f\left(\sigma^i x\right) = \int_Z f\, d\mu_j$$

for μ_j-almost all x. Elementary measure theory says that we must therefore have

$$\int_Z f\, d\mu = \int_Z f\, d\mu_1 = \int_Z f\, d\mu_2.$$

The continuous functions on Z already separate the elements of $\mathfrak{P}(Z)$, and so this is more than enough to deduce that $\mu = \mu_1 = \mu_2$ which contradicts that μ is not an extreme point. □

LEMMA 2.10.4. *For any $m \in \mathfrak{P}(X)$, m^∞ is Σ_∞-ergodic. Thus, ϕ maps the set of extreme points of $\mathfrak{P}(\mathfrak{P}(\mathrm{X}))$ into the set of extreme points of $\mathfrak{I}$.*

PROOF. This proof is based on [Walters, 1982, pp. 32-33]. Suppose that $A \in \mathcal{B}$ is m^∞-a.e. Σ_∞-invariant set. For $n = 1, 2, 3, \ldots$, let

$$A_n = \{x = (x_i) \in Z : \text{ there exists } y = (y_i) \in A \text{ such that } x_i = y_i \text{ for } 1 \le i \le n\}$$

denote the set of sequences whose first n coordinates match the first n coordinates of some element of A. Thus $A = \bigcap_{n=1}^{\infty} A_n$ and so $m^\infty(A) = \lim_{n\to\infty} m^\infty(A_n)$.

Fix $\varepsilon > 0$. Choose n so that $m^\infty(A_n \setminus A) < \varepsilon$. Let $\sigma \in \Sigma_\infty$ be the permutation which swaps the first n coordinates with the next n, and let $B = \sigma A_n$. Since m^∞ is a product measure,

$$m^\infty(A_n \cap B) = m^\infty(A_n)^2.$$

Now (using the general set/measure theoretic inequalities $|m^\infty(F) - m^\infty(G)| \leq m^\infty(F\Delta G) \leq m^\infty(F\Delta H) + m^\infty(H\Delta G)$ for any suitable sets F, G, H),

$$\begin{aligned} |m^\infty(A) - m^\infty(A_n \cap B)| &\leq m^\infty(A\Delta(A_n \cap B)) \\ &\leq m^\infty(A\Delta A_n) + m^\infty(A\Delta B) \\ &\leq m^\infty(A\Delta A_n) + m^\infty(A\Delta\sigma A) + m^\infty(\sigma A\Delta B) \\ &\leq \varepsilon + 0 + m^\infty(A\Delta A_n) \\ &\leq 2\varepsilon. \end{aligned}$$

Also

$$\begin{aligned} \left|m^\infty(A_n \cap B) - m^\infty(A)^2\right| &= \left|m^\infty(A_n)^2 - m^\infty(A)^2\right| \\ &= |m^\infty(A_n) - m^\infty(A)| \cdot |m^\infty(A_n) + m^\infty(A)| \\ &\leq 2\varepsilon. \end{aligned}$$

Thus

$$\left|m^\infty(A) - m^\infty(A)^2\right| \leq |m^\infty(A) - m^\infty(A_n \cap B)| + \left|m^\infty(A_n \cap B) - m^\infty(A)^2\right| \leq 4\varepsilon.$$

Since this is true for all $\varepsilon > 0$, we have that $m^\infty(A) = m^\infty(A)^2$ and hence that $m^\infty(A)$ is either zero or one. $\square$

LEMMA 2.10.5. *If* $\mathbb{P} \in \mathfrak{J}$ *is* Σ_∞*-ergodic, then* $\mathbb{P} = m^\infty = \phi(\delta_m)$ *for some* $m \in \mathfrak{P}(X)$.

PROOF. Suppose that $\mathbb{P} \in \mathfrak{J}$ is Σ_∞-ergodic. For $n = 1, 2, 3, \ldots$ let $\pi_n : Z \to X$ be the coordinate map $\pi_n(x) = \pi_n(x_1, x_2, x_3, \ldots) = x_n$. Define the measure $m_n = m_{n,\mathbb{P}}$ on X by the push-forward measure,

$$m_n(A) = \mathbb{P}\left(\pi_n^{-1}(A)\right).$$

Since $\mathbb{P} \in \mathfrak{J}$, it is clear that these measures each have an identical distribution which we will denote m. If we can show that these measures are independent then (by [Williams, 1991, Section 8.7] for example), $\mathbb{P} = m^{\infty}$.

It suffices to show that for all $n, k \geq 1$ and $a, b \in X$

$$\mathbb{P}(\pi_n(x) = a \text{ and } \pi_k(x) = b) = \mathbb{P}(\pi_n(x) = a)\, \mathbb{P}(\pi_k(x) = b).$$

Let $\mathbb{1}_{n,a}$ denote the indicator function which takes the values 1 if $\pi_n(x) = a$ and 0 otherwise. Thus $\mathbb{P}(\pi_n(x) = a) = \int_Z \mathbb{1}_{n,a}(x)\, d\mathbb{P}(x)$. By the Mean Ergodic Theorem (similar to Theorem 2.2.6) it follows that

$$\begin{aligned}
\mathbb{P}(\pi_n(x) = a)\, \mathbb{P}(\pi_k(x) = b) &= \int_Z \mathbb{1}_{n,a}(x)\, d\mathbb{P}(x) \int_Z \mathbb{1}_{k,b}(y)\, d\mathbb{P}(y) \\
&= \lim_{N\to\infty} \frac{1}{N!} \sum_{\sigma \in \Sigma_N} \int_Z \mathbb{1}_{n,a}(\sigma x)\, \mathbb{1}_{k,b}(x)\, d\mathbb{P}(x).
\end{aligned}$$

Now, let $H_N = \{\sigma \in \Sigma_N : \sigma(n) = k\}$. If $\sigma \in \Sigma_N \setminus H_N$, then there exists $\sigma' \in \Sigma_N$ such that $\sigma'(k) = k$ and $(\sigma' \circ \sigma)(n) = n$. In this case, since $\mathbb{P} \in \mathfrak{J}$

$$\begin{aligned}
\int_Z \mathbb{1}_{n,a}(\sigma x)\, \mathbb{1}_{k,b}(x)\, d\mathbb{P}(x). &= \int_Z \mathbb{1}_{n,a}((\sigma' \circ \sigma)(x))\, \mathbb{1}_{k,b}(\sigma'(x))\, d\mathbb{P}(x) \\
&= \int_Z \mathbb{1}_{n,a}(x)\, \mathbb{1}_{k,b}(x)\, d\mathbb{P}(x) \\
&= \mathbb{P}(\pi_n(x) = a \text{ and } \pi_k(x) = b).
\end{aligned}$$

Clearly $|H_N| = (N-1)!$ and so

$$\begin{aligned}
&\mathbb{P}(\pi_n(x) = a)\,\mathbb{P}(\pi_k(x) = b) \\
&\quad = \lim_{N\to\infty} \frac{1}{N!} \sum_{\sigma\in H_N} \int_Z \mathbb{1}_{n,a}(\sigma x)\,\mathbb{1}_{k,b}(x)\,d\mathbb{P}(x) + \sum_{\sigma\notin H_N} \int_Z \mathbb{1}_{n,a}(\sigma x)\,\mathbb{1}_{k,b}(x)\,d\mathbb{P}(x) \\
&\quad \le \lim_{N\to\infty} \frac{(N-1)!}{N!} + \frac{N! - (N-1)!}{N!}\mathbb{P}(\pi_n(x) = a \text{ and } \pi_k(x) = b) \\
&\quad = \mathbb{P}(\pi_n(x) = a \text{ and } \pi_k(x) = b)
\end{aligned}$$

as required. □

REMARK 2.10.6. This proof of de Finetti's Theorem is for the case where m^∞ in Equation 2.10.1 is an invariant measure under Σ_∞. We can see this because the proof that for any $m \in \mathfrak{P}(X)$, m^∞ is Σ_∞-ergodic assumes that we have a Σ_∞-invariant set. In the other direction suppose that we have a measure which is Σ_∞-ergodic then this measure is m^∞. Note that this also does not apply for the quasi-invariant case. This is because the pushforward measure used in this proof $m_n(A)$ depends on n in the quasi-invariant case and we need it to not depend on n .

2.11. Extensions of de Finetti's Theorem

There are many applications of exchangeability in the literature. Diaconis and Freedman [1980b] discuss a generalisation of exchangeability called partial exchangeability. Similarly, Aldous [1981] discuss arrays of random variables such that the rows or columns are exchangeable and show that these arrays may be represented as functions of underlying i.i.d. random variables. Some of the uses of de Finetti's Theorem are discussed in Aldous [2010].

CHAPTER 3

G-measures And Quasi-Invariant Measures

The space $Z = \prod_{n=1}^{\infty} \mathbb{Z}_2$ forms a compact space which can be identified, via binary expansions, with the interval $[0, 1]$ or the circle $\mathbb{T}$. In this chapter we shall look at the actions of various groups of transformations that act on Z. Our starting point is the work of Brown and Dooley [1991] which addressed the problem of characterising the probability measures on Z which are quasi-invariant and ergodic under the action of the group $\Gamma = \oplus_{n=1}^{\infty} \mathbb{Z}_2$ of finite coordinate changes.

The central concept in Brown and Dooley's work is that of a 'G-measure'. The first step in this direction is to give a correspondence between probability measures on Z and certain families of functions $G = \{G_F\}_F$ on Z. A measure which corresponds to a family G is called a G-measure. Their main result gives conditions on the family G under which the corresponding G-measure is unique and hence ergodic for the action of Γ.

In this chapter we extend this analysis to the case where the group Γ is replaced by the group Σ_∞ of permutations of finitely many components of Z. To this end we shall define 'G-measures' for this permutation group. In addition to the discussion of G-measures for the general case of a finite group. This is original content.

3.1. Brown and Dooley's G-functions

We show how to use a probability measure μ on Z to construct a family $\{G_{F,\mu}\}_F$ of functions on Z. The aim here is to investigate appropriate conditions on $\{G_{F,\mu}\}$ so that μ is quasi-invariant for the action of the group of finite coordinate changes Γ. These conditions will then be adapted to provide an analogous theory for measures which are quasi-invariant for the action of Σ_∞.

Suppose then that μ is a probability measure on Z. Let $\mathcal{F}$ denote the set of finite subsets F of $\mathbb{Z}^+$. For a finite subset $F \in \mathcal{F}$ we can define the group of changes in finitely many coordinates by

$$\Gamma_F = \{\gamma \in \Gamma : \gamma_i = 0, \forall\, i \notin F\}.$$

Corresponding to Γ_F is a tail measure on Z,

$$\mu^F = \mu^F_\Gamma = \frac{1}{|\Gamma_F|} \sum_{\gamma \in \Gamma_F} \mu \circ \gamma, \tag{3.1.1}$$

The tail measure defines a family of functions $G = \{G_{F,\mu}\}$ via the Radon-Nikodým derivative $G_{F,\mu} = \frac{d\mu}{d\mu^F_\Gamma}$. We shall write G_F rather than $G_{F,\mu}$ if the measure μ is understood.

If $\mu \in \mathfrak{P}(Z)$, where $\mathfrak{P}(Z)$ is the set of probability measures on Z, then these functions $\{G_{F,\mu}\}$ satisfy certain natural conditions. To motivate these we shall examine the special case when μ (acting on $[0,1]$) is of the form $f\,dx$ where f is a positive continuous function and dx is Lebesgue measure on $[0,1]$. (Looking at this special case avoids having to deal with making sense of indeterminate forms in the calculations below.) We define

$$f_F(x) = \frac{1}{|\Gamma_F|} \sum_{\gamma \in \Gamma_F} f\left(\gamma^{-1}(x)\right),$$

so that the Radon-Nikodým derivative $\frac{d\mu}{d\mu^F}$ is the function

$$\begin{aligned} G_F(x) &= \frac{d\mu}{d\mu^F}(x) \\ &= \frac{d\mu}{dx}\cdot\frac{dx}{d\mu^F} \\ &= \frac{f(x)}{f_F(x)} \\ &= \frac{|\Gamma_F|\, f(x)}{\sum_{\gamma\in\Gamma_F} f(\gamma^{-1}(x))}. \end{aligned}$$

Note that we are using the fact that dx is invariant under $\gamma \in \Gamma_F$.

For $x \in [0,1]$,

$$\frac{1}{|\Gamma_F|}\sum_{\eta\in\Gamma_F} G_F(\eta(x)) = \sum_{\eta\in\Gamma_F}\frac{f(\eta(x))}{\sum_{\gamma\in\Gamma_F} f(\gamma^{-1}(\eta x))} = \sum_{\eta\in\Gamma_F}\frac{f(\eta(x))}{\sum_{\gamma\in\Gamma_F} f(\gamma^{-1}(x))} = 1.$$

Also, if $\eta \in \Gamma_F$ then

$$G_F(\eta(x)) = \frac{|\Gamma_F|\, f(\eta x)}{\sum_{\gamma\in\Gamma_F} f(\gamma^{-1}(\eta x))} = \frac{|\Gamma_F|\, f(\eta x)}{\sum_{\gamma\in\Gamma_F} f(\gamma^{-1}(x))}.$$

Thus, if $F_1 \subseteq F_2 \in \mathcal{F}$, $\eta \in \Gamma_{F_1}$ and $x \in [0,1]$ then

$$\begin{aligned} G_{F_1}(x)\, G_{F_2}(\eta(x)) &= \frac{|\Gamma_{F_1}|\, f(x)}{\sum_{\gamma\in\Gamma_{F_1}} f(\gamma^{-1}(x))}\cdot\frac{|\Gamma_{F_2}|\, f(\eta(x))}{\sum_{\gamma\in\Gamma_{F_2}} f(\gamma^{-1}(x))} \\ &= \frac{|\Gamma_{F_1}|\, f(\eta(x))}{\sum_{\gamma\in\Gamma_{F_1}} f(\gamma^{-1}(x))}\cdot\frac{|\Gamma_{F_2}|\, f(x)}{\sum_{\gamma\in\Gamma_{F_2}} f(\gamma^{-1}(x))} \\ &= G_{F_1}(\eta x)\, G_{F_2}(x). \end{aligned}$$

These calculations motivate the following definition.

Definition 3.1.1. Suppose that $G = \{G_F\}$ is a family of Borel functions on Z indexed by finite subsets $F \subseteq Z$. Then the family G is

- **compatible** if for $F_1 \subseteq F_2$

$$G_{F_1}(x)\, G_{F_2}(\gamma x) = G_{F_1}(\gamma x)\, G_{F_2}(x) \quad (\gamma \in \Gamma_{F_1}, x \in Z)$$

- **normalised** if for all $F \in \mathcal{F}$

$$\frac{1}{|\Gamma_F|} \sum_{\gamma \in \Gamma_F} G_F(\gamma x) = 1.$$

DEFINITION 3.1.2. Suppose that $G = \{G_F\}_{F \in \mathcal{F}}$ is a normalised compatible family of Borel functions on Z. If $\mu \in \mathfrak{P}(Z)$ has Radon-Nikodým derivatives which satisfy $G_F = \frac{d\mu}{d\mu_\Gamma^F}$ (a.e.) for each $F \in \mathcal{F}$, then we shall say that μ is a G-**measure**.

This definition raises several obvious questions:

(1) Which probability measures are G-measures for some family G?
(2) Which families G admit a G-measure?
(3) If G admits a G-measure, is it unique?

For the group Γ, these questions were addressed by Brown and Dooley [1991].

THEOREM 3.1.3. *Brown and Dooley [1991, p. 280]. If $\mu \in \mathfrak{P}(Z)$ then $\{G_{F,\mu}\}$ is a normalised compatible family.*

THEOREM 3.1.4. *Brown and Dooley [1991, Proposition 2]. If G is a normalised compatible family then there exists $\mu \in \mathfrak{P}(Z)$ such that μ is a G-measure.*

REMARK 3.1.5. The more delicate question is question 3 above. The main point of Brown and Dooley [1991] is to state conditions on a normalised compatible family G which guarantees the uniqueness, and hence the unique ergodicity of the corresponding G-measure.

3.2. g-functions and cocycles

Brown and Dooley [1991] introduce a number of important concepts associated to G-measures, such as g-functions and h-cocycles.

Let $G_n = G_{\{1,\ldots,n\}}$. This sequence determines another sequence of functions $\{g_n\}$ satisfying

$$G_n = g_1 \cdots g_n \text{ for } n = 1, 2, \ldots.$$

Let $\mu^n = \mu^{\{1,\ldots,n\}}$. Then

$$g_n = \frac{d\mu^{n-1}}{d\mu^n}.$$

The normalisation condition on the G-functions implies one for the little g-functions. This gives the normalisation condition for the little g-functions in Brown and Dooley [1991]

$$\frac{1}{\left|\Gamma_{\{i\}}\right|} \sum_{\eta \in \Gamma_{\{i\}}} g_i\left(\eta\left(x\right)\right) = 1. \tag{3.2.1}$$

Example 3.2.1. In the context of a measure $\mu = f\, dx$ considered earlier. Let

$$g_1\left(x\right) = G_1\left(x\right) = \frac{f\left(x\right)}{f_{\{1\}}\left(x\right)}$$

$$g_2\left(x\right) = \frac{G_2\left(x\right)}{G_1\left(x\right)} = \frac{\left|\Gamma_{\{1,2\}}\right| f\left(x\right)}{\sum_{\gamma \in \Gamma_{\{1,2\}}} f\left(\gamma^{-1}\left(x\right)\right)} \cdot \frac{\sum_{\gamma \in \Gamma\{1\}} f\left(\gamma^{-1}\left(x\right)\right)}{\left|\Gamma_{\{1\}}\right|}$$

$$g_3\left(x\right) = \frac{G_3\left(x\right)}{G_2\left(x\right)} = \frac{\left|\Gamma_{\{1,2,3\}}\right| f\left(x\right)}{\sum_{\gamma \in \Gamma_{\{1,2,3\}}} f\left(\gamma^{-1}\left(x\right)\right)} \cdot \frac{\sum_{\gamma \in \Gamma_{\{1,2\}}} f\left(\gamma^{-1}\left(x\right)\right)}{\left|\Gamma_{\{1,2\}}\right|}$$

etc.

Note that the functions g_i satisfy the normalisation constraint in Equation 3.2.1

Now if we fix $i \in \mathbb{N}$ then

$$\sum_{\eta\in\Gamma_{\{i\}}} g_i(\eta(x)) = \frac{\left|\Gamma_{\{1,\ldots,i\}}\right|}{\left|\Gamma_{\{1,\ldots,i-1\}}\right|}\sum_{\eta\in\Gamma_{\{i\}}}\frac{\sum_{\gamma\in\Gamma_{\{1,\ldots,i-1\}}} f(\gamma^{-1}(\eta x))}{\sum_{\gamma\in\Gamma_{\{1,\ldots,i\}}} f(\gamma^{-1}(\eta x))} = \left|\Gamma_{\{i\}}\right| \frac{\sum_{\eta\in\Gamma_{\{i\}}}\sum_{\gamma\in\Gamma_{\{1,\ldots,i-1\}}} f(\gamma^{-1}(\eta x))}{\sum_{\gamma\in\Gamma_{\{1,\ldots,i\}}} f(\gamma^{-1}(\eta x))} = \left|\Gamma_{\{i\}}\right|. \tag{3.2.2}$$

In the case when μ is a quasi-invariant measure for the action of Γ_F on $Z = \prod_{i=1}^{\infty}\mathbb{Z}_l$. We may associate to μ the map $h : Z\times\Gamma_F \to \mathbb{R}$

$$h(x,\gamma) = \left(\frac{d(\mu\circ\gamma)}{d\mu}\right)(x).$$

We can then derive the family (G_F) from h by

$$G_F(x) = \left(\frac{1}{|\Gamma_F|}\sum_{\gamma\in\Gamma_F} h(x,\gamma)\right)^{-1} \qquad (x\in Z). \tag{3.2.3}$$

Note that h may be constructed from G_F as follows

$$h(x,\gamma) = \frac{G_F(\gamma x)}{G_F(x)} \qquad (\gamma\in\Gamma_F,\ x\in Z). \tag{3.2.4}$$

In the case of $\mu = f\,dx$ considered earlier we have

$$h(x,\gamma) = \left(\frac{d(\mu\circ\gamma)}{d\mu}\right)(x) = \frac{f(\gamma^{-1}(x))}{f(x)}.$$

Therefore

$$\begin{aligned}\frac{1}{|\Gamma_F|}\sum_{\gamma\in\Gamma_F} h(x,\gamma) &= \frac{1}{|\Gamma_F|}\sum_{\gamma\in\Gamma_F}\frac{f(\gamma^{-1}(x))}{f(x)}\\ &= \left(\frac{|\Gamma_F|\, f(x)}{\sum_{\gamma\in\Gamma_F} f(\gamma^{-1}(x))}\right)^{-1}\\ &= G_F(x)^{-1.}\end{aligned}$$

An important property of the map h is that it is a cocycle on $Z \times \Gamma_F$.

Definition 3.2.2. **Cocycle formalism**. Let Z be a set and G a group acting on Z. A map $h : Z \times G \to \mathbb{R}$ is a **cocycle** if

$$h(x, g_1 g_2) = h(x, g_2)\, h(g_2 x, g_1) \quad (x \in Z,\ g_1, g_2 \in G).$$

Remark 3.2.3. It is clear that $h(x, \gamma)$ is a cocycle. We can see this as follows

$$\begin{aligned} h(x, \gamma_1\gamma_2) &= \left(\frac{d(\mu \circ \gamma_1 \circ \gamma_2)}{d\mu}\right)(x) \\ &= \left(\frac{d(\mu \circ \gamma_2)}{d\mu}\right)(x) \cdot \left(\frac{d(\mu \circ \gamma_1)}{d\mu}\right)(\gamma_2 x) \qquad (3.2.5) \\ &= h(x, \gamma_2)\, h(\gamma_2 x, \gamma_1) \end{aligned}$$

Note that

$$h(x, \gamma_1\gamma_2) = \frac{G_F(\gamma_1\gamma_2 x)}{G_F(x)},\ h(x, \gamma_2) = \frac{G_F(\gamma_2 x)}{G_F(x)},\ \text{and } h(\gamma_2 x, \gamma_1) = \frac{G_F(\gamma_1\gamma_2 x)}{G_F(\gamma_2 x)},$$

We can see that Equation 3.2.5 holds using the relationship between cocycles and G-measures, since clearly the following equation holds.

$$\frac{G_F(\gamma_1\gamma_2 x)}{G_F(x)} = \frac{G_F(\gamma_2 x)}{G_F(x)} \cdot \frac{G_F(\gamma_1\gamma_2 x)}{G_F(\gamma_2 x)}.$$

Remark 3.2.4. Note that in this chapter we consider $Z = \prod_{n=1}^{\infty} \mathbb{Z}_2 = [0, 1]$ acted on by the group $\Gamma = \oplus_{n=1}^{\infty}\mathbb{Z}_2$ of changes in finitely many coordinates. In the discussion in this chapter we could easily swap $\mathbb{Z}_2$ for $\mathbb{Z}_l$, where $l \geq 2$ is an integer. It is not written in a way that depends on the base of the expansions.

We could also consider $\prod_{k=1}^{\infty} \mathbb{Z}_{l(k)}$ acted on by the group $\Gamma = \oplus_{k=1}^{\infty} \mathbb{Z}_{l(k)}$, that is the possible values of the coordinates can vary depending on k.

3.3. The General Case of a Finite Group

Much of the construction of Brown and Dooley [1991] does not depend on the group Γ of transformations that they consider. In this section we shall reproduce some of the ideas from that construction in the setting of a general group of measurable transformations on a measure space. In particular we shall show that finite groups of transformations always provide an averaging operator which is a conditional expectation operator with respect to the σ-algebra of sets which are invariant under the group.

Suppose that $\mathfrak{G}$ is a finite group of measurable transformations on[1] $Z = \prod_{n=1}^{\infty} \mathbb{Z}_2$, and that μ is a probability measure on $(Z, \mathcal{B})$ which is quasi-invariant for the action of $\mathfrak{G}$. This implies that for all $\tau \in \mathfrak{G}$ the measures $\mu \circ \tau$ and μ are equivalent, and that, at least for bounded measurable f,

$$\int_Z f(x)\ d(\mu \circ \tau)(x) = \int_Z f\left(\tau^{-1} x\right)\ d\mu(x).$$

More generally, if A is a measurable set, then this implies that

$$\int_A f(x)\ d(\mu \circ \tau)(x) = \int_{\tau A} f\left(\tau^{-1} x\right)\ d\mu(x).$$

Since $\mu \circ \tau$ and μ are equivalent, we can form the Radon-Nikodým derivative $\frac{d(\mu \circ \tau)}{d\mu}$.

Consider the map $h : Z \times \mathfrak{G} \to \mathbb{R}$ defined by

$$h(x, \tau) = \frac{d(\mu \circ \tau)}{d\mu}(x)$$

[1]The proof here does not depend on Z being an infinite product. Any measurable space would do!

so that $\mu(\tau A) = \int_A h(x,\tau)\, d\mu$. Suppose that $x \in Z$ and $\tau_1, \tau_2 \in \mathfrak{G}$. For a measurable set A

$$\begin{aligned}
\int_A h(x,\tau_2)\, h(\tau_2 x, \tau_1)\, d\mu &= \int_A h(\tau_2 x, \tau_1)\, d(\mu \circ \tau_2) \\
&= \int_{\tau_2 A} h\left(\tau_2^{-1}\tau_2 x, \tau_1\right)\, d\mu \\
&= \int_{\tau_2 A} h(x, \tau_1)\, d\mu \\
&= \mu(\tau_1 \tau_2 A) \\
&= \int_A h(x, \tau_1\tau_2)\, d\mu.
\end{aligned}$$

It follows that $h(x,\tau_1\tau_2) = h(x,\tau_2)\, h(\tau_2 x, \tau_1)$ (μ-a.e.) and hence that h is a cocycle.

Define

$$\mu_{\mathfrak{G}} = \frac{1}{|\mathfrak{G}|} \sum_{\tau \in \mathfrak{G}} \mu \circ \tau = \frac{1}{|\mathfrak{G}|} \sum_{\tau \in \mathfrak{G}} \mu \circ \tau^{-1}$$

which is again equivalent to μ and so we can find the Radon-Nikodým derivative $G_{\mathfrak{G}} : Z \to \mathbb{R}$,

$$G_{\mathfrak{G}} = \frac{d\mu}{d\mu_{\mathfrak{G}}}$$

which is certainly non-negative.

Note that

$$\frac{d\mu_{\mathfrak{G}}}{d\mu}(x) = \frac{1}{|\mathfrak{G}|} \sum_{\tau \in \mathfrak{G}} h(x,\tau), \qquad x \in Z$$

and so

$$G_{\mathfrak{G}}(x) = \frac{|\mathfrak{G}|}{\sum_{\tau \in \mathfrak{G}} h(x,\tau)}.$$

It follows that if we fix $x \in Z$ and $\tau \in \mathfrak{G}$ then, using the cocycle property,

$$\begin{aligned}
\frac{1}{|\mathfrak{G}|}\sum_{\tau\in\mathfrak{G}} G_{\mathfrak{G}}(\tau x) &= \sum_{\tau\in\mathfrak{G}} \frac{1}{\sum_{\tau'\in\mathfrak{G}} h(\tau x, \tau')} \\
&= \sum_{\tau\in\mathfrak{G}} \frac{1}{\sum_{\tau'\in\mathfrak{G}} h(x, \tau\tau')/h(x,\tau)} \\
&= \sum_{\tau\in\mathfrak{G}} \frac{h(x,\tau)}{\sum_{\tau'\in\mathfrak{G}} h(x, \tau\tau')} \\
&= \sum_{\tau\in\mathfrak{G}} \frac{h(x,\tau)}{\sum_{\tau'\in\mathfrak{G}} h(x, \tau')} \\
&= 1.
\end{aligned}$$

This will of course correspond to Brown and Dooley's normalisation condition.

A very similar calculation to the one for the normalisation constraint shows that

$$\frac{G_{\mathfrak{G}}(\tau x)}{G_{\mathfrak{G}}(x)} = \frac{\sum_{\tau'\in\mathfrak{G}} h(x,\tau')}{\sum_{\tau'\in\mathfrak{G}} h(\tau x,\tau')} = \frac{h(x,\tau)\sum_{\tau'\in\mathfrak{G}} h(x,\tau')}{\sum_{\tau'\in\mathfrak{G}} h(x,\tau')} = h(x,\tau).$$

Suppose that $f \in L^1(Z,\mathcal{B},\mu)$. Then

$$\begin{aligned}
\int_Z f(x)\, d\mu(x) &= \int_Z G_{\mathfrak{G}}(x) f(x)\, d\mu_{\mathfrak{G}}(x) \\
&= \frac{1}{|\mathfrak{G}|}\sum_{\tau\in\mathfrak{G}} \int_Z G_{\mathfrak{G}}(x) f(x)\, d\left(\mu\circ\tau^{-1}\right)(x) \\
&= \frac{1}{|\mathfrak{G}|}\sum_{\tau\in\mathfrak{G}} \int_Z G_{\mathfrak{G}}(\tau x) f(\tau x)\, d\mu(x). \qquad (3.3.1)
\end{aligned}$$

Define $T : L^1(Z,\mathcal{B},\mu) \to L^1(Z,\mathcal{B},\mu)$ by

$$Tf(x) = \frac{1}{|\mathfrak{G}|}\sum_{\tau\in\mathfrak{G}} G_{\mathfrak{G}}(\tau x) f(\tau x).$$

The map is clearly linear and

$$\begin{aligned}\int_Z |Tf(x)|\, d\mu(x) &= \int_Z \left| \frac{1}{|\mathfrak{G}|} \sum_{\tau\in\mathfrak{G}} G_{\mathfrak{G}}(\tau x) f(\tau x) \right| d\mu(x) \\ &\leq \int_Z \frac{1}{|\mathfrak{G}|} \sum_{\tau\in\mathfrak{G}} G_{\mathfrak{G}}(\tau x) |f(\tau x)|\, d\mu(x) \\ &= \int_Z |f(\tau x)|\, d\mu(x)\end{aligned}$$

by Equation 3.3.1.

Then

$$\begin{aligned}T^2 f(x) &= \frac{1}{|\mathfrak{G}|} \sum_{\tau\in\mathfrak{G}} G_{\mathfrak{G}}(\tau x)\, Tf(\tau x) \\ &= \frac{1}{|\mathfrak{G}|^2} \sum_{\tau\in\mathfrak{G}} G_{\mathfrak{G}}(\tau x) \sum_{\tau'\in\mathfrak{G}} G_{\mathfrak{G}}(\tau'\tau x) f(\tau'\tau x) \\ &= \frac{1}{|\mathfrak{G}|^2} \sum_{\tau\in\mathfrak{G}} G_{\mathfrak{G}}(\tau x) \sum_{\gamma'\in\mathfrak{G}} G_{\mathfrak{G}}(\tau' x) f(\tau' x) \\ &= \frac{1}{|\mathfrak{G}|} \sum_{\tau\in\mathfrak{G}} G_{\mathfrak{G}}(\tau x)\, Tf(x) \\ &= Tf(x) \qquad (3.3.2)\end{aligned}$$

using the normalisation property derived above.

Also $T1(x) = \frac{1}{|\mathfrak{G}|}\sum_{\tau\in\mathfrak{G}} G_{\mathfrak{G}}(\tau x) = 1$ μ-a.e. By a theorem of Douglas [1965] then, T is a conditional expectation operator.

Let $\mathcal{C}_{\mathfrak{G}}$ be the σ-algebra of all Borel subsets of Z which are invariant under $\mathfrak{G}$. Then every $\mathcal{C}_{\mathfrak{G}}$ measurable L^1 function is invariant under T and hence is in the range of T, and Tf is always $\mathcal{C}_{\mathfrak{G}}$-measurable. This implies that $T = \mathbb{E}(\cdot|\mathcal{C}_{\mathfrak{G}})$.

Suppose now that $\mathfrak{G}_1$ is a subgroup of $\mathfrak{G}$. The subgroup $\mathfrak{G}_1$ has a corresponding cocycle h_1 and function $G_{\mathfrak{G}_1}$.

Suppose that $x \in Z$ and $\tau \in \mathfrak{G}_1$. From the definition we see that $h_1(x,\tau) = h(x,\tau)$.

Thus, using the cocycle property as above,

$$\begin{aligned} G_{\mathfrak{G}_1}(x)\, G_{\mathfrak{G}}(\tau x) &= \frac{|\mathfrak{G}_1|}{\sum_{\tau_1 \in \mathfrak{G}_1} h_1(x,\tau_1)} \cdot \frac{|\mathfrak{G}|}{\sum_{\tau_2 \in \mathfrak{G}} h(\tau x, \tau_2)} \\ &= \frac{|\mathfrak{G}_1|}{\sum_{\tau_1 \in \mathfrak{G}_1} h_1(x,\tau_1)} \cdot \frac{|\mathfrak{G}|\, h(x,\tau)}{\sum_{\tau_2 \in \mathfrak{G}} h(x, \tau_2 \tau)} \\ &= \frac{|\mathfrak{G}_1|\, h_1(x,\tau)}{\sum_{\tau_1 \in \mathfrak{G}_1} h_1(x,\tau_1 \tau)} \cdot \frac{|\mathfrak{G}|}{\sum_{\tau_2 \in \mathfrak{G}} h(x, \tau_2)} \\ &= \frac{|\mathfrak{G}_1|}{\sum_{\tau_1 \in \mathfrak{G}_1} h_1(\tau x,\tau_1)} \cdot \frac{|\mathfrak{G}|}{\sum_{\tau_2 \in \mathfrak{G}} h(x, \tau_2)} \\ &= G_{\mathfrak{G}_1}(\tau x)\, G_{\mathfrak{G}}(x). \end{aligned}$$

This property then corresponds to the compatibility property from Brown and Dooley [1991].

The results of this section of course allow us to deal with families $\{\mathfrak{G}_F\}_F$ of finite groups of transformations. In the case of Brown and Dooley [1991], the index set ranges over finite subsets F of $\mathbb{Z}^+$, and the group $\mathfrak{G}_F$ is the group Γ_F acting on $Z = \prod_{n=1}^{\infty} \mathbb{Z}_2$ considered earlier. The case that we will be interested in here is where $\mathfrak{G}_F$ is Σ_F, the set of permutations of F.

In either case, given a family $\{\mathfrak{G}_F\}_F$ of finite groups of transformations and a measure μ which is quasi-invariant under all the groups, we obtain a 'compatible, normalised' family of functions $\{G_F\}_F$. The next step is to examine to what extent one can reverse the procedure and construct a quasi-invariant measure from a suitable family of given functions.

In this section we are going to extend the G-measure formalism in Brown and Dooley, 1991 to actions of the symmetric group, Σ_∞. Then almost all of the calculations in the last section go through unchanged. The one big difference is in writing a normalisation constraint for the little g-functions. In Equation 3.2.2 above, that depends on the fact that $\Gamma_{\{1,\ldots,i\}} = \Gamma_{\{1,\ldots,i-1\}} \oplus \Gamma_{\{i\}}$ so that

$$\sum_{\eta\in\Gamma_{\{i\}}} \sum_{\gamma\in\Gamma_{\{1,\ldots,i-1\}}} f\left(\gamma^{-1}(\eta x)\right) = \sum_{\gamma\in\Gamma_{\{1,\ldots,i\}}} f\left(\gamma^{-1}(\eta x)\right).$$

For $i \in \mathbb{N}$ and $1 \le k \le i$, define $\eta_{i,k} \in \Sigma_{\{1,\ldots,i\}}$ to be the permutations which swap the kth and ith elements. Every element $\sigma \in \Sigma_{\{1,\ldots,i\}}$ factors uniquely as $\sigma = \sigma' \circ \eta_{i,k}$ for some k and some $\sigma' \in \Sigma_{\{1,\ldots,i-1\}}$.Thus

$$\sum_{k=1}^{i} \sum_{\sigma\in\Sigma_{\{1,\ldots,i-1\}}} f\left(\sigma^{-1}\left(\eta_{i,k}(x)\right)\right) = \sum_{\sigma\in\Sigma_{\{1,\ldots,i\}}} f\left(\sigma^{-1}\left(\eta_{i,k}(x)\right)\right).$$

It follows then that

$$\begin{aligned}
\sum_{k=1}^{i} g_i\left(\eta_{i,k}(x)\right) &= \frac{\left|\sum_{\{1,\ldots i\}}\right|}{\left|\sum_{\{1,\ldots i-1\}}\right|} \sum_{k=1}^{i} \frac{\sum_{\sigma\in\Sigma_{\{1,\ldots,i-1\}}} f\left(\sigma^{-1}\left(\eta_{i,k}(x)\right)\right)}{\sum_{\sigma\in\Sigma_{\{1,\ldots,i\}}} f\left(\sigma^{-1}\left(\eta_{i,k}(x)\right)\right)} \\
&= i\frac{\sum_{\eta\in\Sigma_{\{i\}}} \sum_{\sigma\in\Sigma_{\{1,\ldots,i-1\}}} f\left(\sigma^{-1}\left(\eta_{i,k}(x)\right)\right)}{\sum_{\sigma\in\Sigma_{\{1,\ldots,i\}}} f\left(\sigma^{-1}\left(\eta_{i,k}(x)\right)\right)} \\
&= i
\end{aligned}$$

or

$$\frac{1}{i}\sum_{k=1}^{i} g_i\left(\eta_{i,k}(x)\right) = 1. \tag{3.3.3}$$

For g-functions the following equation for Σ-normalisation is satisfied.

$$\frac{1}{n}\sum_{\sigma\in\Sigma_n/\Sigma_{n-1}} g_n(\sigma x) = 1.$$

3.4. Formal definitions of Σ-G-measures

In this section we give a formal definition of Σ-G-measures.

DEFINITION 3.4.1. Suppose that $G = \{G_F\}_{F\in\mathcal{F}}$ is a family of non-negative Borel functions on Z. We say that G is

- **Σ-compatible** if

$$G_{F_1}(x)\, G_{F_2}(\sigma x) = G_{F_1}(\sigma x)\, G_{F_2}(x) \quad (\sigma \in \Sigma_{F_1},\, x \in Z,\, F_1 \subseteq F_2)\,. \tag{3.4.1}$$

- **Σ-normalised** if

$$\frac{1}{|\Sigma_F|}\sum_{\sigma\in\Sigma_F} G_F(\sigma x) = 1. \quad x \in Z,\, F \in \mathcal{F}. \tag{3.4.2}$$

DEFINITION 3.4.2. Suppose that $G = \{G_F\}$ is a Σ-normalised, Σ-compatible family of Borel functions on Z. A probability measure μ on Z is called a Σ-G-**measure** if $G_F = \frac{d\mu}{d\mu^F}$ (with respect to μ a.e.) for all finite $F \subseteq \mathbb{Z}^+$.

The challenge now is to construct G-measures from such a family of functions, and to determine the conditions under which such a function is determined by the family. It turns out to be not too difficult to construct Σ-G-measures.

THEOREM 3.4.3. *Suppose that $G = \{G_F\}_{F\in\mathcal{F}}$ is a Σ-normalised, Σ-compatible family of functions on $Z = \Pi_{n=1}^{\infty}\mathbb{Z}_2$. Then there exists a G-measure ν on Z.*

PROOF. Fix $x_0 \in Z$. For $F \in \mathcal{F}$, define the measure ν_F by

$$\int_Z f\, d\nu_F = \frac{1}{|\Sigma_F|}\sum_{\sigma\in\Sigma_F} G_F(\sigma x_0)\, f(\sigma x_0)\,.$$

This forms a net of probability measures on Z (where the index set $\mathcal{F}$ is ordered by inclusion).

By the compactness of $\mathfrak{P}(Z)$, this net must have at least one cluster point, say ν which is the weak-$*$ limit of a subnet $\{\nu_{F_\alpha}\}_\alpha$. By definition,

$$\nu^F = \frac{1}{|\Sigma_F|}\sum_{\sigma\in\Sigma_F}\nu\circ\sigma^{-1}.$$

We want to show that ν is a G-measure, that is, for all F

$$G_F = \frac{d\nu}{d\nu^F} \quad (\nu\text{-a.e.})\,.$$

or equivalently, that for all $f\in\mathcal{C}(Z)$,

$$\int_Z f\,G_F\,d\nu^F = \int_Z f\,d\nu.$$

Now

$$\begin{aligned}
\int_Z f(x)\,G_F(x)\;d\nu^F &= \frac{1}{|\Sigma_F|}\sum_{\sigma\in\Sigma_F}\int_Z f(x)\,G_F(x)\;d\left(\nu\circ\sigma^{-1}\right)\\
&= \frac{1}{|\Sigma_F|}\sum_{\sigma\in\Sigma_F}\int_Z f(\sigma x)\,G_F(\sigma x)\;d\nu\\
&= \frac{1}{|\Sigma_F|}\sum_{\sigma\in\Sigma_F}\lim_\alpha\frac{1}{|\Sigma_{F_\alpha}|}\sum_{\tau\in\Sigma_{F_\alpha}}G_{F\alpha}(\tau x_0)\,G_F(\tau\sigma x_0)\,f(\tau\sigma x_0)\\
&= \lim_\alpha\frac{1}{|\Sigma_{F_\alpha}|}\sum_{\tau\in\Sigma_{F_\alpha}}\frac{1}{|\Sigma_F|}\sum_{\sigma\in\Sigma_F}G_{F\alpha}(\tau x_0)\,G_F(\tau\sigma x_0)\,f(\tau\sigma x_0)\\
&= \frac{1}{|\Sigma_F|}\sum_{\sigma\in\Sigma_F}\lim_\alpha\frac{1}{|\Sigma_{F_\alpha}|}\sum_{\tau\in\Sigma_{F_\alpha}}G_{F\alpha}\left(\tau\sigma^{-1}x_0\right)G_F(\tau x_0)\,f(\tau x_0)\\
&= \lim_\alpha\frac{1}{|\Sigma_{F_\alpha}|}\sum_{\tau\in\Sigma_{F_\alpha}}\frac{1}{|\Sigma_F|}\sum_{\sigma\in\Sigma_F}G_{F\alpha}\left(\tau\sigma^{-1}x_0\right)G_F(\tau x_0)\,f(\tau x_0)
\end{aligned}$$

Suppose now that $F\subseteq F_\alpha$. For each $\tau\in\Sigma_{F_\alpha}$, let $x_\tau=\tau x_0$, and so $\tau\sigma^{-1}x_0 = \tau\sigma^{-1}\tau^{-1}x_\tau$. Thus

$$G_{F\alpha}\left(\tau\sigma^{-1}x_0\right)G_F(\tau x_0)\,f(\tau x_0) = G_{F\alpha}\left(\tau\sigma^{-1}\tau^{-1}x_\tau\right)G_F(x_\tau)\,f(x_\tau)$$

The set $\{\tau\sigma^{-1}\tau^{-1}\}_{\sigma\in F}$ is just $\Sigma_{\tau F}$. Thus

$$\begin{aligned}
\frac{1}{|\Sigma_F|}\sum_{\sigma\in\Sigma_F} G_{F\alpha}\left(\tau\sigma^{-1}x_0\right)G_F\left(\tau x_0\right)f\left(\tau x_0\right) &= \frac{1}{|\Sigma_{\tau F}|}\sum_{\sigma\in\Sigma_{\tau F}} G_{F\alpha}\left(\sigma x_\tau\right)G_F\left(x_\tau\right)f\left(x_\tau\right)\\
&= \frac{1}{|\Sigma_{\tau F}|}\sum_{\sigma\in\Sigma_{\tau F}} G_{F\alpha}\left(x_\tau\right)G_F\left(\sigma x_\tau\right)f\left(x_\tau\right)\\
&= G_{F\alpha}\left(x_\tau\right)f\left(x_\tau\right)\frac{1}{|\Sigma_{\tau F}|}\sum_{\sigma\in\Sigma_{\tau F}} G_F\left(\sigma x_\tau\right)\\
&= G_{F\alpha}\left(\tau x_0\right)f\left(\tau x_0\right)
\end{aligned}$$

using the Σ-normalisation and Σ-compatibility conditions. Thus

$$\begin{aligned}
\int_Z f(x)\,G_F(x)\,d\nu^F &= \lim_\alpha \frac{1}{|\Sigma_{F_\alpha}|}\sum_{\tau\in\Sigma_{F_\alpha}} G_{F\alpha}\left(\tau x_0\right)f\left(\tau x_0\right)\\
&= \lim_\alpha \int_Z f\,d\nu_F\\
&= \int_Z f\,d\nu
\end{aligned}$$

as required. □

Our next aim is to prove a converse of Theorem 3.4.3. Suppose that $\mu \in \mathfrak{P}(Z)$. Then for any finite set F, $\mu \ll \mu^F$ and so we can form the Radon-Nikodým derivative $G_F = \frac{d\mu}{d\mu^F}$. The issue here is that G_F is only defined μ-a.e. and so if μ is supported on a small set, we may have G_F undefined on a large set. Since this function is only determined μ-a.e., and so to produce a function which is actually defined on all of Z we need to find a way of defining the function on the μ-null part of Z.

Recall that the support of a measure μ, denoted $\operatorname{supp}(\mu)$, is defined to be the set of all points $x \in Z$ for which every open neighbourhood of x has positive measure. We therefore adapt our definition of G_F as follows

$$G_F(x) = \begin{cases} \dfrac{d\mu}{d\mu^F}(x), & \text{if } x \in \operatorname{supp}(\mu) \\ 1, & \text{if } x \in Z \setminus \operatorname{supp}(\mu) \end{cases}$$

Note that this is still only defined μ-a.e..

THEOREM 3.4.4. *Given $\mu \in \mathfrak{P}(Z)$, the family $G = \{G_F\}$ with G_F defined above, is Σ-normalised and Σ-compatible, so μ is a Σ-G-measure.*

PROOF. We want to show that $G_F = \frac{d\mu}{d\mu^F}$ is Σ-normalised and Σ-compatible. The following discussion uses cocycle formalism for G-measures which we have discussed in more detail in Section 3.2. Then we have

$$\begin{aligned} \frac{1}{|\Sigma_F|}\sum_{\sigma\in\Sigma_F} G_F(\sigma x) &= \frac{1}{|\Sigma_F|}\sum_{\sigma\in\Sigma_F}\left(\frac{1}{|\Sigma_F|}\sum_{\tau\in\Sigma_F} h(\sigma x,\tau)\right)^{-1} \\ &= \sum_{\sigma\in\Sigma_F}\left(\sum_{\tau\in\Sigma_F}\frac{h(x,\tau\sigma)}{h(x,\sigma)}\right)^{-1} \\ &= \sum_{\sigma\in\Sigma_F} h(x,\sigma)\left(\sum_{\tau\in\Sigma_F} h(x,\tau\sigma)\right)^{-1} \\ &= \sum_{\sigma\in\Sigma_F} h(x,\sigma)\left(\sum_{\tau\in\Sigma_F} h(x,\tau)\right)^{-1} \\ &= \frac{\sum_{\sigma\in\Sigma_F} h(x,\sigma)}{\sum_{\tau\in\Sigma_F} h(x,\tau)} = 1. \end{aligned}$$

Therefore, the normalisation constraint is satisfied.

We now show that the Σ-compatibility condition holds. Let $\omega \in \Sigma_{F_1}$, $x \in Z$, $F_1 \subseteq F_2$, then

$$\begin{aligned}
G_{F_1}(x)\, G_{F_2}(\omega x) &= \left(\frac{1}{|\Sigma_{F_1}|}\sum_{\sigma_1\in\Sigma_{F_1}} h(x,\sigma_1)\right)^{-1}\left(\frac{1}{|\Sigma_{F_2}|}\sum_{\sigma_2\in\Sigma_{F_2}} h(\omega x,\sigma_2)\right)^{-1} \\
&= \left(\frac{1}{|\Sigma_{F_1}|}\sum_{\sigma_1\in\Sigma_{F_1}} h(\omega x,\sigma_1)\right)^{-1}\left(\frac{1}{|\Sigma_{F_2}|}\sum_{\sigma_2\in\Sigma_{F_2}} h(x,\sigma_2)\right)^{-1} \\
&= G_{F_1}(\omega x)\, G_{F_2}(x).
\end{aligned}$$

where the second line holds since by Equation 3.2.4, for $\omega \in \Sigma_{F_1}$ and $F_1 \subseteq F_2$,

$$\begin{aligned}
\frac{G_{F_1}(\sigma_1 x)}{G_{F_1}(x)}\cdot\frac{G_{F_2}(\sigma_2 x)}{G_{F_2}(\omega x)} &= \frac{G_{F_1}(\sigma_1 x)}{G_{F_1}(x)}\cdot\frac{G_{F_2}(\sigma_2 x)}{G_{F_1}(\omega x)\, G_{F_2}(x_{F_1+1},\ldots,x_{F_2})} \\
&= \frac{G_{F_1}(\sigma_1 x)}{G_{F_1}(\omega x)}\cdot\frac{G_{F_2}(\sigma_2 x)}{G_{F_1}(x_1,\ldots,x_{F_1})\, G_{F_2}(x_{F_1+1},\ldots,x_{F_2})} \\
&= \frac{G_{F_1}(\sigma_1 x)}{G_{F_1}(\omega x)}\cdot\frac{G_{F_2}(\sigma_2 x)}{G_{F_2}(x)}.
\end{aligned}$$

Therefore, a probability measure μ is a Σ-G-measure. □

In this section we shall look at ways in which the relationship between normalised compatible families G of functions on Z and measures on Z varies between the two groups of transformations we have been examining.

Starting with a Γ-normalised Γ-compatible family G of functions on Z, Brown and Dooley [1991] showed that one always has a G-measure μ. They did this by considering the net of linear functionals $\phi_{x,F} \in \mathcal{C}(Z)^*$ defined by

$$\phi_{x,F}(f) = \frac{1}{|\Gamma_F|}\sum_{\gamma\in\Gamma_F} G_F(\gamma x)\, f(\gamma x).$$

Brown and Dooley [1991, p. 287-6] showed that this net always converges to a constant whose value is independent of x. (Indeed, one can even let the value of x depend on the finite set F and this remains true.) This in turn allowed them to investigate the uniqueness of the G-measure associated with a Γ-normalised Γ-compatible family of functions.

The situation for the permutation group is different. The main reason for this is that for the group of finite coordinate changes, the orbits of points x, that is, the sets $\{\gamma x\}_{\gamma \in \Gamma_n}$, are spread evenly across Z. The corresponding orbits for the permutation group certainly need not have this property. This will typically result therefore in a lack of uniqueness in the G-measures associated to a Σ-normalised Σ-compatible family of functions.

Example 3.4.5. The simplest Σ-normalised Σ-compatible family of functions is one for which $G_F \equiv 1$ for all F. As usual, we define

$$\psi_{x,F}(f) = \frac{1}{|\Sigma_F|} \sum_{\sigma \in \Sigma_F} G_F(\sigma x) f(\sigma x), \quad x \in Z.$$

Since the family $\{\psi_{x,F}\}_F$ is sitting in the unit ball of $\mathcal{C}(Z)^*$ it must certainly have a nonempty set of weak-$*$ limit points. We showed in Theorem 3.4.3 that these are G-measures.

It is not too difficult to see that even in this case, one can get a variety of different G-measures from this construction, depending on the choice of point x.

Case 1. Take $x = 0 = (0, 0, 0, \ldots)$. Then $\sigma x = x$ for any $\sigma \in \Sigma$. This means that $\psi_{x,F}(f) = f(0)$ for all F, and so there is only one limit point. As a measure, this is just the Dirac measure at $x = 0$. That is, $\mu = \delta_0$. In this case $\mu^F = \mu$ so the the Radon-Nikodým derivative

$\frac{d\mu}{d\mu^F}$ is certainly equal to $G_F = 1$ on the support of μ. The issue here is that the support of μ is only a single point in Z.

Case 2. Take $x = 1 = (1, 1, 1, \ldots)$. In this case we get $\psi_{x,F}(f) = f(1)$ for all F, and the corresponding G-measure is δ_1.

Case 3. Take $x = \frac{1}{2} = (1, 0, 0, 0, \ldots)$. If we fix n in this case, then the permutations $\sigma \in \Sigma_n$ move the entry 1 in this list to each of the first n positions with equal frequency. Hence

$$\psi_{\frac{1}{2},F}(f) = \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{1}{2^k}\right).$$

(The formula for a general finite set F is analogous, but slightly messier to express.) It is clear then that if $f \in \mathcal{C}(Z)$, then $\psi_{\frac{1}{2},F}(f) \to f(0)$ and so as in Case 1, we get that the corresponding measure is $\mu = \delta_0$.

Case 4. Take $x = \frac{2}{3} = (1, 0, 1, 0, \ldots) = x_n + t_n$, where x_n is the n-tuple of the first n entries of x and t_n is the remaining tail.

Abusing notation slightly, we can write $\sigma x = \sigma x_n + t_n$ for $\sigma \in \Sigma_n$. To keep things simpler we will take $n = 2m$ even in what follows. This means that x_n has m zeros and m ones. Fix $k \leq n$ and let $y = (y_1, \ldots, y_k)$ be any of the 2^k possible elements on $\prod_{i=1}^{k} \mathbb{Z}_2$. Let $l = \sum_{i=1}^{k} y_i$ denote the number of ones in y, and so there are $k - l$ zeros. The orbit of x under Σ_n, $\mathcal{O}(n, x) = \{\sigma x_n + t_n : \sigma \in \Sigma_n\}$ contains $\frac{(2m)!}{(m!)^2}$ elements each occuring $(m!)^2$ times. Of the elements of $\mathcal{O}(n, x)$,

$$\frac{(2m-k)!}{(m-l)!\,(m-(k-l))!}$$

have the first k digits equal to y. (This is because once you have fixed the first k elements you have $2m - k$ elements left you may permute, divided by the indistinguishable permutations of the ones and of the zeros you have left.)

Let $D_y = y \oplus \prod_{i=k+1}^{\infty} \mathbb{Z}_2$ be the dyadic interval of Z determined by y. Then

$$\begin{aligned} \#\{\sigma \in \Sigma_n : \sigma x_n = y\} &= \#\{\sigma \in \Sigma_n : \sigma x \in D_y\} \\ &= (m!)^2 \cdot \frac{(2m-k)!}{(m-l)!\,(m-(k-l))!}. \end{aligned}$$

Suppose now that $h : Z \to \mathbb{R}$ and consider

$$A_n(h) = \frac{1}{|\Sigma_n|} \sum_{\sigma \in \Sigma_n} h(\sigma x).$$

If h is the charateristic function of D_y, then

$$A_n(h) = \frac{1}{(2m)!} (m!)^2 \cdot \frac{(2m-k)!}{(m-l)!\,(m-(k-l))!} \to \frac{1}{2^k}$$

as $n = 2m \to \infty$ (for any l).

Let $\mathcal{A}_k$ denote the σ-algebra on Z generated by the dyadic intervals of level k. It follows from above that if h is $\mathcal{A}_k$-measurable, then

$$A_n(h) \to \int_Z h \, d\lambda \qquad \text{as even } n \to \infty$$

where λ denotes Lebesgue measure on Z. Note that this holds for all k.

Since every continuous function h on Z can be uniformly approximated by a sequence h_k where h_k is $\mathcal{A}_k$-measurable, this shows that if $h \in \mathcal{C}(Z)$ then

$$A_n(h) \to \int_Z h \, d\lambda \qquad \text{as even } n \to \infty$$

too. This proves that λ is a weak-$*$ limit point of $\{\psi_{x,F_n}\}_{F\in\mathcal{F}}$.

Note however that this is not the only weak-$*$ limit point for this value of x. If one takes F_n to be the subset of $\{1,2,3,\ldots\}$ which includes the first n odd numbers and the first 10^n even numbers, then σx is almost always near zero for $\sigma\in\Sigma_F$, much like the situation in Case 3. Here, as $n\to\infty$, the linear functional ψ_{x,F_n} converges to the Dirac measure at zero. A similar construction would produce a sequence which converges to δ_1.

3.5. Examples of G-measures for the permutation group

We now give some examples of G-measures for the permutation group. We begin with the example of a Bernoulli Scheme for which we show here that $\{G_n\}=1$. We then discuss the example of a Markov measure, which is of particular importance in ergodic theory. If we have a Markov measure then the dynamical system $(X,\mathcal{B},T,\mu)$ is said to be a Markov odometer if μ satisfies certain conditions given in Dooley and Hamachi [2003, p. 102]. Dooley and Hamachi [2003, p. 94] show that when considered as a G-measure, the Markov odometer may be taken to be uniquely ergodic for the group of finite coordinate changes.

Example 3.5.1. The example of a **Bernoulli Scheme** illustrates that some measures are invariant under Σ_∞, for $Z=\prod_{n=1}^{\infty}\mathbb{Z}_2$ and

$$\mu\left\{\{b_1\}\times\cdots\times\{b_k\}\times\mathbb{Z}_2\times\mathbb{Z}_2\times\cdots\right\} \;=\; \prod_{j=1}^{k}p_{b_j},$$

is clearly invariant under Σ_∞ since $\mu^n=\mu$ hence $\frac{d\mu}{d\mu^n}=G_n=1$. Note that the cocycle, $h(x,\sigma)=1$.

EXAMPLE 3.5.2. Another simple measure is a **Markov measure**. Note that in some cases this reduces to the traditional infinite product measure (see Dooley and Hamachi [2003] for more details). A Markov chain is described here from Dooley and Hamachi [2003] and Walters [1982, p. 22].

A Markov chain is a pair (P, π) consisting of a stochastic matrix P and a stationary probability vector π. In general a stochastic matrix is an $n \times n$ matrix $P = (p_{i,j})_{i,j=0}^{n-1}$ such that $p_{i,j} \geq 0$ for each i, j and such that each row sums to one and a probability vector $\pi = (\pi_0, \ldots \pi_{n-1})$ is one for which $\pi_i \geq 0$ for each i, $\sum \pi_i = 1$ and $\pi P = \pi$. In the present setting, we shall take $n = 2$ and assume that $p_{i,j} > 0$ for each i, j.

An example of such a pair is given by

$$P = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}, \qquad \pi = \left(\frac{4}{13}, \frac{9}{13}\right).$$

Given $(z_1, \ldots, z_m) \in \prod_{j=1}^{m} \mathbb{Z}_2$ let

$$A_{z_1,\ldots,z_m} = \{z_1\} \times \cdots \times \{z_m\} \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \ldots$$

denote the cylinder set consisting of sequences that begin with the specified $m + 1$ binary digits. We define the (P, π) Markov measure of such a set by

$$\mu\left(A_{z_1,\ldots,z_m}\right) = \pi_{z_1} p_{z_1,z_2} \cdots p_{z_{m-1},z_m}. \tag{3.5.1}$$

Thus, in the concrete example above

$$\mu\left(A_{0,1,1,0}\right) = \pi_0 p_{0,1} p_{1,1} p_{1,0} = \frac{4}{13} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{39}.$$

This procedure defines a measure on all cylinder sets of $Z = \prod_{i=1}^{\infty} \mathbb{Z}_2$ and hence a measure (also denoted by μ) on $(Z, \mathcal{B})$. The above assumptions on P ensure that every cylinder set has nonzero measure.

The measure μ here is quasi-invariant under the action of Γ, but it is (in general) not invariant. The corresponding normalised compatible family $G_F = \frac{d\mu}{d\mu^F}$ is rather more complicated to write down explicitly here, but certainly gives an example where the functions G_F are not identically 1.

The importance of this example is that it provides the basis for an important theorem of Dooley and Hamachi [2003]. This theorem concerns systems called Markov odometers.

In the setting above, one can define a partial order on Z by setting $x = (x_i) < y = (y_i)$ if there exists n such that $x_n < y_n$ and $x_i = y_i$ for all $i > n$. For almost all $x \in Z$ there is a smallest sequence $T_O(x) \in Z$ which is larger than x. For example

$$\begin{aligned} T_O(011011011) &= 111011011\ldots \\ T_O(111011011) &= 000111011\ldots \end{aligned}$$

(More specifically, to apply T_O one changes the first 0 in x to a 1 and all the 1's up to that point to 0's.)

The dynamical system $(Z, \mathcal{B}, T_O, \mu)$ is an example of a system called a **Markov odometer**. In general T_O is not measure-preserving, but it is nonsingular. Dooley and Hamachi [2003] define Markov odometers uisng the same ideas, but in somewhat greater gernerality, and prove the following theorem. We refer the reader to Dooley and Hamachi [2003] for the details.

THEOREM 3.5.3. *Every ergodic non-singular dynamical system $(X, \mathcal{B}, T, \nu)$ on a standard measure space is orbit equivalent to a Markov odometer. Furthermore, when considered as a G-measure, the Markov odometer may be taken to be uniquely ergodic.*

From now on we will mainly consider Σ-G-measures and so we will usually drop the Σ and just refer to G-measures.

3.6. An Example

In this section we shall construct some concrete examples of a measure and a corresponding Σ-normalised, Σ-compatible family of functions on $Z = \prod_{i=1}^{\infty} \mathbb{Z}_2$.

EXAMPLE 3.6.1. Let μ_1 be the measure on $\mathbb{Z}_2$, where $\mu_1(1) = \frac{3}{4}$ and for $i \geq 2$, let μ_i be the measure on $\mathbb{Z}_2$ with $\mu_i(1) = \frac{1}{2}$. Let $\mu = \otimes_{i=1}^{\infty} \mu_i$ be the corresponding product measure on Z. This of course corresponds to tossing one biased coin, followed by an infinite sequence of fair coins. Thinking of Z as being modelled by $[0, 1]$, the measure μ equals half Lebesgue measure on $\left[0, \frac{1}{2}\right)$ and $\frac{3}{2}$ times Lebesgue measure on $\left[\frac{1}{2}, 1\right)$. Clearly, μ is not invariant under Σ, but it is quasi-invariant.

For $n = 1, 2, 3$, let

$$\mu^n = \frac{1}{|\Sigma_n|} \sum_{\sigma \in \Sigma_n} \mu \circ \sigma.$$

The measure μ^n could be thought of as a symmetrization of μ with respect to the action of Σ_n. Given any binary string b of length n, the μ_n measure of the set of binary sequence which begin with b depends only on the number of zeros and ones in b and not on the order in which they appear.

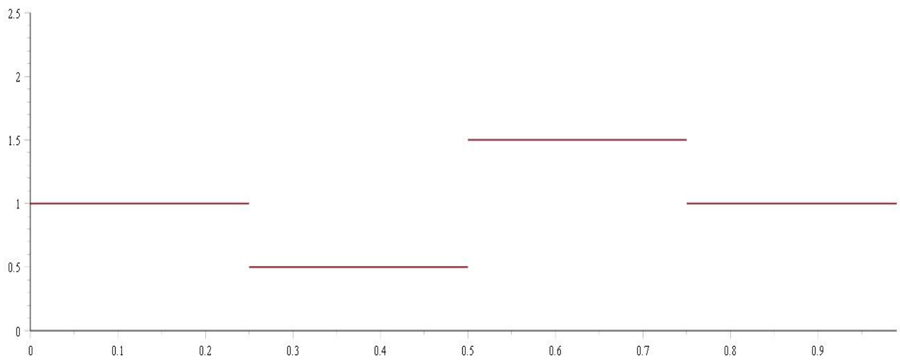
For $n = 2$, it is easy to see that the measures of the four dyadic intervals of length $\frac{1}{4}$ are given in Table 3.6.1.

TABLE 3.6.1

b	00	01	10	11	sum
μ	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	1
μ^2	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{8}$	1

Interpreted another way, restricted to each of these dyadic intervals, these measures are just constant multiples of Lebesgue measure with the scaling constants given in the table. We can find the Radon-Nikodým derivative $G_2 = \frac{d\mu}{d\mu^2}$, which is therefore also constant on these intervals.

FIGURE 3.6.1. The function G_2



We can calculate the functions G_n for larger n in a similar fashion. First we calculate the probability tree for a biased toss followed by two fair ones. Averaging the probabilities over the permutations gives the value of the μ and μ^3 measures of the dyadic intervals, now of length $\frac{1}{8}$.

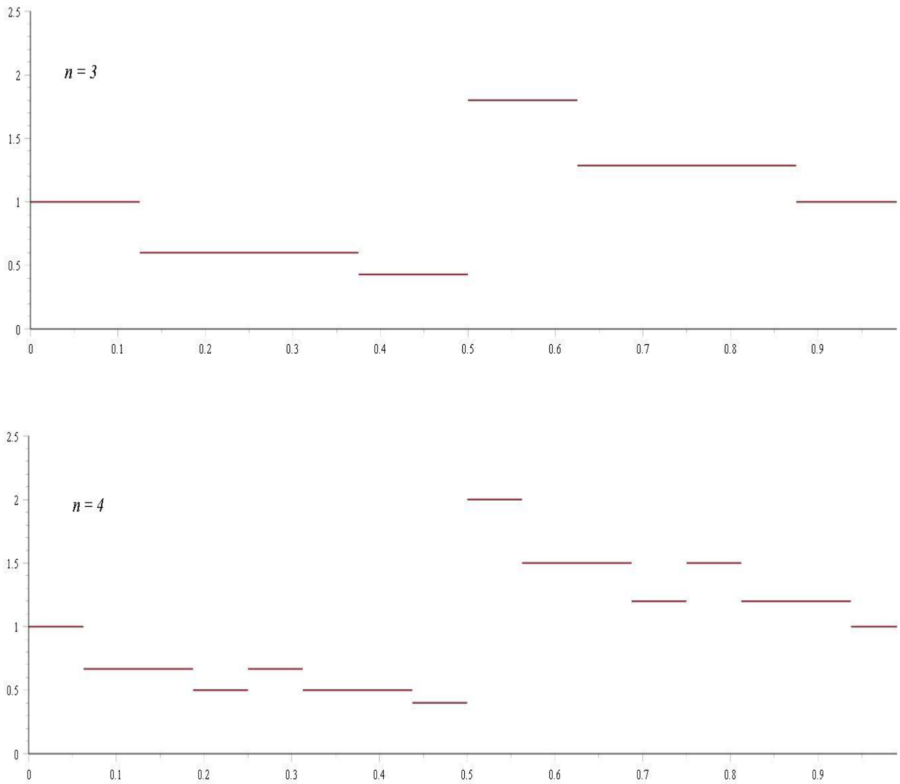
Table 3.6.2

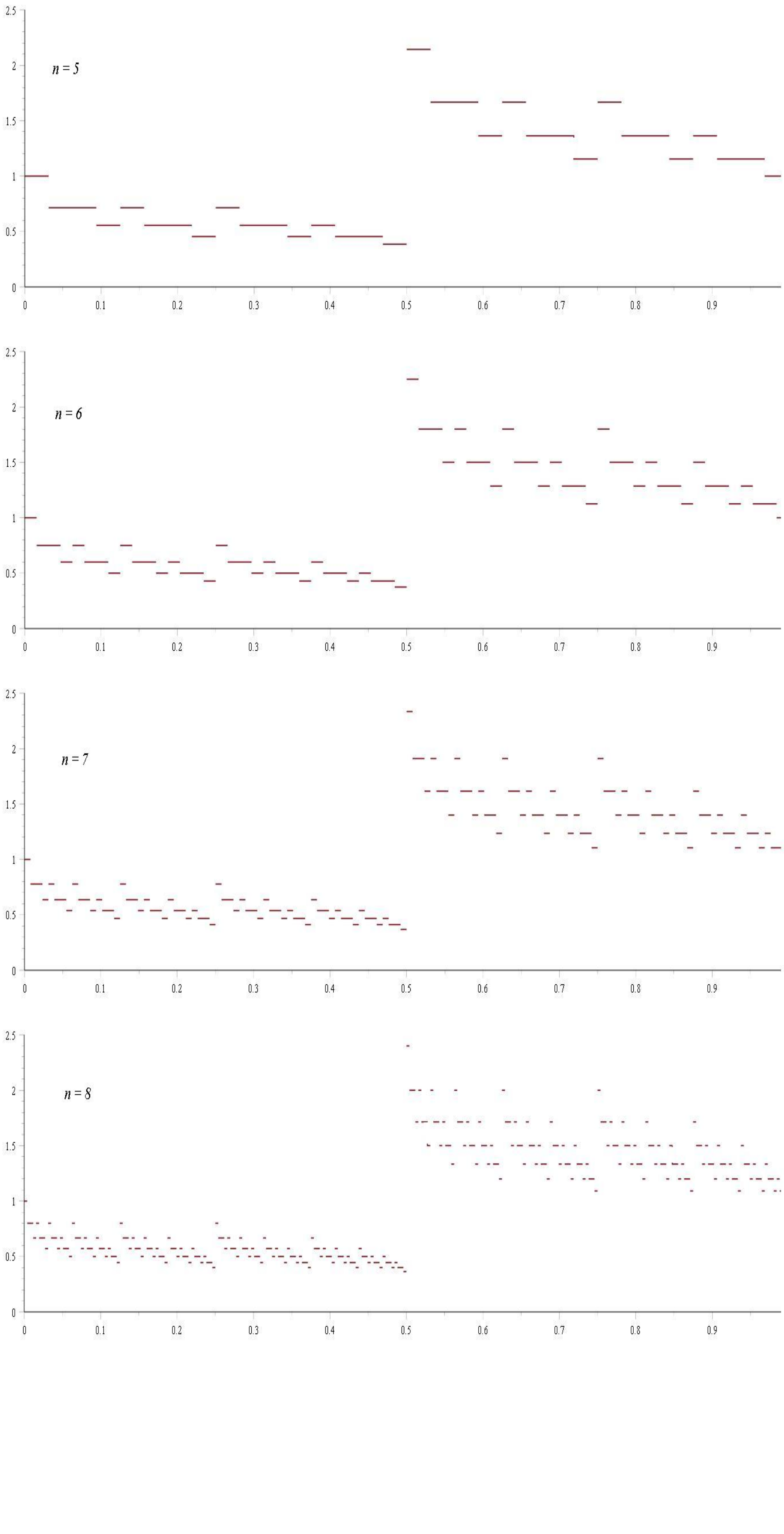
b	000	001	010	011	100	101	110	111	sum
μ	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	1
μ^3	$\frac{1}{16}$	$\frac{5}{48}$	$\frac{5}{48}$	$\frac{7}{48}$	$\frac{5}{48}$	$\frac{7}{48}$	$\frac{7}{48}$	$\frac{3}{16}$	1
$G_3 = \frac{d\mu}{d\mu^3}$	1	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{7}$	$\frac{9}{5}$	$\frac{9}{7}$	$\frac{9}{7}$	1	

This is relatively easily automated by computer. Below are the graphs of $G_3, \ldots G_8$.

As we showed in Section 3.3, the full family $\{G_F\}$ that come from this measure are Σ-normalised and Σ-compatible. For small values of n one can confirm this directly from the graphs in Figure 3.6.2.

Figure 3.6.2. G_n for $n = 3, 4 \ldots 8$




n = 5
n = 6
n = 7
n = 8
0 0.5 1 1.5 2 2.5
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

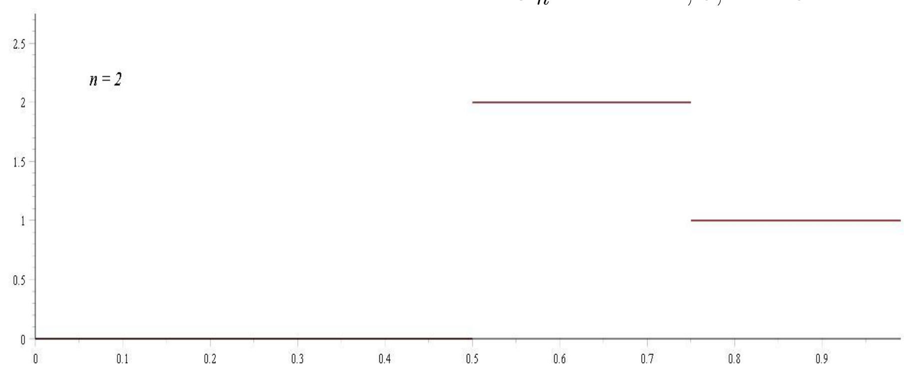
EXAMPLE 3.6.2. An example of G-measures which are not quasi-invariant is given as follows. Let μ_1 be the measure on $\mathbb{Z}_2$, where $\mu_1(0) = 0$ and for $i \geq 2$, let μ_i be the measure on $\mathbb{Z}_2$ with $\mu_i(1) = \frac{1}{2}$. Let $\mu = \otimes_{i=1}^{\infty}\mu_i$ be the corresponding product measure on Z. The same calculations as in the example above can be applied here. The G-measures here are not quasi-invariant since there is an outcome with probability zero, that is the probability of tossing a zero first.

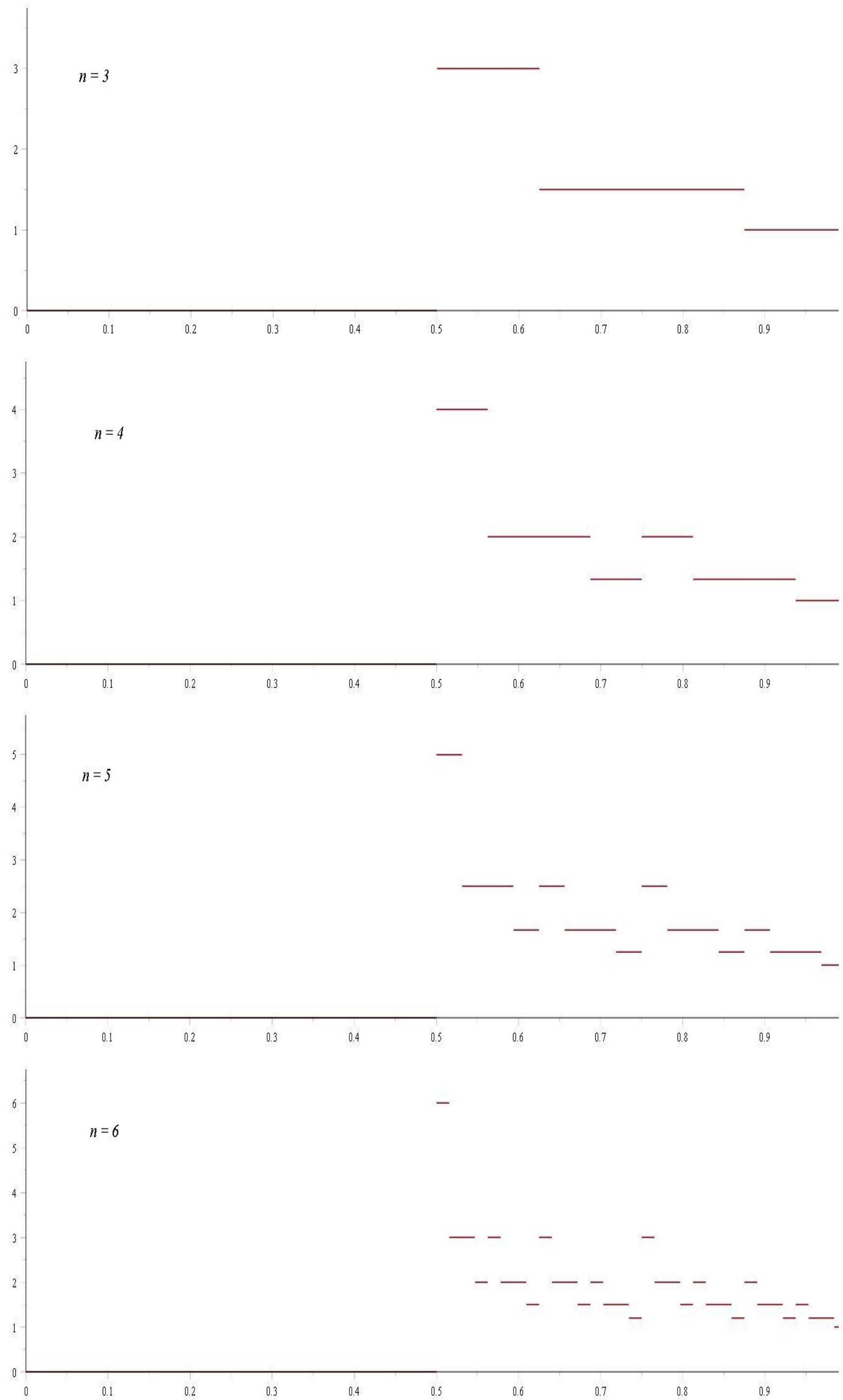
TABLE 3.6.3

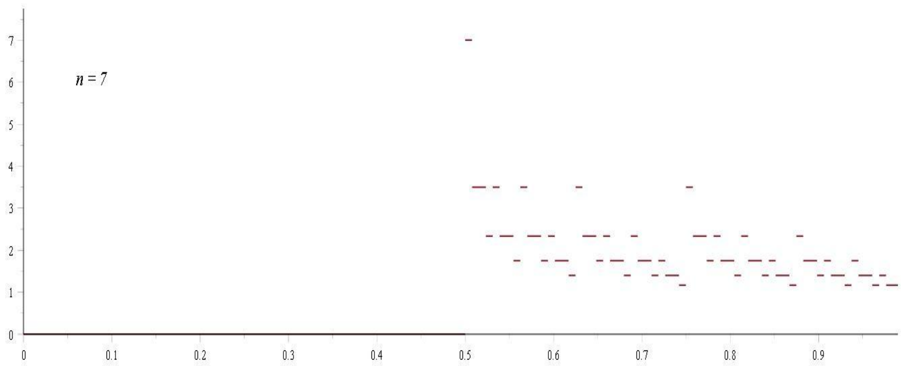
b	00	01	10	11	sum
μ	0	0	$\frac{1}{2}$	$\frac{1}{2}$	1
μ^2	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	1
$G_2 = \frac{d\mu}{d\mu^2}$	0	0	2	1	

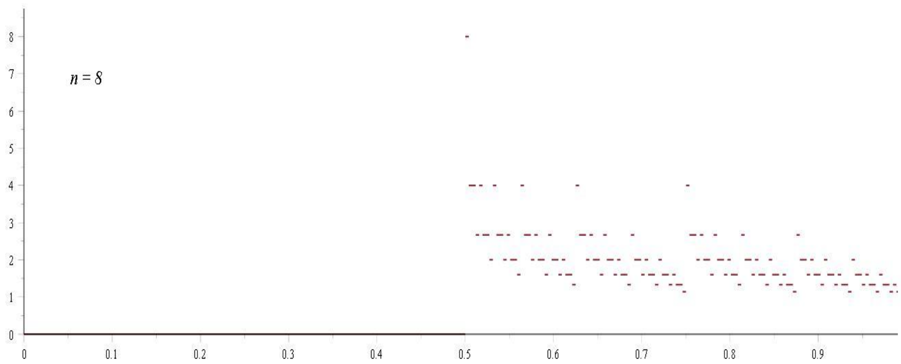
We can calculate the Radon-Nikodým derivatives which are given in the following graphs.

FIGURE 3.6.3. The functions G_n for $n = 2, 3, 4 \ldots 8$




n = 3
n = 4
n = 5
n = 6





For this measure the support of μ is $\left[\frac{1}{2},1\right) = \{1\} \times \prod_{i=2}^{\infty} \mathbb{Z}_2$. To define a G-family on all of Z we need to set $G_n(x) = 1$ for x not in this support.

3.7. An example of a G-measures which is not ergodic

Since we are looking at measures which are unique here. It is interesting to show that there are examples of G-measures which are not unique. We discuss an example here of a G-measure which is not unique, this is given by the rotation of the unit circle. Some of the discussion here of this measure and the proof that the rotation is ergodic under certain conditions is given here from Walters [1982].

Let $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$, let $\mathcal{B}$ be the σ-algebra of Borel subsets of $\mathbb{T}$ and let μ be a normalised Haar measure. Let $a \in \mathbb{T}$ and define the rotation $T : \mathbb{T} \to \mathbb{T}$ by $T(z) = az$. Then T is measure-preserving since μ is Haar measure. Note that $\mathbb{T}$ is a compact topological group. The normalised circular Lebesgue measure μ on $\mathbb{T}$ is a Haar measure. That is, $\mu(az) = \mu(z)$ for all $a \in \mathbb{T}$ and all Borel sets z. Note that T is a transformation since it is a map from a measure space to itself. Therefore, this map is invariant since it is a measure-preserving transformation. We show here that it satisfies the conditions for a G-measure.

The transformation T is called a rotation of $\mathbb{T}$. The rotation $T(z) = az$ of the unit circle is ergodic if and only if a is not a root of unity.

We first need to show that the rotation of the unit circle gives a G-measure. Note that $\mathbb{T} = \mathcal{G}$ of transformations where $\mathcal{G}_n$ is an increasing family of subgroups and

$$\omega_n = \left\{e^{2\pi ik/n} : k = 0, \ldots, n-1\right\} \subseteq \mathcal{G}.$$

Hence $\mathcal{G}_n = \langle \omega_1, \ldots, \omega_n \rangle$. Therefore, $|\mathcal{G}_n| = n$.

By the definition of a Haar measure there exists a probability measure μ defined on the σ-algebra of Borel subsets of $\mathbb{T}$, $\mathcal{B}$, such that

$$\mu(az) = \mu(z)$$

for all $a \in \mathcal{G}$ and $z \in \mathcal{B}$. The tail measure is defined by

$$\begin{aligned}
\mu^n(z) &= \frac{1}{|\mathcal{G}_n|} \sum_{a \in \mathcal{G}} \mu(az) \\
&= \frac{1}{n} \sum_{a \in \mathcal{G}} \mu(az) \\
&= \frac{n\mu(z)}{n} \\
&= \mu(z).
\end{aligned}$$

Therefore, $\mu \in \mathfrak{P}(Z)$ has Radon-Nikodým derivatives which satisfy $G_n = \frac{d\mu}{d\mu^n} = 1$ (a.e.) since $\mu^n = \mu$. We can show that $G_n = 1$ by the Radon-Nikodým Theorem since

$$\begin{aligned} \mu(z) &= \int_z \frac{d\mu}{d\mu^n} d\mu^n \\ &= \int_z 1\, d\mu^n \\ &= \mu^n(z), \end{aligned}$$

this implies that $G_n = 1$.

To show that μ is a G-measure we need to show that $G = \{G_n\}_n$ is a normalised compatible family of Borel functions on $\mathbb{T}$. The normalisation constraint is satisfied since

$$\begin{aligned} \frac{1}{|\mathcal{G}_n|} \sum_{a \in \mathcal{G}} G_n(az) &= \frac{1}{n} n G_n(z) \\ &= G_n(z) \\ &= 1. \end{aligned}$$

Since $G_n(z) = 1$, the family $\{G_n\}$ is Σ-compatible.

We can now prove that the rotation of the unit circle gives an example of a G-measure which is ergodic if and only if a is not a root of unity.

Suppose that a is a root of unity, then $a^p = 1$ for some integer $p > 0$. Let $f(z) = z^p$. Then $f \circ T = f$ and f is not constant a.e.. Therefore T is not ergodic.

Conversely, suppose that a is not a root of unity and $f \circ T = f$, $f \in L^2(\mu)$. Let $f(z) = \sum_{n=-\infty}^{\infty} b_n z^n$ be its Fourier series. Then $f \circ T(z) = \sum_{n=-\infty}^{\infty} b_n a^n z^n$

and therefore, $b_n (a^n - 1) = 0$ for each n. If $n \neq 0$ then $b_n = 0$, and so f is a constant a.e. hence T is ergodic.

3.8. Quasi-invariant Measures on $Z = \prod_{i=1}^{\infty} \mathbb{Z}_2$.

In this section we look at the quasi-invariant measures for permutations on the infinite product space. We show that every G-measure is quasi-invariant.

PROPOSITION 3.8.1. *Every measure associated to a family of G_n's is quasi-invariant.*

PROOF. Suppose that $G = \{G_n\}$ is a Σ-normalised, Σ-compatible family. We want to show here that if μ is a G-measure such that $G_n = \frac{d\mu}{d\mu^n}$ is a Σ-normalised, Σ-compatible family then μ is quasi-invariant for Σ_∞. Note that for any $\tau \in \Sigma_\infty$, $\mu \circ \tau \sim \mu \circ \tau^{-1}$.

Let $A \subseteq Z$ where A is measurable. Suppose that $G_n = \frac{d\mu}{d\mu^n}$. Then $\mu \ll \mu^n$. Hence by the definition of equivalent measures (Definition 2.1)

$$\text{If } \mu^n (A) = 0 \text{ then } \mu (A) = 0. \tag{3.8.1}$$

We say that μ^n is quasi-invariant for the action of Σ_∞ (by Definition 2.1.5) if for $\sigma \in \Sigma_\infty$,

$$\mu^n (A) = 0 \text{ if and only if } \mu (\sigma A) = 0.$$

We first show that $\mu \circ \sigma \sim \mu^n$.

The tail measure defines a family of functions $G = \{G_n\}$ via the Radon-Nikodým derivatives $G_n = \frac{d\mu}{d\mu^n}$ is defined as

$$\mu^n (A) = \frac{1}{|\Sigma_n|} \sum_{\sigma \in \Sigma_n} \mu (\sigma A) . \tag{3.8.2}$$

Therefore, if $\mu^n(A) = 0$ then $\mu(\sigma A) = 0$ since none of the terms in the sum are negative. Suppose that $\mu(\sigma A) = 0$ then $\mu^n(A) = 0$. Hence $\mu \circ \sigma \sim \mu^n$.

We want to show that $\mu \circ \sigma \sim \mu$. To do this we show that $\mu \sim \mu^n$. As given in Equation 3.8.1, if $\mu^n(A) = 0$ then $\mu(A) = 0$. For the other direction, suppose that $\mu(A) = 0$ for all $A \subseteq Z$ then $\mu(\sigma A) = 0$ for all $\{\sigma A\} \subseteq Z$, hence $\mu^n(A) = 0$ for all $\sigma \in \Sigma_n$.

Therefore, $\mu \circ \sigma \sim \mu^n$ and $\mu \sim \mu^n$, hence $\mu \circ \sigma \sim \mu$. That is, μ is quasi-invariant for Σ_∞. $\square$

CHAPTER 4

A Proof of de Finetti's Theorem using Martingales

In this chapter we define a bounded linear operator A_n as follows

$$A_n\left(f\right)\left(x\right)=\frac{1}{\left|\Sigma_n\right|}\sum_{\sigma\in\Sigma_n}G_n\left(\sigma x\right)f\left(\sigma x\right),\ x\in Z.$$

In the last chapter we used this operator to define G-measures. With respect to a suitable measure, this operator turns out to be a conditional expectation operator.

In this chapter, we give some background material on martingales. Then we give another proof of de Finetti's Theorem where we use the theory of martingales. This version of de Finetti's Theorem states that if an infinite sequence is exchangeable then it is conditionally i.i.d. . The link between the theory of martingales and G-measures is original.

We also show that the analogue of the Brown and Dooley (1991) Theorem, that is the unique ergodicity of the finite coordinate changes on the circle, does not occur for the bounded linear operator for the symmetric group. Since this may not converge to a constant and there may not be a unique G-measure. This part of the thesis is original work.

4.1. Background to the Theory of Martingales

The definitions and theorem given here are from Ash and Doléans-Dade [2000], Sousi [2013] and Lalley [2014, p. 19].

Definition 4.1.1. Let $(Z, \mathcal{B}, \mu)$ be a measurable space. A **reverse or backward filtration** is a sequence $\{\mathcal{C}_n\}_{n\leq 0} = \ldots \subseteq C_{-2} \subseteq \mathcal{C}_{-1} \subseteq \mathcal{C}_0$ of σ-algebras on $(Z, \mathcal{B}, \mu)$. That is, it is a decreasing sequence of sub σ-algebras of $\mathcal{C}$, indexed by $n = \{0, -1, -2, \ldots\}$ such that $\mathcal{C}_{n-1} \subseteq \mathcal{C}_n$ for each $n \leq 0$. A sequence of random variables $\{X_n\}_{n\leq 0}$ is said to be adapted to the filtration $\{\mathcal{C}_n\}_{n\leq -1}$ if X_n is $\mathcal{C}_n$-measurable for every n. An adapted sequence X_n is such that for every $n \leq -1$,

$$\mathbb{E}\left(X_{n+1}|\mathcal{C}_n\right) = X_n \text{ a.e..}$$

We say that $\{X_n, \mathcal{C}_n\}$ is a **reverse martingale**.

The **Reverse Martingale Convergence Theorem** is as follows. For the Theorem below see for example Sousi [2013, pp. 23-24].

Theorem 4.1.2. *Let $\{X_n\}_{n\leq 0}$ be a reverse martingale relative to the reverse filtration $\{\mathcal{C}_n\}_{n\leq 0}$. Then*

$$\lim_{n\to-\infty} X_n = \mathbb{E}\left(X_0|\bigcap_{n\leq 0}\mathcal{C}_n\right)$$

almost everywhere and in L^1.

This discussion is based on Miermont [2006]. Sometimes backward martingales are defined as a **forward process** $\{Y_n\}_{n\geq 0}$ with respect to a backwards filtration $\mathcal{C}_0 \supseteq \mathcal{C}_1 \supseteq \mathcal{C}_2 \supseteq \cdots$ such that Y_n is adapted in L^1 and

$$\mathbb{E}\left(Y_n|\mathcal{C}_{n+1}\right) = Y_{n+1}.$$

This is equivalant to our definition of a reverse martingale if we let $Y_n = X_{-n}$ and $\mathcal{C}_n = \mathcal{C}_{-n}$ for all $n \geq 0$. For example, $\mathbb{E}\left(Y_n|\mathcal{C}_{n+1}\right) = Y_{n+1}$ gives $\mathbb{E}\left(X_{-n}|\mathcal{C}_{-(n+1)}\right) = X_{-(n+1)}$, therefore for $n = 0$, $\mathbb{E}\left(X_0|\mathcal{C}_{-1}\right) = X_{-1}$.

Definition 4.1.3. Let $f \in L^1(Z, \mathcal{B}, \mu)$ and $B_0 \subseteq \mathcal{B}$, there is a (up to sets of μ measure zero) function $h \in L^1(Z, B_0, \mu)$ such that

$$\int_A h \, d\mu = \int_A f \, d\mu \quad \text{for all } A \in B_0.$$

We denote h by $\mathbb{E}(f|B_0)$. We call this the **conditional expectation operator**.

Remark 4.1.4. As an aside we consider the case of a unique G-measure for the group of finite coordinate changes as discussed in Brown and Dooley [1991]. Let G be a Γ-normalised Γ-compatible family and $Z = \prod_{n=1}^{\infty} \mathbb{Z}_2$.

As discussed in Brown and Dooley [1991] for the group of finite coordinate changes we let the conditional expectation operator, $\mathbb{E}(f|\mathcal{C}_n)$, be a reverse martingale. We discuss the conditional expectation operator in more detail in Definition 4.1.3. For the group of finite coordinate changes, it is easy to identify $\mathbb{E}(f|\mathcal{C}_n)$ in terms of the linear functionals in Section 3.4. For $n = 1, 2, 3, \ldots$ and $x \in Z$ let

$$\phi_{x,n}(f) = \frac{1}{|\Gamma_n|} \sum_{\gamma \in \Gamma_n} G_n(\gamma x) f(\gamma x). \tag{4.1.1}$$

Then

$$\mathbb{E}(f|\mathcal{C}_n) = \phi_{x,n}(f).$$

Then there exists a unique G-measure, μ, if and only if for all $f \in L^1(Z, \mathcal{B}, \mu)$ and $x \in Z$, Equation 4.1.1 holds. Since the unit ball of $M(Z)$ is weak-$*$ compact, there exists a weak-$*$ convergent sequence, $\phi_{x,n}$, whose limit we shall denote as μ. That is, $\phi_{x,n} \overset{w}{\to} \mu \in M(Z)$. This means that for all $f \in \mathcal{C}(Z)$,

$$\lim_{n\to\infty} \phi_{x,n}(f) = \lim_{n\to\infty} \frac{1}{|\Gamma_n|} \sum_{\gamma\in\Gamma_n} G_n(\gamma x) f(\gamma x) = \int f \, d\mu.$$

We intend to define $\phi_{x,n}(f) = \mathbb{E}(f|\mathcal{C}_n)(x)$.

4.2. Linear Operators for the Symmetric Group

Following on from the examples of $A_n(f)(x)$ discussed in Section 3.4, in this section we discuss the properties of a bounded linear operator A_n for the symmetric group. Suppose that $G = \{G_n\}$ is a Σ-normalised, Σ-compatible family of functions on Z. For each $n \in \mathbb{Z}^+$, we define the averaging map $A_n : \mathcal{C}(Z) \to \mathrm{Borel}(Z)$ where $\mathrm{Borel}(Z)$ is the set of bounded Borel measurable functions by

$$A_n(f)(x) = \frac{1}{|\Sigma_n|} \sum_{\sigma\in\Sigma_n} G_n(\sigma x) f(\sigma x), \; x \in Z. \tag{4.2.1}$$

This map is clearly linear and bounded. We show that the operator norm is one (as G is a Σ-normalised family).

Proposition 4.2.1. *If $m \geq n$ and $A_n(f)(x)$ is given by Equation 4.2.1 then $A_n A_m = A_m A_n = A_{\max(m,n)}$ and $A_n^2 = A_n$.*

PROOF. We now show that for $m \geq n$, $A_n A_m = A_m$.

$$\begin{aligned}
A_n\left(A_m\left(f\right)\right)\left(x\right) &= \frac{1}{\left|\Sigma_n\right|}\sum_{\sigma\in\Sigma_n} G_n\left(\sigma x\right)A_m\left(f\right)\left(\sigma x\right) \\
&= \frac{1}{\left|\Sigma_n\right|}\sum_{\sigma\in\Sigma_n} G_n\left(\sigma x\right)\frac{1}{\left|\Sigma_m\right|}\sum_{\sigma'\in\Sigma_m} G_m\left(\sigma'\sigma x\right)f\left(\sigma'\sigma x\right) \\
&= \frac{1}{\left|\Sigma_n\right|}\sum_{\sigma\in\Sigma_n} G_n\left(\sigma x\right)\frac{1}{\left|\Sigma_m\right|}\sum_{\sigma'\in\Sigma_m} G_m\left(\sigma' x\right)f\left(\sigma' x\right) \\
&= \frac{1}{\left|\Sigma_n\right|}\sum_{\sigma\in\Sigma_n} G_n\left(\sigma x\right)\left(A_m\left(f\right)\left(x\right)\right) \\
&= A_m\left(f\right)\left(x\right) \text{ as } \{G_n\} \text{ is } \Sigma\text{-normalised.}
\end{aligned}$$

Note that similar working shows that $A_m A_n = A_m$. Therefore, $A_n^2 = A_n$. □

PROPOSITION 4.2.2. *Suppose that A_n is given by Equation 4.2.1. Then $\left\|A_n\right\| = 1$.*

PROOF. If $f \in \mathcal{C}\left(Z\right)$ and $x \in Z$ then

$$\begin{aligned}
\left|A_n\left(f\right)\left(x\right)\right| &= \left|\frac{1}{\left|\Sigma_n\right|}\sum_{\sigma\in\Sigma_n} G_n\left(\sigma x\right)f\left(\sigma x\right)\right| \\
&\leq \frac{1}{\left|\Sigma_n\right|}\sum_{\sigma\in\Sigma_n} \left|G_n\left(\sigma x\right)f\left(\sigma x\right)\right| \\
&\leq \frac{1}{\left|\Sigma_n\right|}\sum_{\sigma\in\Sigma_n} G_n\left(\sigma x\right)\left\|f\right\|_\infty \\
&= \left\|f\right\|_\infty \text{ as } \{G_n\} \text{ is } \Sigma\text{-normalised.}
\end{aligned}$$

So $\left\|A_n\left(f\right)\right\|_\infty = \sup_x \left|A_n\left(f\right)\left(x\right)\right| \leq \left\|f\right\|_\infty$ so $\left\|A_n\right\| \leq 1$. Taking $f \equiv 1$ shows that $\left\|A_n\right\| = 1$. □

The properties shown in the previous two propositions are exactly those that one gets with a conditional expectation operator. Note that for any Borel measurable function f on Z, the function $A_n(f)$ is Σ_n-invariant. That is, if we let $\mathcal{C}_n$ denote the σ-algebra of Σ_n-invariant Borel subsets of Z, then $A_n(f)$ is necessarily $\mathcal{C}_n$-measurable.

For each n, let ν_n by the measure $G_n\, d\lambda$, where λ is the Lebesgue measure. Then, for $f \in L^1(Z, \mathcal{B}, \nu_n)$ and $A \in \mathcal{C}_n$

$$\begin{aligned}
\int_A A_n(f)(x)\, d\nu_n(x) &= \int_A \frac{1}{|\Sigma_n|} \sum_{\sigma \in \Sigma_n} G_n(\sigma x) f(\sigma x) G_n(x)\, d\lambda \\
&= \int_A \frac{1}{|\Sigma_n|} \sum_{\sigma \in \Sigma_n} G_n(x) f(x) G_n(\sigma^{-1} x)\, d\lambda \\
&\quad \text{(as } A \text{ is } \Sigma_n\text{-invariant)} \\
&= \int_A G_n(x) f(x) \frac{1}{|\Sigma_n|} \sum_{\sigma \in \Sigma_n} G_n(\sigma^{-1} x)\, d\lambda \\
&= \int_A G_n(x) f(x)\, d\lambda \qquad \text{(as } \{G_n\} \text{ is normalised)} \\
&= \int_A f(x)\, d\nu_n(x).
\end{aligned}$$

Thus $A_n = \mathbb{E}(\cdot|\mathcal{C}_n)$ on this measure space. A similar calculation in fact shows that $\{A_n f\}$ forms a reverse martingale with respect to the nest of σ-algebras $\ldots \mathcal{C}_3 \subseteq \mathcal{C}_2 \subseteq \mathcal{C}_1$.

This martingale will converge almost everywhere and in L^1. In the case of the group of finite coordinate changes, the limit σ-algebra is trivial, just $\{\emptyset, Z\}$, and so the martingale converges to $\int_Z f\, d\lambda$. This is not the case here since the exchangeable σ-algebra $\mathcal{C} = \bigcap_n \mathcal{C}_n$ is more complicated. However it seems likely that $A_n(f)(x)$ does converge to $\int_Z f\, d\lambda$ for almost all $x \in Z$.

Brown and Dooley [1991] prove that the following proposition holds for the group of finite coordinate changes. Note that this proposition does not hold for the symmetric group.

Proposition 4.2.3. *Let $(Z, \mathcal{B}, \mu)$ be a measurable space and $G = \{G_F\}$ be a Γ-normalised Γ-compatible family. The following are equivalent:*

(1) There is a unique G-measure μ - which is therefore ergodic.
(2) The n-net
$$A_n(f)(x) = \frac{1}{|\Gamma_n|} \sum_{\gamma \in \Gamma_n} G_n(\gamma x) f(\gamma x)$$
converges uniformly to a constant for every $f \in \mathcal{C}(X)$.
(3) The n-net
$$A_n(f)(x) = \frac{1}{|\Gamma_n|} \sum_{\gamma \in \Gamma_n} G_n(\gamma x) f(\gamma x)$$
converges pointwise (for all $x \in Z$) to a constant for every $f \in \mathcal{C}(X)$.

4.3. A Proof of de Finetti's Theorem using Martingales

We now give a proof of de Finetti's Theorem using martingales. We begin this section by defining the conditional expectations operator which we show is given by $A_n(f)(x)$. We then prove de Finetti's Theorem using the theory of martingales. The following definition is similar to that of an exchangeable σ-algebra in Definition 2.5.3.

Definition 4.3.1. Let $Z = \prod_{i=1}^{\infty} \mathbb{Z}_2$, suppose $n \geq 1$. Let $\mathcal{C}_n$ be the set of σ-algebra of all Borel subsets of Z which are invariant under Σ_n and $\mathcal{C} = \cap_{n=0}^{\infty} \mathcal{C}_n$.

LEMMA 4.3.2. *Let* $Z = \prod_{i=1}^{\infty} \mathbb{Z}_2$ *and* $\mathcal{C}_n$ *be given in Definition 4.3.1 then for* $f \in L^1(Z, \mathcal{B}, G_n)$,

$$\mathbb{E}(f|\mathcal{C}_n) = A_n(f),$$

where the measure is G_n times Lebesgue measure.

PROOF. By a theorem of Douglas [1965], A_n is a conditional expectation operator. Every $\mathcal{C}_n$ measurable L^1 function is invariant under A_n and hence in the range of A_n. In additon, $A_n(f)$ is always $\mathcal{C}_n$-measurable. This implies that $A_n = \mathbb{E}(\cdot|\mathcal{C}_n)$. □

We define a conditionally i.i.d. sequence which we use in de Finetti's Theorem. This is also discussed in Section 2.7.

DEFINITION 4.3.3. Let $X = (X_1, X_2, \ldots)$ be an infinite exchangeable sequence and $\mathcal{C}$ be a σ-algebra as defined in Definition 4.3.1. The conditions for a sequence X to be **conditionally i.i.d.** given $\mathcal{C}$ are

$$\mathbb{P}(X_i \in A|\mathcal{C}) = \mathbb{P}(X_j \in A|\mathcal{C}) \text{ a.e. for each Borel } A \subseteq \mathbb{R},\ i \neq j.$$

and

$$\mathbb{P}(X_1 \in A_1, \ldots X_k \in A_k|\mathcal{C}) = \mathbb{P}(X_1 \in A_1|\mathcal{C}) \cdots \mathbb{P}(X_k \in A_k|\mathcal{C}).$$

We now give a version of de Finetti's Theorem.

THEOREM 4.3.4. *If an infinite sequence is exchangeable then it is conditionally i.i.d.* .

That is, if we let $\{X_n\}_{n\in\mathbb{Z}^+}$ *be an exchangeable sequence and* $\mathcal{C}$ *is the corresponding exchangeable* σ*-algebra. Then, conditionally on* $\mathcal{C}$*,* $\{X_n\}_{n\in\mathbb{Z}^+}$ *are i.i.d.* .

We can use the theory of martingales to prove de Finetti's Theorem. We give a general outline here of how to prove de Finetti's Theorem. Let f be any bounded Borel function $f_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \dots, k$.

We show that if $X = (X_1, X_2, \dots)$ is exchangeable then

$$\mathbb{E}\left(f_1\left(X_1\right) f_2\left(X_2\right) \cdots f_k\left(X_k\right) |\mathcal{C}\right) = \mathbb{E}\left(f_1\left(X_1\right) |\mathcal{C}\right) \cdots \mathbb{E}\left(f_k\left(X_k\right) |\mathcal{C}\right), \tag{4.3.1}$$

for all $\mathcal{C} = \cap_{n=0}^{\infty} \mathcal{C}_n$ and if we let, $f_i = \chi_{A_i}$ for $A_i \in \mathcal{B}(\mathbb{R})$ then Equation 4.3.1 becomes

$$\mathbb{P}\left(X_1 \in A_1, \dots X_k \in A_k|\mathcal{C}\right) = \mathbb{P}\left(X_1 \in A_1|\mathcal{C}\right) \cdots \mathbb{P}\left(X_k \in A_k|\mathcal{C}\right). \tag{4.3.2}$$

Therefore, we have a conditionally i.i.d. sequence (X_i). Showing that Equation 4.3.1 holds will require us to use the Martingale Convergence Theorem (Theorem 4.1.2).

Remark 4.3.5. The converse of Theorem 4.3.4, that is that a conditionally i.i.d. sequence is exchangeable, is false. In the other direction to Theorem 4.3.4, suppose that the random variables $X_1, \dots, X_n, \dots$ are conditionally i.i.d. then we can give a counterexample to show that this is not necessarily exchangeable. This is given by the Polya's Urn model, which is discussed in Example 2.5.4 . This gives an example of exchangeable random variables that are statistically dependent. Therefore, by the contrapositive, a conditionally i.i.d. sequence is not necessarily exchangeable.

We now consider de Finetti's Theorem for the symmetric group. Some of these details here are from OpenCourseWare [2013] and Žitković [2010].

DEFINITION 4.3.6. For the symmetric group, let $\mathcal{C}_n$ denote the tail σ-algebra generated by all cylinder sets $\Pi_n A_n$. Then $\mathcal{C}_n$ is a **reverse or backward filtration**. Let the **forward process** $\{f\}_{n\geq 0}$ be defined with respect to the backward filtration $\mathcal{C}_0 \supseteq \mathcal{C}_1 \supseteq \mathcal{C}_2 \supseteq \cdots$ such that $\{f_n\}$ is adapted on L^1 and

$$\mathbb{E}\left(f_n|\mathcal{C}_{n+1}\right) = f_{n+1}.$$

DEFINITION 4.3.7. For $k \leq n \in \mathbb{N}$, let Σ_n^k denote the set of injections $\sigma : \{1, 2, \ldots, k\} \to \{1, 2, \ldots, n\}$. Then Σ_n^k has $k!\binom{n}{k} = \dfrac{n!}{(n-k)!}$ elements. The set of all permutations of the set $\{1, 2, \ldots, n\}$ is given by Σ_n^n . Given a function $f : \mathbb{R}^k \to \mathbb{R}$, for $n \geq k$, we let the function $f_n^{\mathrm{sim}} : \mathbb{R}^n \to \mathbb{R}$ be defined as

$$f_n^{\mathrm{sim}}\left(x_1, \ldots, x_n\right) = \frac{(n-k)!}{n!} \sum_{\sigma \in \sum_n^k} f\left(x_{\sigma(1)}, \ldots, x_{\sigma(k)}\right).$$

This is called the n-**symmetrization** of f.

The proof of de Finetti's Theorem requires the following result due to Žitković [2010, Lemma 11.29].

PROPOSITION 4.3.8. *Let $\{X_n\}_{n\in\mathbb{Z}^+}$ be an i.i.d. sequence, let $f : \mathbb{R}^k \to \mathbb{R}$, $k \in \mathbb{Z}^+$ be a bounded Borel function and $\underset{\sim}{X}_n = (X_1, \ldots, X_n)$.*

$$\begin{aligned} f_n^{sim}\left(\underset{\sim}{X}_n\right) &= \mathbb{E}\left(f_n^{sim}\left(\underset{\sim}{X}_k\right)|\mathcal{C}_n\right), \ a.e. \\ &= \frac{1}{n(n-1)\cdots(n-k+1)} \sum_{\sigma\in\sum_n^k} \mathbb{E}\left(f\left(X_{\sigma(1)}, \ldots, X_{\sigma(k)}\right)|\mathcal{C}_n\right). \end{aligned} \tag{4.3.3}$$

and

$$f_n^{sim}\left(\underset{\sim}{X}_n\right) \to \mathbb{E}\left(f\left(\underset{\sim}{X}_k\right)|\mathcal{C}\right), \quad a.e. \text{ and in } L^1 \text{ as } n \to \infty. \tag{4.3.4}$$

PROOF. The first step in the proof is to show that for all $\sigma \in \Sigma_n^k$

$$\mathbb{E}(f(X_{\sigma(1)}, \ldots, X_{\sigma(k)}|\mathcal{C}_n) = \mathbb{E}\left(f(X_1, \ldots, X_k)|\mathcal{C}_n\right).$$

Žitković [2010, Lemma 11.29] does this by showing that for any symmetric function $g : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}(g(X_1, \ldots, X_n)(f(X_{\sigma(1)}, \ldots, X_{\sigma(k)}) - f(X_1, \ldots, X_k)) = 0.$$

The identity in Equation 4.3.3 follows easily from this by the linearity of the conditional expectation operator. From Equation 4.3.4 we can deduce that $\{f_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right)\}_{n=k}^{\infty}$ is a reverse martingale with respect to the nest of σ-algebras $\{\mathcal{C}_n\}_{n=k}^{\infty}$.

To prove Equation 4.3.4 holds we then combine

$$f_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) = \mathbb{E}\left(f\left(\underset{\sim}{X}_k\right)|\mathcal{C}_n\right),$$

the definition of $\mathcal{C} = \cap_{n=0}^{\infty}\mathcal{C}_n$ of the exchangeable σ-algebra and the backward martingale convergence theorem, Theorem 4.1.2. $\square$

We now give the details of a proof of de Finetii's Theorem (Theorem 4.3.4) which uses the theory of martingales.

PROOF. Let f be of the form $f = gh$, where $g : \mathbb{R}^{k-1} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ are bounded Borel functions and $n \geq k$ for all $x \in \mathbb{R}$. We now consider the following term from Žitković [2010], which we use to derive the results that we want about the relationship between the terms f_n^{sim}, g_n^{sim} and h_n^{sim}.

$$\begin{aligned}
&\frac{n!}{(n-k+1)!} g_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) n h_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) \\
&= \sum_{\sigma \in \sum_n^{k-1}} g\left(X_{\sigma(1)}, \ldots, X_{\sigma(k-1)}\right) \sum_{i \in \{1,\ldots,n\}} h\left(X_i\right) \\
&= \sum_{\sigma \in \sum_n^{k}} g\left(X_{\sigma(1)}, \ldots, X_{\sigma(k-1)}\right) h\left(X_{\sigma(k)}\right) \\
&\quad + \sum_{\sigma \in \sum_n^{k-1}} g\left(X_{\sigma(1)}, \ldots, X_{\sigma(k-1)}\right) \sum_{j=1}^{k-1} h\left(X_{\sigma(j)}\right). \qquad (4.3.5)
\end{aligned}$$

It is easy to show that if we set $f^j\left(\underset{\sim}{X}_{k-1}\right) = g\left(\underset{\sim}{X}_{k-1}\right) h\left(X_j\right)$, $1 \leq j \leq k-1$, then

$$\frac{n!}{(n-k+1)!} g_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) n h_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) = \frac{n!}{(n-k)!} f_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) + \frac{n!}{(n-k+1)!} \sum_{j=1}^{k-1} f_n^{j,\mathrm{sim}}\left(\underset{\sim}{X}_n\right)$$

Dividing by $\frac{n!}{(n-k)!}$ gives

$$\frac{n}{n-k+1} g_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) h_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) = f_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) + \frac{1}{n-k+1} \sum_{j=1}^{k-1} f_n^{j,\mathrm{sim}}\left(\underset{\sim}{X}_n\right)$$

Therefore,

$$f_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) = \frac{n}{n-k+1} g_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) h_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) - \frac{1}{n-k+1} \sum_{j=1}^{k-1} f_n^{j,\mathrm{sim}}\left(\underset{\sim}{X}_n\right)$$

The sum $\sum_{j=1}^{k-1} f_n^{j,\mathrm{sim}}$ is bounded so letting $n \to \infty$ gives

$$\lim_n \left| f_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) - g_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) h_n^{\mathrm{sim}}\left(\underset{\sim}{X}_n\right) \right| = 0.$$

Therefore, by Equation 4.3.4 applied to f_n^{sim}, g_n^{sim} and h_n^{sim} gives

$$\mathbb{E}\left(f\left(X_1, \ldots, X_k\right) | \mathcal{C}\right) = \mathbb{E}\left(g\left(X_1, \ldots, X_{k-1}\right) | \mathcal{C}\right) \mathbb{E}\left(h\left(X_k\right) | \mathcal{C}\right).$$

Repeating this procedure with $g = g'h'$, for bounded Borel function $g' : \mathbb{R}^{k-1} \to \mathbb{R}$ and $h' : \mathbb{R} \to \mathbb{R}$ we can split the conditional expectation into the product $\mathbb{E}\left(g'\left(X_1, \ldots, X_{k-2}\right)|\mathcal{C}\right)\mathbb{E}\left(h'\left(X_{k-1}\right)|\mathcal{C}\right)$. After $(k-1)$ such steps, we get

$$\mathbb{E}\left(h_1\left(X_1\right)h_2\left(X_2\right)\cdots h_k\left(X_k\right)|\mathcal{C}\right) = \mathbb{E}\left(h_1\left(X_1\right)|\mathcal{C}\right)\cdots\mathbb{E}\left(h_k\left(X_k\right)|\mathcal{C}\right).$$

This holds for any bounded Borel function $h_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, k$. If we let, $h_i = \chi_{A_i}$ for $A_i \in \mathcal{B}(\mathbb{R})$ then

$$\mathbb{P}\left(X_1 \in A_1, \ldots X_k \in A_k|\mathcal{C}\right) = \mathbb{P}\left(X_1 \in A_1|\mathcal{C}\right)\cdots\mathbb{P}\left(X_k \in A_k|\mathcal{C}\right).$$

Therefore, we have a conditionally i.i.d. sequence (X_i).

We have shown here that any infinite sequence is exchangeable then it is conditionally i.i.d. . Therefore, we have shown that a version of de Finetti's Theorem applies here (see Theorem 4.3.4). $\square$

Bibliography

J. Aaronson. *An Introduction to Infinite Ergodic Theory.* Number 50 in Mathematical Surveys and Monographs (Book 50). American Mathematical Society, 1997.

D Aldous. Representations for Partially Exchangeable Arrays of Random Variables. *J. Multivariate Anal.*, 11, Issue 4:581–598, 1981.

D. Aldous. *Exchangeability and related topics*, volume 1117. Lecture Notes in Mathematics, 1985.

D. Aldous. Exchangeability and Continuum Limits of Discrete Random Structures. *Proceesings of the International Congress of Mathematicians, Hyderabad, India*, 2010.

D. Aldous and J. Pitman. On the Zero-One Law for Exchangeable Events. *Ann. Probab.*, 7(4):553–743, 1979.

R. Ash and C. Doléans-Dade. *Probability and Measure Theory.* 2nd edition, 2000.

W. Basener. *Topology and Its Applications.* John Wiley & Sons, 2006.

G. Birkhoff. Proof of the ergodic theorem. *Proc Natl Acad Sci USA*, 17(12): 656–660, 1931.

V. Bogachev. *Measure Theory*, volume 1 and 2. Berlin; New York; Springer, 2007.

G. Brown. Integrability of Fourier transforms under an ergodic hypothesis. *J. Austral. Math. Soc. Ser. A*, 26(2):129–153, 1978.

G. Brown and A. Dooley. Odometer actions on G-measures. *Ergodic Theory Dynam. Systems*, 11:279–307, 1991.

G. Brown and A. Dooley. Dichotomy theorems for G-measures. *Internat. J. Math*, 5:827–834, 1994.

G. Brown and W. Moran. On orthogonality of Riesz products. *Proc. Camb. Phil. Soc.*, 76(1):173–181, 1974.

M Capiński and E. Kopp. *Measure, Integral and Probability.* London, Springer Undergraduate Mathematics Series, 2nd edition, 2004.

B. de Finetti. *Probability, Induction and Statistics.* Wiley, New York, 1972a.

B. de Finetti. *Probability, Induction and Statistics.* Wiley, New York, 1972b.

B. de Finetti. La probabilità e la statistica nei rapporti con l'induzione; Secondo i diversi Punti di vista. *Lectures given at a Summer School of the Centro Internazionale Matematico Estivo (C.I.M.E.) held in Varenna (Como), Italy, June 1-10, 1959 Corso C.I.M.E su Induzione e Statistica, Rome*, 18: 1–122, 2011.

P. Diaconis. Finite forms of de Finetti's theorem on exchangeability. *Synthese*, 36:271–281, 1977.

P. Diaconis and D. Freedman. Finite exchangeable sequences. *Ann. Probab.*, 8 No. 4:639–859, 1980a.

P. Diaconis and D. Freedman. De Finetti's Theorem for Markov Chains. *Ann. Probab*, 8(1):115–130, February 1980b.

A. Dooley and A. Fan. Chains of Markovian projections and (G, Gamma)-measures. In N. Kono and N. Shieh, editors, *Trends in Probability and Related Fields, Proceedings of SAP'96. Singapore: World Scientific*, pages 101–116, 1997.

A. Dooley and T. Hamachi. Nonsingular dynamical systems, Bratteli diagrams and Markov odometers. *Israel J. Math*, 138(1):93–123, 2003.

T. Dooley. G-measures. University of New South Wales, 2007. URL `http://www.maths.usyd.edu.au/u/donaldc/gbrown/dooley.pdf`.

R. G. Douglas. Contractive projections on l1 space. *Paci*, 15(2), 1965.

D. Dummit and R. Foote. *Abstract Algebra*. New York: Wiley, 3rd ed edition, 2004.

G. Elliott, P. Fillmore, and B. Rajarama. *Lectures on operator theory*. American Mathematical Society, 1999.

W. Ericson. A Bayesian Approach to 2-Stage Sampling. Technical Report 26, Department of Statistics, University of Michigan, Ann Arbor., 1973.

D. Freedman. A remark on the diference between sampling with and without replacement. *Journal of the American Statistical Association*, 72(359):681, 1977.

E. Glasner. *Ergodic Theory via Joinings*. Mathematical surveys and monographs. American Mathematical Society, 2003. Issue 101.

A. Haar. Der massbegriff in der theorie der kontinuierlichen gruppen. *Ann. Math.*, 2(34):147–169, 1933.

P. Halmos. *Measure Theory*. Van Nostrand, 1950.

P. Halmos. *Lectures on Ergodic Theory*. American Mathematical Soc., 1956.

R. Haydon. A New Proof that every Polish Space is the Extreme Boundary of a Simplex. *Bull. London Math. Soc.*, 7(1):97–100, 1975.

E. Hewitt and L. Savage. Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.*, 80(2):470–501, November 1955.

E. Hewitt and K. Stromberg. *Real and Abstract Analysis: A Modern Treatment of the Theory of Functions of a Real Variable*, volume 25 of Graduate Texts in Mathematics. Springer-Verlag Berlin, 1965.

J. Hoffmann-Jorgensen, J. Kuelbs, and M. Marcus. *Probability in Banach Spaces, 9.* Springer Science & Business Media, 2012.

J-P. Kahane. Commutative Harmonic Analysis. *A Panorama of Hungarian Mathematics in the Twentieth Century*, 14:159–192, 2006. Bolyai Society Mathematical Studies.

M. Keane. Strongly mixing g-measures. *Inventiones math.*, 16(4):309–324, 1972.

G. Kerns and G. Székely. De finetti's Theorem for Abstract Finite Exchangeable Sequences. *J. Theoret. Probab.*, 19(3):589–608, December 2006.

A. Kolmogorov. A new metric invariants of transitive dynamical systems and automorphisms of Lebesgue spaces. *Trudy Mat. Inst. Steklov.*, 169:861–864, 1958. (in Russian).

A. Kolmogorov. Entropy per unit time as a metric invariant of automorphism. *Dokl. Akad. Nauk SSSR*, 124:754–755, 1959. (in Russian).

E. Kreyszig. *Introductory Functional Analysis with Applications.* John Wiley & Sons, 1989.

S. Lalley. Conditional expectations and martingales, 2014. URL `http://galton.uchicago.edu/~lalley/Courses/385/Martingales.pdf`.

A. Lauriničikas. *Weak convergence of measures.* 2012.

S. Lauritzen. Exchangeability and de Finetti's Theorem, April 2007. URL `http://www.stats.ox.ac.uk/$\mathtt{\sim}$steffen/teaching/grad/definetti.pdf`.

J.L Lebowitz and O. Penrose. Modern ergodic theory. *Physics Today*, 26: 23–29, 1973.

O. Makarenkov. Ergodic Theory Lecture Notes. URL `http://www2.imperial.ac.uk/$\mathtt{\sim}$omakaren/ErgodicTheory/$\newline$ErgodicTheoryNotes.pdf`.

G. Miermont. Advanced probability michaelmas, 2006. URL `https://www.andrew.cmu.edu/user/calmost/pdfs/probabilitylec.pdf`.

K. Olav. *Probabilistic Symmetries and Invariance Principles.* Dordrecht : Springer-Verlag New York Inc, 2005.

G. Olshanski. *An introduction to harmonic analysis on the infinite symmetric group.* Springer LNM, 2008.

MIT OpenCourseWare. Martingale convergence theorem, 2013. URL `http://ocw.mit.edu/courses/sloan-school-of-management/`.

D. Ornstein. *Ergodic theory, randomness and dynamical systems*, volume 5. New Haven abd London: Yale University Press, 1974.

K. Petersen. *Ergodic theory*, volume 2. Cambridge University Press, 1983. Cambridge studies in advanced mathematics.

R. Phelps. *Lectures on Choquet's Theorem*, volume 1757 of *Lecture Notes in Mathematics*. Springer-Verlag Berlin Heidelberg New York, 2nd edition, 2001.

M. Riesz. Sur les maxima des formes bilinèaires et sur les fonctionelles linéaires. *Acta Math.*, 49(3-4):465–497, 1927.

H. Royden. *Real Analysis.* MacMillan: New York, 2nd ed. edition, 1969.

W. Rudin. *Real and Complex Analysis.* McGraw-Hill International Editions, 3rd edition, 1966.

B. Simon. *Convexity : An Analytic Viewpoint.* Camrbridge, UK ; New York : Cambridge University Press, 2011.

Y. Sinaǐ. On the Concept of Entropy of a Dynamical System. *Dokl. Akad. Nauk SSSR*, 124:768–771, 1959. (in Russian).

P. Sousi. Advanced probability, 2013. URL `http://www.statslab.cam.ac.uk/~ps422/mynotes.pdf`.

R. Sun. Lecture notes, 2014. URL `http://www.math.nus.edu.sg/~matsr/ProbII/Lec1.pdf`.

T. Tao. 245b, notes 12: Continuous functions on locally compact hausdorff spaces, 2009. URL `https://terrytao.wordpress.com/`.

S. Taylor. Simplices, 2009. URL `http://personal.colby.edu/personal/s/sataylor/teaching/S09/MA331/Simplices.pdf`.

G. Žitković. Theory of probability - measure theory, classical probability and stochastic analysis lecture notes, 2010. URL `http://www.ma.utexas.edu/users/gordanz/notes/theoryofprobabilityI.pdf`.

J. von Neumann. Proof of the quasi-ergodic hypothesis. *Proc Natl Acad Sci U S A.*, 18(1):70–82, Jan 1932a.

J. von Neumann. Physical applications of the ergodic hypothesis. *Proc Natl Acad Sci U S A.*, 18(3):263–266, Mar 1932b.

P. Walters. *An Introduction to Ergodic Theory*, volume 79 of Graduate Texts in Mathematics. Springer-Verlag: New York, 1982.

D. Williams. *Probability with Martingales.* Cambridge University Press, 1991.